https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Differentiable Hybrid Hydrological Model for Short-Term Flood Forecasting with Future Meteorological Information

Leijing Li[1,2,3], Liutianjiao Hong[1], Jianzhu Li[1], Peng Shi[3], Shuaihang Wang[3], Jiyang Tian[4]

[1]State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

[2]Guizhou Water Conservation Science and Research Institute, Guiyang 550002, Guizhou, China

[3]Guizhou Provincial Hydrology and Water Resources Bureau, Guiyang 550002, Guizhou, China

[4] China Institute of Water Resources and Hydropower Research, Beijing 100038, China

*Corresponding author: Liutianjiao Hong, PhD

E-mail: hongliutianjiao@tju.edu.cn

Abstract: Short-term flood forecasting is essential for flood control operation and risk warning in small- and medium-sized basins, yet its accuracy is constrained by rainfall uncertainty, hydrological model structural limitations, and watershed physical response characteristics. To address these challenges, this study developed a differentiable hybrid hydrological forecasting framework that coupled future meteorological information with physically representations of runoff generation, baseflow recession, flow concentration, and channel routing. A physics-informed neural architecture search (PINAS) routing scheme was further proposed and systematically compared with differentiable Muskingum routing and convolutional unit-hydrograph routing, while a purely data-driven LSTM model was used as the benchmark. The framework was evaluated in three representative watersheds in Hebei Province, China, namely the Shahe basin (2210 km^2), Jumahe basin (1760 km^2), and Liulin basin (57.4 km^2), using rainfall–runoff observations from the flood

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

seasons of 2000–2023 and radar-echo data from 2018–2023. Multi-lead flood forecasting experiments were conducted for lead times of 0.5–2.0 h. The results showed that the LSTM benchmark achieves high numerical accuracy in some high-peak flood events, but lacks explicit physical constraints. In contrast, the model with PINAS routing exhibited the most stable overall performance among the hybrid hydrological models, achieving a more balanced representation of flood-peak propagation, hydrograph smoothing, and recession preservation. The data-dependency analysis indicated that increasing the amount of training data improved forecasting stability, but model skill does not increase strictly monotonically with sample size, sample representativeness and flood-type coverage were also critical. SHAP interpretability analysis further revealed that forecasts in the Shahe and Jumahe basins were mainly controlled by radar echoes and their extrapolated features, whereas the small Liulin Basin was more strongly influenced by measured areal rainfall and gauge rainfall. These findings demonstrated that integrating physical constraints, differentiable runoff–routing structures, and future meteorological information can improve the stability, physical consistency, and hydrological interpretability of short-term flood forecasting in small- and medium-sized basins.

Keywords: Flood forecasting; Hybrid hydrological model; Neural architecture search; Radar echo extrapolation; SHAP interpretability

1 Introduction

Floods are one of the most destructive natural hazards worldwide, and the impacts are further amplified by rapid hydrometeorological variability and the pronounced spatiotemporal nonlinearity of rainfall–runoff transformation (Matanó et al., 2024). Short-term flood forecasting is useful for

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

emergency response, reservoir operation, flood management and decision making. However, errors in calculation of runoff generation and channel routing may lead to large inaccuracies in forecasted peak discharge and arrival time, making accurate and reliable flood forecasting highly challenging. (Xiang et al., 2025; Tanhapour et al., 2025).

Two main types of flood forecasting techniques are data-driven model (DDM) and process-driven model (PDM). Although PDM can characterize runoff generation and routing with strong physical interpretability, the forecasting capability is limited by structural deficiencies and parameter uncertainty, particularly when calibration primarily relies on outlet measured discharge in basins with limited data (Gan et al., 2018; Cibin et al., 2014). With strong non-linear representational capabilities, DDM, especially deep learning techniques, show significant potential in rainfall-runoff forecasting. However, purely data driven models may show reduced generalization and lack clear physical consistency and interpretability with extreme data (Amini et al., 2024). These shortcomings have motivated rapid development of hybrid hydrological modeling, which aims to combine the explanatory structure of physical mechanism with the learning-based components (Xu et al., 2024). Differentiable modeling explicitly constructs hydrological process equations and embeds them into an automatic differentiation framework, enabling end-to-end training of the models (Gao et al., 2024; Huynh et al., 2025; Jamaat et al., 2025; Feng et al., 2022). In contrast to Physical Informed Neural Networks (PINNs), which impose physical constraints through loss functions, differentiable models retain a clear mechanistic framework and introduce learning components that are difficult to parameterize (Herat et al., 2025). Recently, several studies have developed distributed hybrid hydrological models for flood simulation. Zhong et al. (2024) developed a distributed physics-informed deep learning model under a differentiable parameter

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

learning framework, highlighting how regionalized deep learning can map watershed attributes to parameters and improve forecasting in data-scarce contexts. Wang et al. (2024) advanced physics-encoded deep learning for distributed modeling and proposed that routing remains difficult to embed in neural architectures in a physically consistent manner. Song et al. (2025) further discussed national-scale differentiable modeling and showed that routing choices can substantially affect computational cost and scalability in high-resolution settings. These studies showed that, building reliable distributed hybrid hydrologic models requires not only a physically meaningful runoff-generation backbone, but also a routing design that is computationally feasible and physically defensible.

For distributed hybrid hydrologic models, channel routing is becoming more widely acknowledged as a bottleneck (Mendoza et al., 2023; Collins et al., 2025). Although Muskingum routing is still the default option, it necessitates recursive calculations from upstream to downstream, which significantly raise computational costs during iterative training (Ayvaz et al., 2017). When the routing is trained in a differentiable manner, the data-driven routing operator performs exceptionally well and show promise in learning complex spatiotemporal relationships (He et al., 2025a). However, the learned routing can turn into a black-box operation that may violate hydrologic constraints, fail to maintain physically meaningful wave dynamics, strictly conserve water mass, and degrade robustness and out-of-sample generalization, which are particularly important for flood forecasting and model transfer across basins (He et al., 2025b). Such routing deficiencies can most obviously show up as distorted rising limbs and biased peak timing in event-based flood nowcasting for small-to-medium size watersheds with short lead times, decreasing the operational effectiveness of hybrid models. As a result, creating a routing approach that is both

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

physically consistent and learnable continues to be a major research issue.

The selection of a routing architecture that achieves the best balance between expressiveness, stability, and physical plausibility remains a fundamental methodological challenge even after the idea of physics informed learnable routing has been adopted (Feng et al., 2022). The components of differentiable models are usually determined by manual design, which can be challenging to transfer across basins, temporal resolutions, and event regimes (Sawadekar et al., 2025). Manually selected designs may quickly become either under-parameterized or over-parameterized in short-term lead time (Astagneau et al., 2024). Neural architecture search (NAS) has recently been used in the field of hydrological forecasting to automate model structure design, such as water level prediction (Ng et al., 2025) and flood forecasting (Yao et al., 2023; Li et al., 2024), thereby reducing reliance on manual expertise and iterative parameter tuning. Similar studies have also been carried out in related hydro-meteorological forecasting, such as NAS-assisted rainfall forecast post-processing (Wen et al., 2024). In the field of hydrological machine learning, architecture search has been recognized as a more reproducible and comparable technique for model construction compared to empirically selecting network architectures (Chadalawada et al., 2020). Furthermore, the routing design must satisfy conservation and non-negativity constraints, merely using general network architecture search techniques may produce correct but physically incomprehensible solutions. Therefore, incorporating NAS into routing design while limiting searches within the physics-informed routing operator space to yield convergence structures with better forecasting capabilities and greater conformity to hydrological physical plausibility.

Apart from the structure of the hydrological model, the main limiting factor in short-term event forecasting is typically the uncertainty in meteorological parameters. (Cafiero et al., 2025). Effective

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

forecasting requires some depiction of future precipitation, as measured rainfall and radar data can only provide information about the past and present (Li et al., 2024). Because meteorological radar can provide high-resolution and rapidly updated observations of convective systems, it plays an important role in precipitation nowcasting. To generate future reflectivity fields and accompany precipitation nowcasting, radar echo extrapolation techniques have been developed (Ravuri et al., 2021). These techniques range from conventional motion-field advection to deep-learning sequence models (Zheng et al., 2024; Song and Han, 2025). However, the practical application value of radar in hydrological forecasting is still unclear. This is because nowcasting accuracy usually decreases quickly with lead time, and hydrological reaction filters meteorological uncertainty in a scale-dependent way. Therefore, current research focuses on determining the extent to which various data derived from radar echo analysis can improve the accuracy of flood forecasts.

Motivated by the challenges in hydrological forecasting, A hybrid multiscale flood forecasting framework is proposed, which focuses on the runoff generation and concentration, employing a modular routing representation for forecasting of multi-period hydrograph, and attempt to incorporate future meteorological information based on radar echo extrapolation to improve forecast accuracy. The Physics-Informed Neural Architecture Search (PINAS) routing has been proposed, which is guided by physics-consistency constraints and learned in a physically motivated architecture space, and compared it with learnable convolution unit hydrograph routing and differentiable Muskingum routing. In summary, the objective of this study is to investigate the following scientific questions:

(1) How does routing design affect hydrological processes throughout entire watershed areas during short lead times?

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

(2) Is it possible for PINAS to improve hydrographs simulation while maintaining physical consistency and transferability across events and basins?

(3) In the short-term forecasting of flood events, which characteristics will have a decisive impact? Will the radar echo data bring about positive contributions?

