https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Physics-Informed Spatiotemporal Graph Convolutional Network for Flash Flood Forecasting and Inundation Simulation

Ziqi Zhang[1,2], Qian Zhang[1], Shaokun Fu[2], Zhenghang Chen[2], Mengfei Zhang[2], Naliang Guo[2], Fubao Sun[2], Feng Wu[2]
[1]College of Land Science and Technology, China Agricultural University, Beijing 100193, China
[2]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Correspondence to: Feng Wu (wufeng@igsnrr.ac.cn)

Abstract. Flash floods pose a significant threat in mountainous regions, where rapid-onset inundation requires timely and reliable forecasting for effective emergency response. Physically based hydrodynamic models provide strong interpretability but require detailed terrain, roughness, rainfall, and initial-condition information, and their computational cost limits real-time forecasting and repeated scenario analysis. Data-driven models offer higher efficiency, but grid-based architectures may distort irregular hydrodynamic meshes, and purely data-driven graph models can generate physically implausible predictions during long-horizon autoregressive simulation. To address these limitations, this study proposes a physics-informed spatiotemporal graph convolutional network, termed PI-STGCN, for rapid flash flood inundation simulation over irregular triangular control-volume meshes. The model represents the hydrodynamic domain as a directed dual graph, where triangular cells are treated as nodes and shared interfaces are treated as directed edges with geometric and hydraulic attributes. PI-STGCN integrates edge-conditioned spatial graph convolution, rainfall encoding, causal temporal convolution, multi-step residual prediction, and rollout-aware training through scheduled sampling and pushforward rollout. Shallow-water-equation residuals, water-volume consistency, and near-channel boundary constraints are used as soft physical regularization terms to improve local hydrodynamic plausibility. The model was trained using hydrodynamic simulation outputs from design storm events and evaluated on an unseen 100a24h extreme design storm and an independent real rainfall-driven flash flood event. For the 100a24h event, PI-STGCN achieved $R^2 = 0.9142$, NSE = 0.9088, KGE = 0.8565, MAE = 0.3390, and RMSE = 0.7892. Ablation and residual diagnostics show that spatial graph convolution and residual prediction are critical for accuracy and rollout stability, while physics-informed regularization improves momentum-equation consistency, although its effect on conventional predictive metrics is event- and metric-dependent. The trained model reduced inference time from 63–105 s for hydrodynamic simulations to 1.06–1.08 s. These results indicate that PI-STGCN provides an efficient physics-regularized surrogate for rapid inundation prediction on irregular meshes in mountainous catchments.

1 Introduction

In recent years, flash floods in mountainous regions have become increasingly frequent and severe under the influence of extreme rainfall events (Wasko et al., 2021). Mountainous catchments usually have steep terrain gradients and short flow-

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

concentration paths (Hu and Song, 2018). Rainfall in these areas is highly variable, and runoff responses are often rapid. These features make flood forecasting and inundation simulation difficult, particularly when predictions must be produced within the limited time available for emergency response (Liao et al., 2023). Therefore, accurate and efficient flood forecasting and inundation simulation are essential for flash-flood risk assessment, early warning, and emergency decision making.

Traditionally, flood forecasting and inundation simulation have relied on physically based hydrological and hydrodynamic models, including distributed hydrological models, one-dimensional and two-dimensional hydrodynamic models, and coupled hydrological-hydrodynamic models (Kumar et al., 2023). These models explicitly represent rainfall-runoff generation, flow concentration, and flood-routing processes, and therefore provide strong physical interpretability (Zhou et al., 2021). However, they typically require high-resolution terrain data, accurate rainfall boundary conditions, reliable roughness parameters, and appropriate initial hydrological states (Moghim et al., 2023). In mountainous catchments, the pronounced spatial variability of rainfall, topography, land-surface properties, and channel networks can introduce substantial parameter and boundary-condition uncertainties (Yoganand et al., 2025; Zhang et al., 2024). In addition, high-resolution hydrodynamic simulations over irregular terrain are computationally expensive, which limits their use in real-time forecasting and repeated scenario analysis (Albano et al., 2024; Fraehr et al., 2023; Ni et al., 2020).