2 Materials and Methods

2.1 Study area and data

Three representative basins in Hebei Province, China, namely the Jumahe, Shahe, and Liulin basins, have been selected as case study areas. The Shahe basin covers an area of 2210 km^2, a main-channel length of 124 km, and an average basin width of 34.5 km. Six meteorological stations are located within the basin, and the Fuping hydrological station is situated at the basin outlet. The Jumahe basin covers an area of 1760 km^2, with a main channel length of 81.5 km and an average basin width of 25.4 km. It contains eight meteorological stations, and discharge is measured at the Zijingguan hydrological station at the outlet. The Liulin basin with a drainage area of 57.4 km^2. Five meteorological stations are distributed within the basin, and the Liulin hydrological station serves as the outlet control station.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

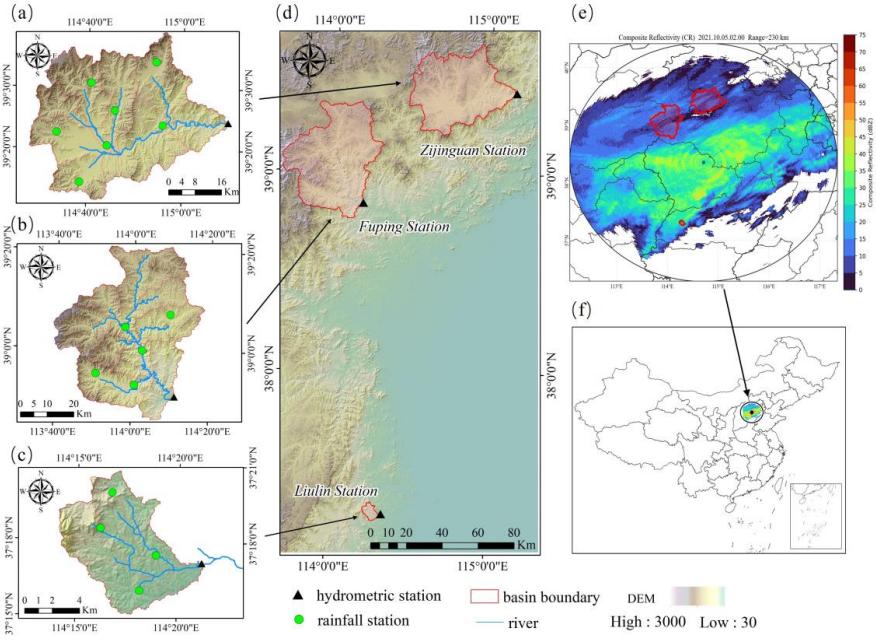


Figure 1. Overview of the study area. (a) Jumahe basin, (b) Shahe basin, (c) Liulin basin, (d) Location of the research area, (e) Radar scanning range, (f) The position of the radar.

The data were obtained from the Hebei Provincial Hydrology Bureau and the Hebei Provincial Meteorological Bureau. The rainfall-runoff records during the flood seasons from 2000 to 2023 and radar echo data during the flood seasons from 2018 to 2023 for the three study basins have been collected. The rainfall-runoff records were temporally interpolated to a 30 min resolution. Typical flood events were identified based on the criterion that each event should exhibit a distinct and complete rising and recession process. The radar-echo data were temporally matched with the flood events and used as optional meteorological inputs to the forecasting models. In total, 38 typical flood events were identified in the Shahe Basin, with 4250 h of available radar-echo maps, 25 typical flood events were identified in the Jumahe Basin, with 1512 h of radar-echo maps, and 45 typical flood events were identified in the Liulin Basin, with 498 h of radar-echo maps.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

2.2 Flood forecasting model

2.2.1 Model framework

The differentiable hybrid hydrological model suitable for short-term flood forecasting in small and medium-sized basins were developed. The model was constructed based on key hydrological processes and integrates runoff generation, baseflow recession, lateral inflow aggregation, channel routing, and multi-lead state correction into an end-to-end trainable framework. Different to purely data-driven models, the proposed framework first generated initial flow with physical meaning through a differentiable hydrological core, and then nonlinearly corrected the physical prior by combining historical state memory, future meteorological characteristics, and lead-time-specific information.

The model consisted of two main components. The first part is the differentiable hydrological core, which produced the physical prior discharge Q_{phys}. This core was guided by hydrological mechanisms, in which watershed topographic characteristics were represented in a differentiable form, runoff generation was described using physical formulations, and baseflow dynamics were represented through a differentiable recession process. Within this framework, channel routing was treated as a key learnable component. Three routing schemes, namely Muskingum routing, convolutional unit-hydrograph (CNN Unit) routing, and Physics-Informed Neural Architecture Search (PINAS) routing, were structurally substituted and compared to evaluate the influence of different routing formulations on flood hydrograph representation. The second part was the multi-lead forecasting head, which integrated the physical-prior discharge, historical hydrological memory, and future meteorological characteristics to generate the final forecast discharge. This design enables the model to retain the hydrological process constraints while adapting to the nonlinear

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

response differences between different basins and flood events through data-driven corrections.

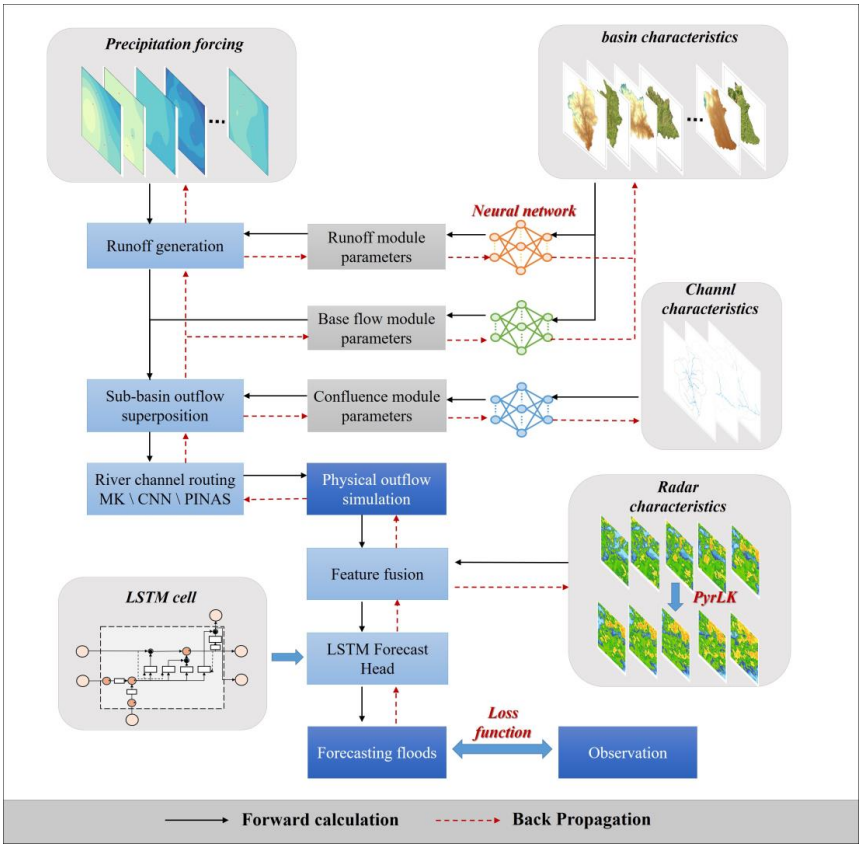


Figure 2. The diagram of the hybrid hydrological model

2.2.2 Differentiable hybrid hydrological core

The differentiable hybrid hydrological core was designed to simulate the main process of rainfall converting into downstream river flow. This module consisted of four components, runoff generation, baseflow recession, lateral inflow aggregation, and channel routing. All components were formulated in a differentiable manner, allowing the model to be trained end-to-end through backpropagation while maintaining hydrologically meaningful process constraints.

2.2.2.1 Runoff generation

The runoff generation adopted differentiable SCS-CN expression, which linked rainfall loss

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

with effective runoff. The curve number parameters were used to describe the combined influence of land surface conditions, previous wetness status, and rainfall-runoff threshold behavior. To make the model better adapted to different flood generation conditions, the runoff generation parameters were treated as learnable variables and restricted within physically reasonable ranges through bounded differentiable mappings. This prevents nonphysical parameter values and numerical instability during training process. In this module, rainfall inputs were first transformed into effective rainfall after accounting for hydrological losses, and the resulting effective rainfall was then converted into fast lateral inflow. The differentiable SCS-CN expression retained the threshold and nonlinear characteristics of the rainfall-runoff conversion, thereby improving the hydrological rationality of the model under limited sample conditions.

2.2.2.2 Baseflow calculation

The baseflow component was represented by a differentiable recession formulation to describe slow-response runoff. The baseflow state was updated recursively over time, and its recession rate was controlled by a recession coefficient. To ensure the stability of the numerical and hydrological values, the recession coefficient was limited within a physically reasonable range, so that the baseflow dynamics followed an exponential decay behavior. The baseflow component was then combined with the routed fast-flow component at the basin outlet, forming part of the physical-prior discharge. This design preserved the influence of antecedent hydrological conditions on current discharge in short-term flood forecasting, and was particularly important for maintaining recession flows and representing post-peak hydrograph behavior.

2.2.2.3 Lateral inflow aggregation and river-network topology

For spatial representation, the basins were divided into multiple subbasin units, and the

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

generated rapid runoff was allocated as lateral inflow according to the area fraction of each subbasin. The lateral inflow volume of each subbasin was further mapped to the corresponding river section as the local input for the calculation of river network paths. The river network was represented as a directed topological structure composed of reach nodes. The inflow volume of each river segment includes the local lateral inflow volume and the outflow volume of its upstream river segment. Flow propagation was then performed reach by reach from upstream to downstream according to a predefined topological order. This structure ensured physically consistent flow accumulation and propagation along the river network.

2.2.3 Channel routing schemes

Channel routing was a key component linking subbasin runoff generation to outlet discharge forecasting. To assess the impact of different routing structures on the performance of flood hydrographs, three differentiable channel routing schemes were embedded into a unified hydrological core: Muskingum routing, convolutional unit-hydrograph routing, and Physics-Informed Neural Architecture Search routing. The formulas for the channel routing were shown in Table 1.

2.2.3.1 Muskingum routing

The Muskingum method represented flood-wave propagation through a storage–discharge relationship. For each river reach r, two parameters K_r and X_r were respectively defined, where K_r denoted the characteristic travel-time scale and X_r controlled the relative weighting of inflow and outflow in channel storage. After standard finite-difference discretization, the routed outflow was expressed as a recursive update, in which the current outflow was represented as a linear combination of the current inflow, previous inflow, and previous outflow. To enable end-to-end

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

differentiable training while avoiding numerical oscillations, K_r and X_r were treated as learnable parameters but constrained within physically plausible ranges through differentiable mappings. In addition, the routing stability condition $K_r(1-X_r) \geqslant \Delta t/2$ was imposed.

2.2.3.2 Convolutional unit-hydrograph routing

In the convolutional unit-hydrograph routing scheme, channel routing was formulated as a linear time-invariant impulse-response process. Each reach r was assigned a discrete causal unit-hydrograph kernel $w_{r,j}$. Non-negativity and volume-conservation constraints are imposed on the kernel to ensure physical interpretability, preventing negative discharge generation and preserving runoff volume in the convolutional sense while allowing flood peak smoothing and propagation delay. The routed outflow $Q_r(t)$ was obtained by the discrete convolution between the reach inflow $I_r(t)$ and the unit-hydrograph kernel. Because both the convolution operation and the normalized kernel parameterization were differentiable, this routing operator was jointly trained with the upstream rainfall–runoff module through end-to-end backpropagation.

2.2.3.3 Physics-Informed Neural Architecture Search routing

To enhance the adaptability of the routing structure to different basin scales and flood wave propagation characteristics, a routing scheme based on physical information called Physical Informed Neural Architecture Search (PINAS) was proposed. This method combined neural architecture search (NAS) with physics-informed constrained to automatically identify appropriate combinations and scales of routing operators under hydrological constraints. The channel routing was represented as a routing stack composed of multiple differentiable operators, and the appropriate operator combinations and scale parameters were selected from the predefined search space to form a routing structure that was adaptable to different river network configurations.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

The search space included causal convolution, dilated convolution, multi-scale convolution, kernel mixing, gated connections, pure-delay mapping, and identity mapping. During structural computation, the input sequence was transformed layer by layer, while differentiable mixing between kernel output and residual gated connections was allowed. To maintain physical consistency, smoothness regularization was introduced. The regularization terms from different layers were aggregated into a routing regularization term, and a volume-consistency residual was calculated. During the training process, physical constraints were incorporated into the loss function in a weighted manner. Candidate architectures were first screened using limited training and validation batches, and the selected architectures were then evaluated more thoroughly. A complexity penalty was also included in the scoring function to avoid unnecessarily complicated routing structures. Compared with Muskingum and CNN-Unit, PINAS learned routing operator combinations that were adaptive to different reaches and flood events under physical constraints, thereby enhancing the representation of complex flood-wave propagation processes.

Table 1. The formula for the channel routing

Type	Formula	Description
Unit-hydrograph kernel	$w_{r,j}=\frac{softplus(w_{r,j})}{\sum_{i=0}^{k-1} softplus(w_{r,i})}$	Non-negative and mass-conserving parameterization of the convolutional unit kernel.
Unit-hydrograph routing	$Q_r(t)=\sum_{j=0}^{k-1} w_{r,j} I_r(t-j)$	Channel routing based on the discrete convolution.
Muskingum denominator	$D_r=K_r(1-X_r)+\frac{\Delta t}{2}$	Denominator used to calculate the Muskingum routing coefficients.
Muskingum routing coefficients	$C_{0,r}=\frac{-K_r X_r+\Delta t/2}{D_r}$ $C_{1,r}=\frac{K_r X_r+\Delta t/2}{D_r}$ $C_{2,r}=\frac{K_r(1-X_r)-\Delta t/2}{D_r}$	Stable recurrence coefficients of the differentiable Muskingum routing scheme.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Muskingum routing recurrence	$Q_r(t+\Delta t) = C_{0,r} I_r(t+\Delta t) + C_{1,r} I_r(t) + C_{2,r} Q_r(t)$	Recursive update of routed outflow in the Muskingum scheme.
PINAS routing stack	$Q_r(t) = \left(f^{(B)} \circ f^{(B-1)} \circ \cdots \circ f^{(1)} \right)\left(I_r(t)\right)$	Differentiable routing stack composed of multiple operator layers.
Operator selection	$f^{(b)}(x) = \sum_{o \in \mathcal{O}} z_{b,o}\, o(x), \quad z_{b,o} \in \{0,1\}, \quad \sum_{o \in \mathcal{O}} z_{b,o} = 1$	Differentiable relaxation of operator selection
Routing regularization	$R_{\text{route}} = \frac{1}{N_r} \sum_{r=1}^{N_r} \sum_{b=1}^{B} \text{reg}\left(\theta_{r,b}\right)$	Physics-inspired regularization imposed on routing parameters.

Note: r indexes river reaches and s indexes sub-basins. t denotes the discrete time index, and Δt is the hydrological model time step. $Ir(t)$, $Lr(t)$, and $Qr(t)$ denote reach inflow, lateral inflow, and reach outflow, respectively. $U(r)$ is the set of immediate upstream reaches of reach r. Sub-basin runoff $Qs(t)$ is mapped to the channel network through the assignment function m(s). In the convolutional unit-hydrograph routing scheme, $w_{\{r,j\}}$ denotes the normalized kernel weight. In Muskingum routing, Kr is the propagation time-scale parameter, Xr is the weighting coefficient, and $C_{0,r}$, $C_{1,r}$, and $C_{2,r}$ are the recurrence coefficients. In PINAS, B is the number of differentiable routing layers, $f^{(b)}(\cdot)$ denotes the operator layer at depth b, $\theta_{r,b}$ denotes the routing parameters of reach r at layer b, and R_{route} is the routing regularization term.

2.2.4 Multi-lead forecasting head and state-correction mechanism

The differentiable hydrological core generated a physically meaningful prior discharge, denoted as Q_{phys}. However, in short-term flood forecasting, uncertainties associated with future rainfall evolution, initial hydrological states, and model structural deficiencies may lead to deviations between the physical-prior discharge and the actual discharge. Therefore, a multi-lead forecasting head was constructed after the hydrological core to dynamically correct the physical-prior discharge through a state-correction mechanism, thereby producing the final multi-lead flood forecasts.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Let t denote the current forecasting time, and $t+k$ denote the target time corresponding to the k forecast lead, where $k=1,2,\ldots,K$. In this study, the rainfall data were represented at a 30 min temporal resolution, and $K=4$, corresponding to lead times of 0.5, 1.0, 1.5, and 2.0 h. For each lead time, the hydrological core first generated the corresponding physical-prior discharge Q_{t+k}^{phys}. To perform state-dependent correction of this physical prior, the model further constructed an integrated context vector C_t, which consisted of three types of information: the historical memory state h_t^{mem}, the encoded future meteorological forcing f_t^{fut}, and the hydrological state features s_t^{hyd}. Specifically, h_t^{mem} was encoded from historical hydrometeorological sequences and historical discharge records, and was used to represent antecedent wetness conditions, rainfall–runoff response memory, and hydrograph persistence. The term f_t^{fut} was derived from future meteorological forcing information, mainly including extrapolated radar echoes and their basin-scale statistical descriptors, and was used to describe the intensity, spatial position, and movement tendency of short-term future rainfall systems. The vector s_t^{hyd} represents hydrological state information, including the current discharge state, discharge variation rate, antecedent precipitation index, rainfall statistics, and physical-prior discharge.

Due to rainfall uncertainty, watershed response lag, and routing phase errors differ among forecast lead times, a lead-aware mechanism was introduced. The forecasting head formulations was shown in Table 2. For the k lead time, the model integrated the context vector C_t, the lead-time embedding e_k, the future meteorological forcing feature x_{t+k}^{fut}, and the physical-prior discharge Q_{t+k}^{phys} to construct the lead aware feature representation $z_{t,k}$. Here, e_k was a learnable lead-time embedding that represented lead-dependent error growth, information decay, and response delay. The variable x_{t+k}^{fut} denoted the future meteorological input at the k lead time, mainly derived from

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

radar-echo extrapolation and its statistical features. The lead-aware feature $z_{t,k}$ was then used as the direct input for residual correction. After obtaining $z_{t,k}$, the model estimated the discharge correction term ΔQ_{t+k} through a nonlinear residual correction function $f_\theta(\,)$. The residual term was designed to compensate for structural biases in the hydrological core arising from localized intense rainfall, multi-peak flood processes, radar extrapolation errors, and complex nonlinear routing dynamics. Finally, the residual correction ΔQ_{t+k} was added to the physical-prior discharge Q_{t+k}^{phys}, and a non-negative mapping function $g^+(\cdot)$ was applied to obtain the final forecast discharge Q_{t+k}, ensuring that the output satisfies the basic physical constraint of non-negative discharge. Accordingly, the multi-lead forecasting head could be interpreted as a physical-prior-driven state-correction mechanism. The core idea is to learn a residual correction ΔQ_{t+k} conditioned on the physical prior discharge Q_{t+k}^{phys}, historical memory, future meteorological information, and lead-time-specific characteristics. This structure preserved the fundamental physical constraints associated with runoff generation, flow concentration, and channel routing, while enhancing the adaptability to hydrograph shape, peak magnitude, and recession behavior across different forecast lead times.

Table 2. Forecasting-head formulations

Type	Formula	Description
Integrated context vector	$C_t=[h_t^{mem};\, f_t^{fut};\, s_t^{hyd}]$	Integrated context vector combining historical memory, future forcing, and hydrological state information.
Lead-aware feature	$z_{t,k}=\phi(C_t, e_k, x_{t+k}^{fut}, Q_{t+k}^{phys})$	Lead-aware feature representation for the k-th forecast horizon.
Residual correction	$\Delta Q_{t+k}=f_\theta(z_{t,k})$	Data-driven correction term for the physical-prior discharge.
Final forecast	$Q_{t+k}=g^+(Q_{t+k}^{phys}+\Delta Q_{t+k})$	Final discharge forecast constrained by a non-negative mapping.

Note: t denotes the current forecasting time, and k is the forecast lead step. C_t is the integrated context vector, where $[\,;\,]$ denotes vector concatenation. h_t^{mem} is the historical memory state,

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

329 f_t^{fut} is the encoded future meteorological forcing, and s_t^{hyd} is the hydrological state feature
330 vector. e_k is the learnable lead-time embedding, $\varphi(\cdot)$ is the nonlinear feature-fusion function, $f_\theta(\cdot)$ is
331 the residual correction function with parameters θ, and $g^+(\cdot)$ is a non-negative mapping function
332 used to ensure physically feasible discharge predictions.

333 **2.3 Radar echo extrapolation**

334 The future meteorological information was obtained through short-term extrapolation of the
335 radar echo sequence. The Pyramidal Lucas–Kanade (PyrLK) optical-flow method was used to
336 estimate a pixel-wise two-dimensional motion field $v(x)=[u(x),v(x)]$. The formulations were shown
337 in Table 3. Under a Lagrangian persistence assumption, the most recent radar echo was advected
338 forward to produce multi-step extrapolated fields. To accommodate large storm displacements,
339 PyrLK adopted a coarse-to-fine image pyramid, where large-scale motion was first estimated at
340 coarser levels and then iteratively refined at finer levels. In practice, outliers were rejected by a
341 combination of forward–backward consistency checking and displacement-magnitude thresholding,
342 followed by spatial interpolation and mild smoothing to obtain a continuous motion field.
343 Extrapolation was implemented through backward warping with bilinear interpolation for non-
344 integer source locations. The procedure was applied directly to dBZ echoes, and the extrapolated
345 echoes were used as future forcing inputs for the multi-lead-time flood forecasting model.

346 Table 3. Lucas-Kanade optical-flow formulations

Type	Formula	Description
Brightness conservation	$R(\mathbf{x},t) \approx R(\mathbf{x}+\mathbf{v}(\mathbf{x})\delta, t+\delta), \quad \delta = 6\text{min}$	Approximate conservation of radar-echo intensity between two consecutive frames.
Optical-flow constraint	$R_x u + R_y v + R_t = 0$	Linearized optical-flow constraint equation.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Local least-squares estimation	$\min_{\Delta\mathbf{d}} \sum_{\mathbf{p}\in\mathcal{W}} \left(\nabla R(\mathbf{p},t)\cdot\Delta\mathbf{d} + R_t(\mathbf{p},t)\right)^2$	Estimation of local displacement increments in the Lucas-Kanade window.
Pyramid displacement update	$\mathbf{d}^{(\ell)} = 2\mathbf{d}^{(\ell+1)} + \Delta\mathbf{d}^{(\ell)}$	Coarse-to-fine displacement refinement in the image pyramid.
Frozen-advection extrapolation	$\hat{R}(\mathbf{x},t+k\delta) = R\left(\mathbf{x}-k\mathbf{v}(\mathbf{x})\delta, t\right), \quad k=1,2,\ldots$	Backward-mapping extrapolation under a frozen advection field.

Note: $R(x,t)$ denote radar reflectivity in dBZ at location $x=(x,y)$ and time t, $v(x)=[u(x),v(x)]$ is the two dimensional motion field, R_x, R_y, R_t denote spatial and temporal gradients, W denote the local Lucas–Kanade window used for least-squares estimation, ℓ denote pyramid levels in the coarse-to-fine scheme, $\hat{R}(\mathbf{x},t+k\delta)$ is the extrapolated reflectivity, where k is the lead step.

2.4 Control experiment

To evaluate the effectiveness of the proposed hybrid hydrological model, the Long Short-Term Memory model (LSTM) (Kratzert et al., 2018) was implemented as a control experiment. This model directly learned the nonlinear input-output relationship from historical information sequences and future meteorological forcing. Therefore, its forecasting results were used to represent the baseline performance achievable by a black-box deep-learning model under the same input conditions, providing a reference for assessing the added value of physical constraints in the hybrid hydrological framework.

2.5 Loss function

A hybrid loss function was constructed by combining data-fitting terms with physics-consistency constraints to improve multi-lead flood hydrograph forecasting, peak-flow forecasting accuracy, and physical consistency. To enhance model sensitivity to flood peaks, a weighted error term was introduced in the logarithmic discharge domain. Physical consistency was enforced through a physics fitting loss, a residual consistency loss, and a relative volume error constraint

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

during end-to-end training. In addition, the regularization residuals produced by the PINAS routing module were incorporated into the objective function with corresponding weights, thereby constraining the learned routing structure while maintaining training stability. The Loss-function formulations were shown in Table 4.

Table 4. Loss-function formulations

Type	Formula	Description
Total objective	$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\text{vol}}\mathcal{L}_{\text{vol}} + s(e)\lambda_{\text{route}}R_{\text{route}}$	The training objective combines a data-fitting term with physically motivated constraints.
Data-fitting loss	$\mathcal{L}_{\text{data}} = \frac{1}{BK}\sum_{b=1}^{B}\sum_{k=1}^{K}\sum_{\tau\in Q} W_{bk}\,\rho_{\tau}\left(y_{bk} - \hat{q}_{bk}^{(\tau)}\right)$	Quantile-based data-fitting loss.
Volume-consistency loss	$\mathcal{L}_{\text{vol}} = \frac{1}{B}\sum_{b=1}^{B}\frac{\left\lvert\sum_k \hat{q}_{bk}^{+} - \sum_k y_{bk}^{+}\right\rvert}{\left\lvert\sum_k y_{bk}^{+}\right\rvert + \varepsilon}, \quad x^{+} = \max(x,0)$	Constrains the cumulative forecast discharge volume to match the measured volume.
Ramp-up schedule	$s(e) = \min\left(1, \frac{e}{e_0}\right)$	Epoch-dependent ramp-up factor used to gradually introduce physics-based constraints.

Note: B is the batch size, K is the lead times, y_{bk} is the observation, W_{bk} denotes a composite weighting matrix, λ_{vol} and λ_{route} are hyperparameters controlling the strength of the volume-consistency and routing-regularization terms, $s(e)$ is a ramp-up schedule used to gradually introduce physics constraints during training.

2.6 Shapley Additive Explanation

The Shapley Additive Explanation (SHAP) was used to interpret the forecasting results. SHAP is based on the concept of Shapley values from cooperative game theory, where a single prediction is decomposed into the marginal contributions of each input feature to the model output. Therefore, it provides local explanations at the sample level and can be aggregated over multiple flood events

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

to quantify the importance of global features. Given the temporal input structure, including gauge rainfall, areal rainfall, antecedent precipitation index, historical radar echoes, and future radar-echo features, individual feature values of different time steps were used as basic explanatory units to calculate the Shapley values. To enhance hydrological interpretability, grouped SHAP analysis was further conducted by aggregating Shapley values according to variable category and temporal scale. Specifically, the contributions of the same hydrological variable over the full historical window were summed to obtain variable-level importance. This enabled diagnosis of the relative contribution of different hydrometeorological factors to flood forecasting.

3 Experiment Design

3.1 Hydrometeorological fusion for flood forecasting

A land–atmosphere coupled forcing configuration was established to integrate hydrological information with future meteorological inputs. Three channel-routing modules, namely CNN-Unit, Muskingum, and PINAS, were comparatively evaluated. In addition, a purely deep-learning LSTM model was implemented as the benchmark for flood hydrograph forecasting. The comparison focused on the ability of different configurations to reproduce flood peak magnitude, runoff volume, and the rising and recession limbs of the hydrograph, thereby quantifying the contribution of different routing mechanisms to short-term flood forecasting.

3.2 Data dependency analysis for flood forecasting

To assess the dependence of model performance on data availability and its applicability under limited data conditions, training sample reduction experiments were conducted. By gradually reducing the proportion of training data from 25% to 100%, this study repeatedly trained and

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

evaluated the model on the same test set. This experiment was designed to characterize how forecasting performance changes with training sample size, thereby providing guidance for data requirements and model transfer in operational applications.

3.3 Interpretability analysis of forecasting results

In order to determine the factors that have the greatest impact on the accuracy of flood forecasting, an interpretability analysis based on Shapley values to attribute the causes of the forecast results were employed. The PINAS model was used for interpreting representative flood events, and the direction and degree of each contribution to the forecasting were quantified. To accommodate the temporal structure of the input data, feature contributions were further aggregated along the time dimension, yielding both variable-level importance and temporal contribution patterns. These results were used to diagnose the key influencing factors of flood forecasts at different lead times.

3.4 Model Training and Evaluation Metrics

All models were implemented in PyTorch and trained on the NVIDIA GeForce RTX 4060 Ti GPU. The AdamW optimizer was used with a batch size of 32 and a maximum training epoch of 200. Cosine annealing learning rate scheduling was employed, with an initial learning rate of 0.001. The optimal model weights were selected through an early stopping strategy based on validation-set performance. The accuracy of flood forecasting was evaluated using four deterministic metrics. The relative peak error (REP) was used to quantify the bias in flood-peak magnitude, while the relative runoff volume error (REV) was used to measure the bias in total flood volume. Root-mean-square error (RMSE) and Nash–Sutcliffe efficiency (NSE) were further employed to comprehensively assess hydrograph fitting performance. The evaluation metrics were shown in

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Table 5.

Table 5. Evaluation metrics

Type	Formula	Description
NSE	$\mathrm{NSE}=1-\frac{\sum_{t=1}^{N}\left(\hat{Q}(t)-Q(t)\right)^2}{\sum_{t=1}^{N}\left(Q(t)-\bar{Q}\right)^2}$	Hydrograph skill relative to the observed.
REP	$REP=\frac{\hat{Q}_p-Q_p}{Q_p}\times 100\%$	Relative error of peak discharge.
REV	$REV=\frac{\hat{V}-V}{V}\times 100\%$	Relative error of runoff volume.
RMSE	$RMSE=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\hat{Q}(t)-Q(t)\right)^2}$	Root mean square error of the forecast hydrograph.

Note: $Q(t)$ is the measured discharge hydrograph，$\hat{Q}(t)$ is the forecasted discharge hydrograph，Q_p and $\hat{Q}_p$ denote the measured and forecasted peak discharges，V and $\hat{V}$ denote the measured and forecasted runoff volume，T_p and $\hat{T}_p$ denote the measured and forecasted peak time，$\bar{Q}$ represent the average value of the measured discharge.

4 Results

4.1 Evaluation of Flood Forecasting Accuracy

As shown in Figure 3, the forecasting accuracy of all models deteriorated to varying degrees as the lead time increased. NSE generally decreased and RMSE increased for most events, indicating that uncertainties associated with future rainfall information, watershed state evolution, and routing related errors accumulated progressively. The LSTM model performed well in several high-peak flood events, particularly in controlling peak-discharge and runoff-volume errors, which suggested that LSTM was able to learn complex nonlinear rainfall–runoff relationships from historical samples. However, because it lacked explicit hydrological process constraints, LSTM was

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

still prone to phase displacement and local hydrograph distortion, especially for multi-peak floods in small basins and under longer lead times. Among the three hybrid hydrological models, PINAS exhibited the most stable overall performance. Its NSE was generally higher than those of CNN-Unit and Muskingum across most events and lead times, which indicated that the physically constrained searchable routing structure was more flexible in adapting to different basin scales and flood propagation characteristics. By contrast, CNN-Unit represented channel response through unit-hydrograph convolution and was able to capture smoothing and lag effects to some extent, but it was susceptible to peak amplification, spurious peaks, or phase shifts under strongly nonstationary flood conditions. The Muskingum model had clear physical interpretability. However, its linear storage–discharge assumption limited its ability to represent complex nonlinear flood-routing processes, particularly during high-peak or rapidly varying flood events, where peak attenuation and runoff-volume underestimation were more likely to occur.

Model performance also varied substantially among basins and flood types. As shown in Figure 4, the 20230729 event in the Shahe Basin exhibited a well defined dominant peak, and all models reproduced the hydrograph reasonably well, which indicated that both the LSTM and the hybrid hydrological models could achieve high accuracy when the rainfall–runoff response was relatively regular. For the 20200802 event in the Shahe Basin, PINAS preserved the hydrograph shape more effectively than CNN-Unit and Muskingum, suggesting that the searchable routing structure helped reduce the accumulation of phase errors. The 20230729 event in the Jumahe Basin was characterized by a high peak and large discharge. LSTM provided the best fit for peak magnitude and runoff volume, while PINAS performed best among the hybrid models. In contrast, Muskingum substantially underestimated the hydrograph, reflecting the limited applicability of traditional linear

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

routing under strongly nonlinear flood propagation conditions. The two flood events in the Liulin Basin were characterized by rapid small basin response and multi-peak hydrographs, which considerably increased forecasting difficulty for all models. PINAS showed relatively stable performance for the 20230730 event, indicating its advantage over fixed form physically based routing schemes in representing multi-peak responses. Hydrograph comparisons further revealed distinct sources of model error. LSTM showed strong capability in fitting peak magnitude, but its hydrograph evolution was mainly sample-driven and lacked physical constraints on peak timing and recession behavior. CNN-Unit could simulate propagation delay and hydrograph smoothing to some extent, but it was sensitive to nonstationary inputs. Muskingum provided clear physical interpretation but tended to over-smooth complex flood hydrographs. PINAS achieved a better balance among peak propagation, hydrograph smoothing, and recession preservation by searching routing-operator combinations under physical constraints, making it the best-performing scheme among the three hybrid hydrological models.

In summary, PINAS showed stronger cross event robustness and long lead adaptability than CNN-Unit and Muskingum, which indicated that incorporating a physically constrained searchable routing structure into a differentiable hydrological framework improved the ability to represent complex watershed response processes in flood forecasting.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

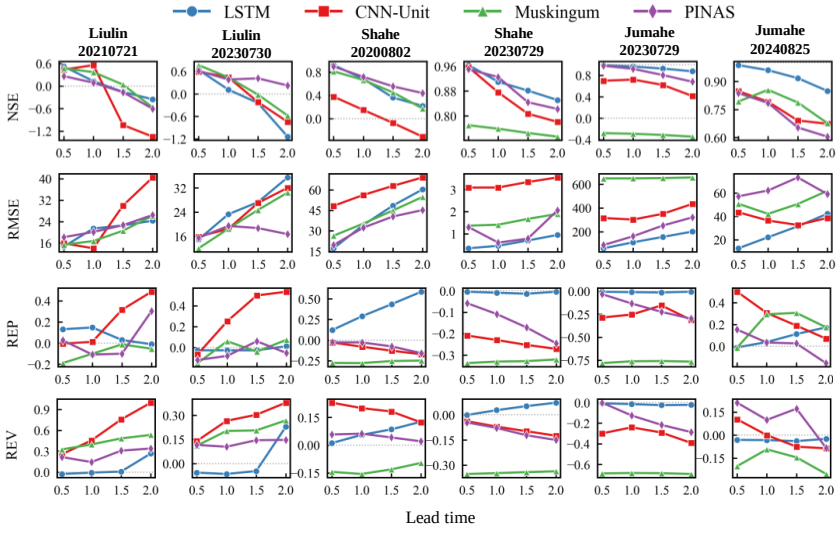


Figure 3. Comparison of accuracy evaluation metrics across models

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

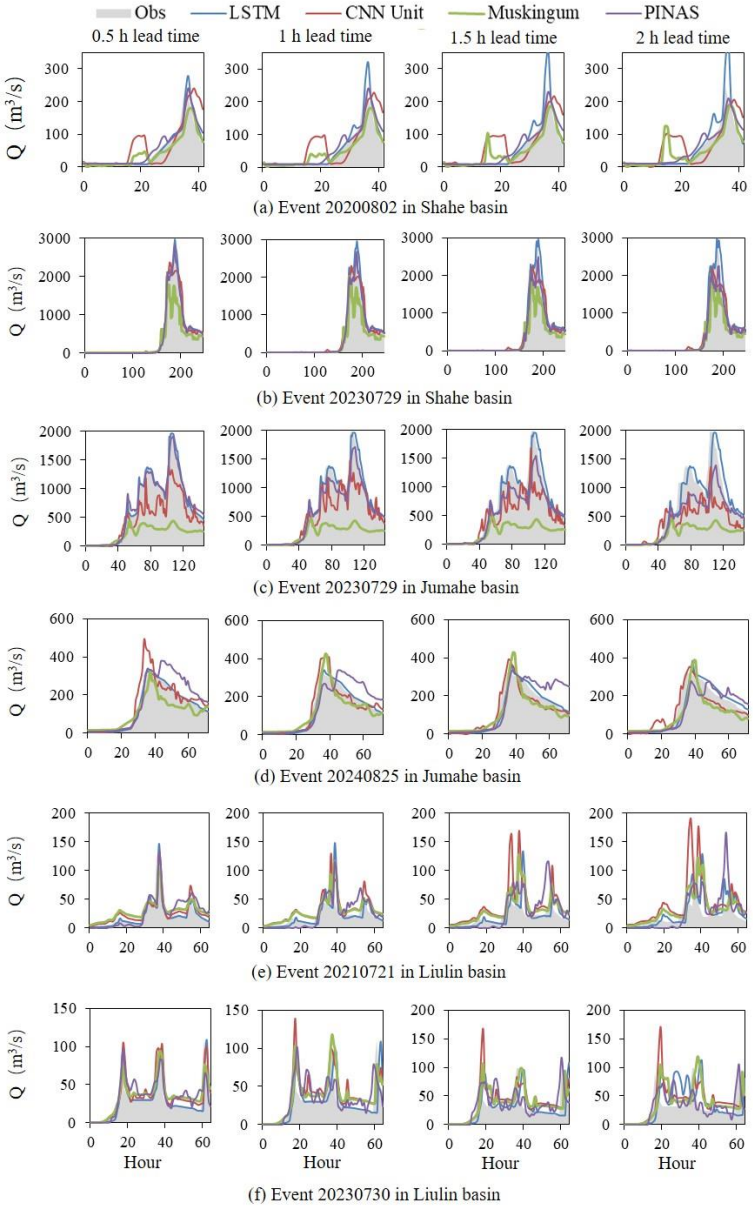


Figure 4. Forecast of Flood Process

4.2 Data dependency analysis

To evaluate the applicability and sample efficiency of the model under conditions of limited data, the PINAS model was trained using 25%, 50%, 75% and 100% of the data to quantify the

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

impact of training sample size on flood forecasting skill. As shown in Figure 5, increasing the training data volume significantly improved the forecasting stability of PINAS, although model performance did not strictly show a monotonic improvement trend across all events and indicators. When using the complete training dataset, the model typically achieved higher NSE and lower RMSE in most flood events, which indicated that sufficient training samples enable the model to more comprehensively learn the relationship between basin runoff generation and path response, thereby improving the overall fitting effect of the hydrological map. In contrast, the results obtained using 25% and 50% of the training data showed greater variability, particularly at longer lead times, where NSE declined more rapidly and RMSE increased more significantly, which suggested that insufficient samples amplified the effects of future meteorological input errors, state evolution uncertainty, and routing-related phase errors on forecast performance.

The benefits of increasing training data were mainly reflected in three aspects. First, the ability to identify the dominant flood peak was improved. As shown in Figure 6, for flood events with a pronounced main peak, such as the 20230729 events in the Shahe and Jumahe basins, the hydrographs obtained using 100% of the training data were closer to the measured flood processes, with better preservation of peak magnitude, peak timing, and post-peak recession behavior. Second, systematic bias was reduced. As the training data increased, REP and REV generally became smaller, indicating that more sample coverage helped reduce overestimation or underestimation of flood-peak discharge and total runoff volume. Third, long lead times became more stable. Low-data configurations showed more evident error growth at the 1.5 and 2.0 h lead times, whereas training with the full dataset partially suppressed the accumulation of errors with increasing lead times. Nevertheless, in some events, the 75% or 50% data configurations achieved results that were

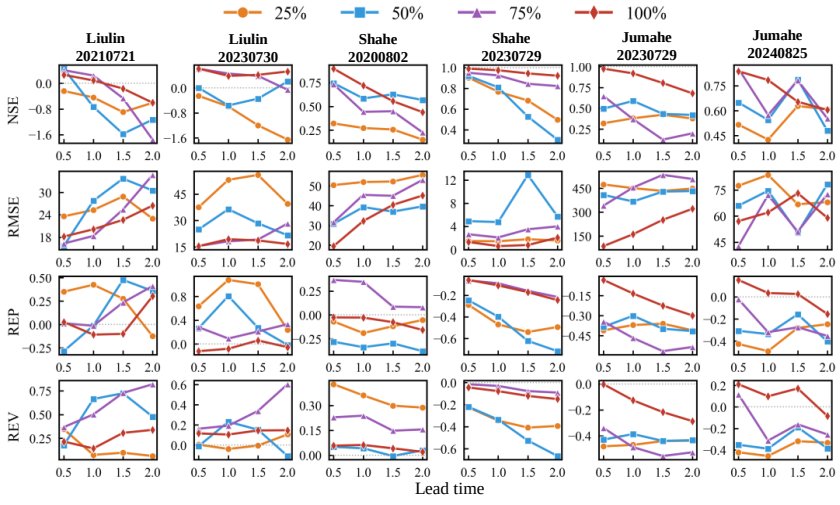
https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

comparable to those obtained using 100% of the training data at specific lead times. This behavior indicated that model performance was governed not only by the number of training samples, but also by sample representativeness, flood type coverage, event composition in the training set, and uncertainty in parameter optimization. For flood forecasting with limited event samples, simply increasing the sample size may not directly translate into improved accuracy if the added samples differ substantially from the hydrological response characteristics of the testing events. The sensitivity to training data also varied among basins. The typical large-flood events in the Shahe and Jumahe basins showed a relatively clear positive response to data expansion, especially for high-peak events with a well defined dominant peak, where full-data training substantially improved peak and volume estimation. By contrast, the Liulin Basin was characterized by rapid hydrological response, multi-peak hydrographs, and localized intense rainfall forcing, making model performance more sensitive to the composition of the training data. Even under relatively large training data fractions, some long-lead forecasts still exhibited considerable variability, suggesting that forecasting errors in small basins were controlled more by peak-timing displacement and multi-peak structure mismatch than by sample size alone.

Overall, PINAS demonstrated a certain degree of sample efficiency. In some events, acceptable forecasting performance was achieved using 50%–75% of the training data, indicating that the physical constraints and searchable routing structure partly alleviated uncertainty caused by data scarcity. However, from the perspective of cross basin and cross event stability, a larger training dataset remained essential for achieving high accuracy. These results indicated that the performance of PINAS depended not only on the model architecture, but also on whether the training samples adequately represented typical flood types, extreme flood processes, and flash response

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

527 characteristics in small basins.



528

529 Figure 5. Forecasting accuracy indicators of the model trained with 25% - 100% of the data volume

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

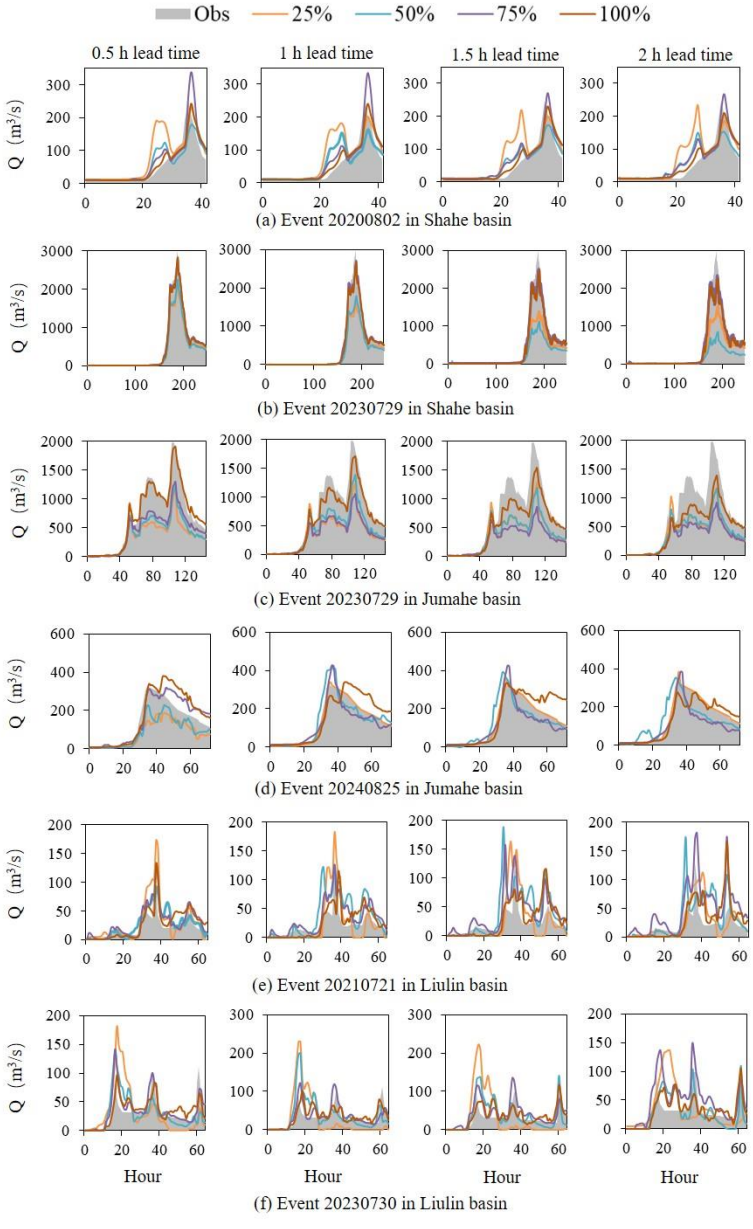


Figure 6. Training model to forecast flood process with 25% - 100% of data volume

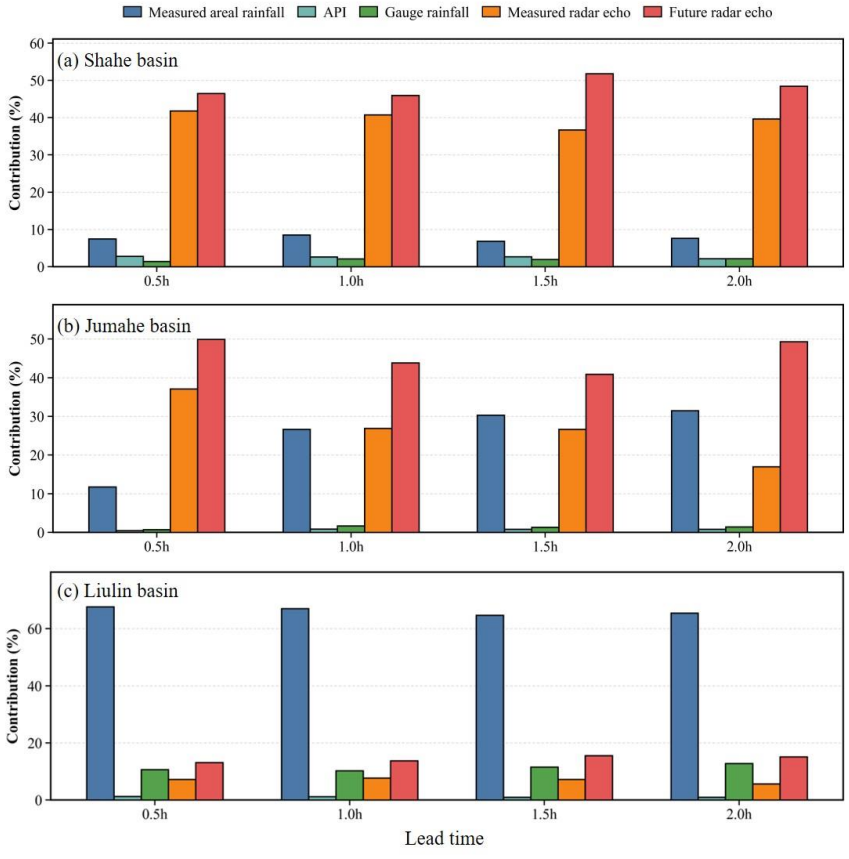
4.3 Interpretability analysis

To further interpret how the PINAS model utilized hydrometeorological information for flood forecasting, SHAP analysis was conducted to quantify the contributions of different input factors.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

SHAP values were aggregated by variable category and temporal dimension to identify both global and local contributions of different input information to the model outputs. As shown in Figure 7, the grouped SHAP results revealed clear basin differences in the dominant information sources. Forecasts in the Shahe and Jumahe basins were mainly controlled by radar echo information, with future radar echoes contributing the most and observed radar echoes ranking second. In the Shahe Basin, future radar echoes accounted for approximately 50% of the total contribution across different lead times, while observed radar echoes contributed approximately 40%, which indicated that, for the Shahe Basin, the rainfall spatial structure and short-term evolution represented by radar echoes provided critical information for inferring future flood dynamics. In the Jumahe Basin, as the lead time increased, the contribution of areal rainfall gradually increased, whereas the contribution of radar echo gradually decreases, which suggested that, in the larger basin, flood response was controlled by antecedent accumulated rainfall and the future spatial evolution of rainfall systems. In contrast, the dominant factor in the Liulin Basin was measured areal rainfall, whose contribution remained above 60%, while radar echo variables contributed relatively less. This pattern was consistent with the short concentration time and rapid flood response of the Liulin Basin, indicating that short-term forecasting in this basin relied more strongly on measured rainfall as a direct indicator of watershed saturation and rapid runoff-generation conditions.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026



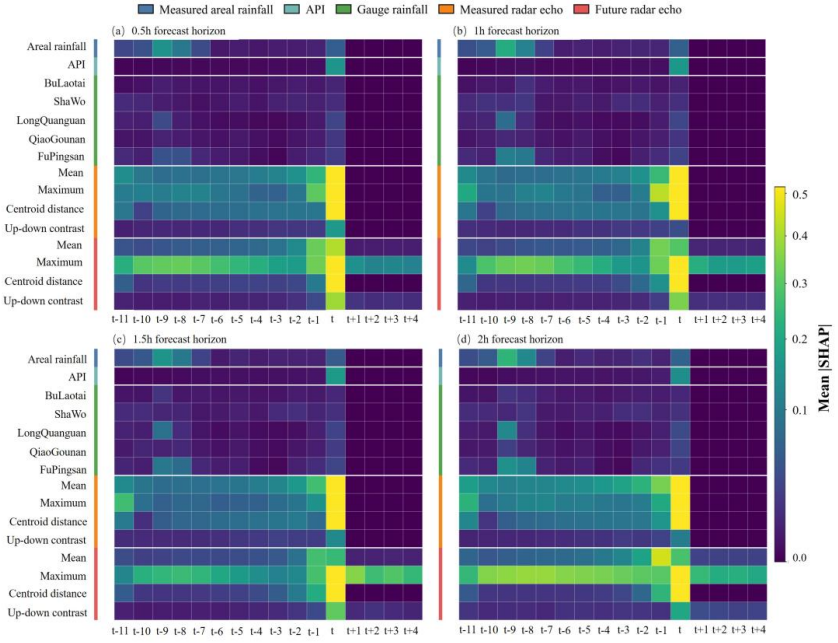
552

553 Figure 7 SHAP Importance Grouping

554 The SHAP heatmaps further revealed the temporal roles of different information sources during
555 the forecasting process. As shown in Figure 8, for the Shahe Basin, the maximum value features of
556 both measured and future radar echoes made strong contributions at the current time step and the
557 adjacent future periods, which indicated that the model primarily used radar echo intensity extremes
558 to identify intense rainfall cells that could potentially trigger flood peaks. Meanwhile, features such
559 as mean radar echo intensity and centroid distance also contributed to the forecasting, suggesting
560 that the model considered not only rainfall intensity, but also the spatial displacement of rainfall
561 systems relative to the basin when inferring flood wave propagation. As shown in Figure 9, for the

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

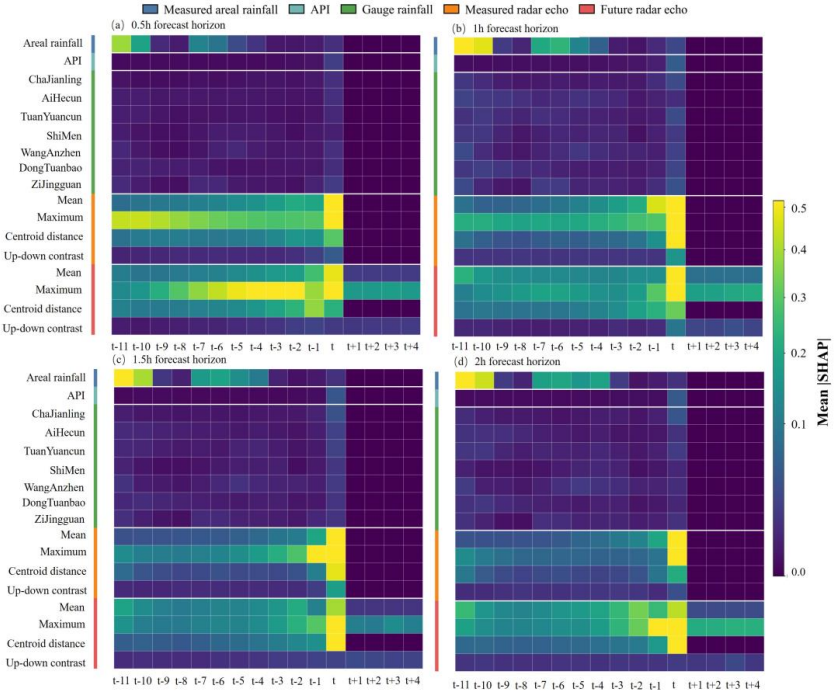
562 Jumahe Basin, areal rainfall showed relatively high contributions during earlier historical periods,
563 whereas radar echo features contributed more strongly at the current time and over future periods,
564 which indicated that flood forecasting in this basin depended simultaneously on antecedent rainfall
565 accumulation and short-term information on rainfall movement. As shown in Fig. 10, for the Liulin
566 Basin, areal rainfall maintained a substantial contribution throughout the entire historical window,
567 while gauge rainfall also showed relatively high importance at certain time periods and stations. In
568 contrast, the contribution of radar-echo information was concentrated mainly around the current
569 time and adjacent periods, which suggested that flood forecasting in small basins was more strongly
570 controlled by rainfall accumulation and localized gauge rainfall, whereas the additional value of
571 radar extrapolation information was relatively limited.



572

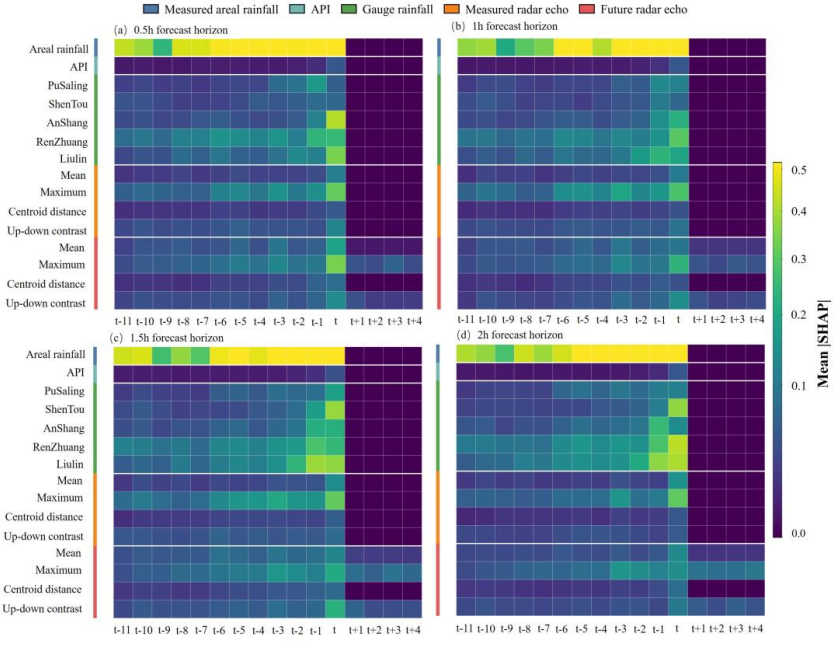
573 Figure 8. SHAP heatmap of feature contributions for the Shahe Basin model

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026



574

575 Figure 9. SHAP heatmap of feature contributions for the Jumahe Basin model



576

577 Figure 10. SHAP heatmap of feature contributions the Liulin Basin model

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Overall, the SHAP results indicated that the forecasting produced by the hybrid hydrological model based on PINAS were not dependent on a single dominant input variable. Instead, the model adaptively assigned importance to different information sources based on the scale of the basin and the characteristics of the flood event. For medium-sized basins such as the Shahe and Jumahe basins, radar echoes and the short-term extrapolated features played a central role in flood forecasting, particularly in characterizing the movement of future rainfall systems and the subsequent evolution of flood peaks. In contrast, for the Liulin Basin, measured areal rainfall and gauge rainfall remained the primary information sources, suggesting that in rapidly responding watersheds, measured areal rainfall and local rainfall observations were more crucial for short-term forecasting. These findings provided interpretability evidence for the necessity of hydrometeorological data fusion and demonstrated that the hybrid hydrological model can learn the contribution patterns of hydrological significance features within the framework of physical constraints.

5 Discussion

The differentiable hybrid hydrological forecasting framework coupled with future meteorological information was developed and its short-term flood forecasting capability was evaluated across representative basins and flood events. The results indicated that physically constrained routing, particularly PINAS, can improve the stability of hydrograph forecasting. Nevertheless, several limitations need to be further discussed.

The uncertainty of future meteorological forcing remained the main limiting factor for flood forecasting. In this study, radar echoes were extrapolated to provide rainfall information within the next 0-2 hours. However, the optical flow field calculation relied on the continuity of echo

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

movement and cannot fully reflect the rapid deformation of convection. Previous studies have shown that radar based nowcasting was generally effective at short lead times, but forecast errors increased rapidly for intense convective rainfall, and traditional advection based approaches struggle to represent nonlinear precipitation evolution (Ravuri et al., 2021). Therefore, the decline in accuracy at longer lead times may not only be attributed to the errors of the hydrological model, but also to the propagation of radar echo calculation errors to runoff forecasting. Future work should adopt ensemble radar extrapolation methods (Chung et al., 2025), deep generative nowcasting (Franch et al., 2025), or multi-source numerical weather prediction (Wang et al.,2024) to replace deterministic future forcing with probabilistic precipitation inputs, thereby explicitly propagating meteorological uncertainty through the hydrological forecasting chain.

The hybrid hydrological model mainly covered aspects such as runoff generation, baseflow recession, basin confluence, and channel routing, but does not explicitly account for spatially heterogeneous soil moisture, dynamic land-surface conditions, reservoir and gate operations, and urbanization effects. These processes can significantly alter the rainfall-runoff relationship in small and medium-sized watersheds, especially during extreme rainstorms or under different antecedent wetness conditions. Recent studies have suggested that differentiable hydrological modelling provides a promising approach for connecting process-based models with deep learning, because it enables end-to-end training while retaining physically meaningful states and parameters (Tsai et al., 2021). Therefore, future extensions should introduce learnable soil-water states, dynamic infiltration parameters, and hillslope concentration parameters, shifting the framework from physically constrained routing toward a distributed forecasting model with interpretable hydrological states(Ettalbi et al., 2025).

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Data dependency analysis indicated that model performance was more influenced by the representativeness of the samples and the coverage of flood types. Similar findings have been observed in studies on differentiable parameter learning and physics-based deep learning, where a larger training dataset can enhance generalization ability and physical consistency, provided that the samples cover a sufficiently diverse set of hydro-meteorological conditions (Feng et al., 2024; Huynh et al., 2025). This study collected flood events from three different scales and response characteristic basins, but the total number of events was still limited, especially for extreme floods and multi-peak hydrographs. Future work should expand the number of basins and climatic geomorphological settings, while also using physically consistent data augmentation or extreme-event resampling to improve model learning for rare high-peak and flash flood events.

The assessment in this study mainly relies on deterministic indicators, which are useful for comparing hydrological fit, peak flow error, and flood volume error, but they are insufficient to comprehensively describe the risk uncertainty in operational flood forecasting. Operating short-term flood forecasting not only requires the hydrological map, but also requires confidence intervals, the probability of exceeding the warning threshold, and the estimation of false alarm and missed alarm risks. Recent studies have shown that quantile regression, mean variance estimation, extreme quantile regression, and generative probability models can provide useful uncertainty information for flood or runoff forecasting (Pasche and Engelke, 2024; Jahangir and Quilty, 2024). Therefore, future work should expand the current framework by incorporating methods such as quantile regression, deep ensembles, or conditional generative models, moving towards probabilistic prediction, so that the model can provide deterministic hydrological maps and reliable flood risk intervals at different lead times. (Xiang et al., 2025; Ou et al., 2025)

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

In general, this study demonstrated that integrating future meteorological information with physically informed channel routing can improve short-term flood forecasting for small and medium-sized basins. However, the model performance was still affected by the uncertainty of future rainfall, the representativeness of the samples, and the simplified process representation. Future research should focus on probabilistic meteorological forcing, distributed physical state representation, and enhanced training for extreme events, in order to support the transition of hybrid intelligent flood forecasting models from experimental research to reliable operational applications.

6 Conclusion

This study developed a differentiable hybrid hydrological forecasting framework coupled with future meteorological information for short-term flood forecasting in small and medium-sized basins. Flood forecasting experiments were conducted in the Shahe, Jumahe, and Liulin basins. The main conclusions are as follows.

(1) Forecasting accuracy decreased as lead time increaseing, indicating the progressive accumulation of uncertainty from future forcing, hydrological state evolution, and channel routing phase errors. Among the three hybrid hydrological models, PINAS showed the most stable overall performance. Compared with Muskingum and CNN unit, PINAS used a physically constrained searchable routing structure to more flexibly represent basin-scale differences and flood-wave propagation characteristics, achieving a better balance among peak propagation, hydrograph smoothing, and recession representation.

(2) The size of the training sample had a significant impact on the performance of PINAS. Increasing the amount of training data usually leads to an improvement in NSE, a reduction in

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

RMSE, and a decrease in the deviations of peak flow and runoff. However, the improvement was not entirely a monotonic trend, indicating that sample representativeness, flood-type coverage, and the inclusion of extreme events are also critical for model generalization.

(3) SHAP interpretability analysis revealed significant differences in the main sources of information among different basins. Forecasts in the Shahe and Jumahe basins were mainly influenced by measured data and future radar echoes, indicating that the spatial structure of rainfall and the short-term evolution of storms were of great significance in medium-sized basins. In contrast, forecasts in Liulin Basin were primarily driven by measured areal rainfall and gauge rainfall, reflecting the strong dependence of small basin floods on measured rainfall accumulation and localized rainfall measurements.

Overall, integrating the hybrid hydrological model with future meteorological information can enhance the stability and hydrological rationality of short-term flood forecasting for small and medium-sized basins. The proposed framework provides a feasible approach for intelligent flood forecasting models that combine forecasting accuracy, physical consistency, and interpretability.

CRediT authorship contribution statement

Leijing Li: Conceptualization, Methodology, Writing–original draft, Writing–review & editing. **Liutianjiao Hong:** Conceptualization, Methodology, Software, Validation, Writing–original draft. **Jianzhu Li:** Validation, Writing – review & editing. **Peng Shi:** Methodology, Visualization. **Shuaihang Wang:** Supervision, Data curation. **Jiyang Tian:** Data curation.

Declaration of Competing Interest

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

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Acknowledgment

This work was funded by the National Key Research and Development Program of China (No. 2024YFC3082200), the National Natural Science Foundation of China [No. 52279022].

Data availability

Data will be made available on request.

References

Astagneau, P. C., Bourgin, F., Andréassian, V., & Perrin, C. (2024). Lead-time-dependent calibration of a flood forecasting model. Journal of Hydrology, 644, 132119.

Amini, A., Dolatshahi, M., & Kerachian, R. (2024). Real-time rainfall and runoff prediction by integrating BC-MODWT and automatically-tuned DNNs: Comparing different deep learning models. Journal of Hydrology, 631, 130804.

Ayvaz, M. T., & Gurarslan, G. (2017). A new partitioning approach for nonlinear Muskingum flood routing models with lateral flow contribution. Journal of Hydrology, 553, 142-159.

Cafiero, L., Bertola, M., Mazzoglio, P., Blöschl, G., Laio, F., & Viglione, A. (2025). How changes in future precipitation impact flood frequencies: A quantile‐quantile mapping approach. Water Resources Research, 61(7), e2024WR038471.

Chadalawada, J., Herath, H. M. V. V., Babovic, V. (2020). Hydrologically informed machine learning for rainfall‐runoff modeling: A genetic programming‐based toolkit for automatic model induction. Water Resources Research, 56(4), e2019WR026933.

Chernozhukov, V., Fernández‐Val, I., & Galichon, A. (2010). Quantile and probability curves without crossing. Econometrica, 78(3), 1093-1125.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

Chung, K. S., Hsu, Y. C., Tsou, Y. H., & Lin, H. H. (2025). Investigating the spatiotemporal characteristics of motion fields using three-dimensional radar echoes to construct an ensemble nowcasting system. Quarterly Journal of the Royal Meteorological Society, 151(768), e4935.

Cibin, R., Athira, P., Sudheer, K. P., & Chaubey, I. (2014). Application of distributed hydrological models for predictions in ungauged basins: a method to quantify predictive uncertainty. Hydrological Processes, 28(4), 2033-2045.

Collins, E. L., Altenau, E. H., Pavelsky, T. M., David, C. H., Getirana, A., & Johnson, E. (2025). River network routing and discharge partitioning on a multichannel river network. Water Resources Research, 61(11), e2025WR041417.

Ettalbi, M., Garambois, P.-A., et al. (2025). Improving parameter regionalization learning for spatialized differentiable hydrological models by assimilation of satellite-based soil moisture data. Journal of Hydrology, 660, 133300.

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process‐based models with multiphysical outputs can approach state‐of‐the‐art hydrologic prediction accuracy. Water Resources Research, 58(10), e2022WR032404.

Feng, D., Beck, H., de Bruijn, J., Sahu, R. K., Satoh, Y., Wada, Y., Liu, J., Pan, M., Lawson, K., & Shen, C. (2024). Deep dive into hydrologic simulations at global scale: harnessing the power of deep learning and physics-informed differentiable models (δHBV-globe1.0-hydroDL). Geoscientific Model Development, 17, 7181–7198.

Franch, G., Tomasi, E., Wanjari, R., Poli, V., Cardinali, C., Alberoni, P. P., & Cristoforetti, M. (2025). GPTCast: a weather language model for precipitation nowcasting. Geoscientific Model Development, 18, 5351–5371.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

729 Gan, Y., Liang, X. Z., Duan, Q., Ye, A., Di, Z., Hong, Y., & Li, J. (2018). A systematic assessment
730 and reduction of parametric uncertainties for a distributed hydrological model. Journal of
731 hydrology, 564, 697-711.
732 Gao, S., Zhang, S., Huang, Y., Han, J., Zhang, T., & Wang, G. (2024). A hydrological process-based
733 neural network model for hourly runoff forecasting. Environmental Modelling & Software,
734 176, 106029.
735 He, M., Jiang, S., Ren, L., et al. (2025a). Distributed hybrid flood modeling framework: Integrating
736 physical mechanisms with deep learning for enhanced efficiency and accuracy. Water
737 Resources Research, 61(10), e2025WR039932.
738 He, M., Jiang, S., Ren, L., et al. (2025b). Exploring the performance and interpretability of hybrid
739 hydrologic model coupling physical mechanisms and deep learning. Journal of Hydrology, 649,
740 132440.
741 Herath, H. M. V. V., Marshall, L., Saha, A., Rasnayaka, S., & Seneviratne, S. (2025). Subgrid
742 informed neural networks for high-resolution flood mapping. Journal of Hydrology, 133329.
743 Huynh, N. N. T., Garambois, P. A., Renard, B., Colleoni, F., Monnier, J., & Roux, H. (2025). A
744 distributed hybrid physics–AI framework for learning corrections of internal hydrological
745 fluxes and enhancing high-resolution regionalized flood modeling. Hydrology and Earth
746 System Sciences, 29(15), 3589-3613.
747 Jamaat, A., Song, Y., Rahmani, F., Liu, J., Lawson, K., & Shen, C. (2025). Update hydrological
748 states or meteorological forcings? Comparing data assimilation methods for differentiable
749 hydrologic models. arXiv preprint arXiv:2502.16444.
750 Jahangir, M. S., & Quilty, J. (2024). Generative deep learning for probabilistic streamflow

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

751 forecasting: Conditional variational auto-encoder. Journal of Hydrology, 629, 130498. doi:
752 10.1016/j.jhydrol.2023.130498.
753 Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling
754 using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences,
755 22(11), 6005-6022.
756 Kratzert, F., Gauch, M., Klotz, D., & Nearing, G. (2024). HESS Opinions: Never train a Long Short-
757 Term Memory (LSTM) network on a single basin. Hydrology and Earth System Sciences,
758 28(17), 4187-4201.
759 Li, W., Liu, C., Xu, Y., Niu, C., Li, R., Li, M., ... & Tian, L. (2024). An interpretable hybrid deep
760 learning model for flood forecasting based on Transformer and LSTM. Journal of Hydrology:
761 Regional Studies, 54, 101873.
762 Li, J., Li, L., Zhang, T., Xing, H., Shi, Y., Li, Z., ... & Liu, J. (2024). Flood forecasting based on
763 radar precipitation nowcasting using U-net and its improved models. Journal of Hydrology,
764 632, 130871.
765 Matanó, A., Berghuijs, W. R., Mazzoleni, M., de Ruiter, M. C., Ward, P. J., & Van Loon, A. F. (2024).
766 Compound and consecutive drought-flood events at a global scale. Environmental Research
767 Letters, 19(6), 064048.
768 Mendoza, P. A., Cortes Salazar, N., Vasquez, N., Mizukami, N., & Vargas, X. (2022, December). To
769 What Extent Does River Routing Matter in Hydrological Modeling?. In AGU Fall Meeting
770 Abstracts (Vol. 2022, pp. H22O-1018).
771 Ng, K. W., Huang, Y. F., Koo, C. H., et al. (2025). Optimizing dam water level prediction through a
772 one-shot neural architecture search. Ain Shams Engineering Journal, 16(11), 103683.

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

773 Ou, Z., Nai, C., Pan, B., Zheng, Y., Shen, C., Jiang, P., Liu, X., et al. (2025). Probabilistic Diffusion
774 Models Advance Extreme Flood Forecasting. Geophysical Research Letters, 52,
775 e2025GL115705. doi: 10.1029/2025GL115705.
776 Pasche, O. C., & Engelke, S. (2024). Neural networks for extreme quantile regression with an
777 application to forecasting of flood risk. The Annals of Applied Statistics, 18(4), 2818–2839.
778 doi: 10.1214/24-AOAS1907.
779 Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., ... & Mohamed, S. (2021).
780 Skilful precipitation nowcasting using deep generative models of radar. Nature, 597(7878),
781 672-677.
782 Sattari, A., & Moradkhani, H. (2025). Coping with data scarcity in extreme flood forecasting: A
783 deep generative modeling approach. Advances in Water Resources, 204, 105063.
784 Sawadekar, K., Song, Y., Pan, M., Beck, H., McCrary, R., Ullrich, P., ... & Shen, C. (2025).
785 Improving differentiable hydrologic modeling with interpretable forcing fusion. Journal of
786 Hydrology, 659, 133320.
787 Song, Y., Bindas, T., Shen, C., et al. (2025). High‐resolution national‐scale water modeling is
788 enhanced by multiscale differentiable physics‐informed machine learning. Water Resources
789 Research, 61(4), e2024WR038928.
790 Song, L., & Han, W. (2025). The impact of assimilating radar reflectivity observations based on the
791 1D Bayesian retrieval and valid-time-shifting method on the short-term severe weather
792 forecasts. Atmospheric Research, 108370.
793 Tanhapour, M., Soltani, J., Shakibian, H., Malekmohammadi, B., Hlavcova, K., Kohnova, S., &
794 Valent, P. (2025). The enhanced integration of proven techniques to quantify the uncertainty of

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

795 forecasting extreme flood events based on numerical weather prediction models. Weather and
796 Climate Extremes, 48, 100767.
797 Tsai, WP., Feng, D., Pan, M. et al. (2021).From calibration to parameter learning: Harnessing the
798 scaling effects of big data in geoscientific modeling. Nat Commun 12, 5988
799 Wang, C., Jiang, S., Zheng, Y., et al. (2024). Distributed hydrological modeling with physics‐
800 encoded deep learning: A general framework and its application in the Amazon. Water
801 Resources Research, 60(4), e2023WR036170.
802 Wang, R., Fung, J. C. H., & Lau, A. K. H. (2024). Skillful precipitation nowcasting using physical-
803 driven diffusion networks. Geophysical Research Letters, 51(24), e2024GL110832.
804 Wen, Y., Yu, W., Zheng, F., et al. (2024). AdaNAS: Adaptively Postprocessing With Self-Supervised
805 Neural Architecture Search for Ensemble Rainfall Forecasts. IEEE Transactions on Geoscience
806 and Remote Sensing, 62, 1-10.
807 Xiang, X., Guo, S., Li, C., Sun, B., & Liang, Z. (2025). Deep learning model for flood probabilistic
808 forecasting considering spatiotemporal rainfall distribution and hydrologic uncertainty. Journal
809 of Hydrology, 655, 132879.
810 Xu, C., Zhong, P. A., Zhu, F., Xu, B., Wang, Y., Yang, L., ... & Xu, S. (2024). A hybrid model
811 coupling process-driven and data-driven models for improved real-time flood forecasting.
812 Journal of Hydrology, 638, 131494.
813 Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., ... & Wang, S. (2022). Research on particle swarm
814 optimization in LSTM neural networks for rainfall-runoff simulation. Journal of hydrology,
815 608, 127553.
816 Yao, Z., Wang, Z., Wang, D., Wu, J., & Chen, L. (2023). An ensemble CNN-LSTM and GRU

https://doi.org/10.5194/egusphere-2026-3781
Preprint. Discussion started: 7 July 2026

adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as input. Journal of Hydrology, 625, 129977.

Zheng, K., Tan, Q., Ruan, H., Zhang, J., Luo, C., Tang, S., ... & Cheng, J. (2024). GAN-argcPredNet v2. 0: a radar echo extrapolation model based on spatiotemporal process enhancement. Geoscientific Model Development, 17(1), 399-413.

Zhong, L., Lei, H., Yang, J. (2024). Development of a distributed physics‐informed deep learning hydrological model for data‐scarce regions. Water Resources Research, 60(6), e2023WR036333.