To overcome these computational constraints, data-driven approaches have been increasingly adopted for flood forecasting because of their computational efficiency and ability to approximate complex nonlinear hydrological responses (Bentivoglio et al., 2022). Deep-learning models, including convolutional neural networks, recurrent neural networks, long short-term memory networks, and Transformer-based architectures, can automatically extract spatiotemporal features from continuous hydrological sequences and have shown promising performance in flood prediction (Li et al., 2024; Maity et al., 2024; Yin et al., 2022; Zhou et al., 2022). However, many deep-learning models rely on grid-based representations, which require rasterization of the computational domain (Fraehr et al., 2023). This process may simplify or distort the original mesh topology and local geometric relationships, especially in hydrodynamic models based on unstructured triangular meshes. Graph convolutional networks (GCNs) provide a flexible alternative for representing flood dynamics over irregular spatial structures (Peng et al., 2024; Tam et al., 2022). Unlike grid-based convolutional networks, GCNs can operate directly on unstructured meshes, river networks, and irregular topological relationships. This property is particularly suitable for hydrodynamic models based on triangular irregular networks. In a triangular control-volume mesh, each cell can be represented as a graph node, and each shared interface between adjacent cells can be represented as a graph edge (Sun et al., 2022). This graph representation is consistent with finite-volume hydrodynamic discretization and enables the model to learn spatial interactions between physically connected control volumes (Brandstetter et al., 2023).

Despite this advantage, purely data-driven graph models may produce physically inconsistent predictions during long-horizon autoregressive simulation (Shao et al., 2024). In flood modeling, small errors in water depth and momentum can accumulate over time, causing unrealistic water-depth fluctuations, storage imbalance, and unstable momentum evolution. These problems are particularly important when the model is initialized from dry or partially wet conditions and then

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

recursively predicts future hydraulic states (Zhang et al., 2025). Therefore, a reliable flood surrogate model should not only learn spatiotemporal patterns from simulation data but also be regularized by physically meaningful constraints (Yuan et al., 2020). Physics-informed learning provides a potential route for incorporating hydrodynamic knowledge into data-driven flood simulation (He et al., 2025). In this study, the term "physics-informed" refers to the use of shallow-water-equation (SWE) residuals and water-volume consistency terms as soft regularization constraints during model training. This differs from physics-informed neural networks (PINNs) that solve governing equations over continuous space-time domains using automatic differentiation (Raissi et al., 2019; Shen et al., 2023). Although PINNs can improve physical consistency, they often face challenges in flood forecasting and inundation simulation, including high computational cost, unstable training under steep hydraulic gradients, difficulty in treating wetting-drying processes, and limited flexibility for irregular meshes (Gao et al., 2021). These limitations motivate the development of a graph-based physics-informed surrogate model that can preserve the irregular triangular mesh structure while incorporating hydrodynamic constraints (Lu et al., 2025).

The central hypothesis of this study is that an irregular-mesh spatiotemporal graph convolutional network constrained by shallow-water-equation residuals and water-volume consistency can improve the extrapolative stability and physical plausibility of rapid flood forecasting and inundation simulation compared with purely data-driven graph models (Bentivoglio et al., 2023). To test this hypothesis, this study proposes a physics-informed spatiotemporal graph convolutional network, referred to as PI-STGCN, for flood forecasting and inundation simulation over irregular triangular control-volume meshes. The model is trained using hydrodynamic simulation outputs as reference data and evaluated under both design rainfall scenarios and an independent real rainfall-driven flash-flood event (Zhang et al., 2023). The proposed PI-STGCN has three main features:

1. It represents the computational domain using the dual graph of the triangular irregular network, where triangular control volumes are treated as nodes and shared cell interfaces are treated as directed edges. Static terrain, land-use, roughness, and channel-related attributes are assigned to nodes, while geometric and hydraulic connectivity is encoded through edge attributes.

2. It integrates edge-conditioned spatial graph convolution, a rainfall encoder, causal temporal convolution, multi-step residual prediction, and physics-informed regularization to learn spatiotemporal flood evolution patterns. Shallow-water-equation residuals, water-volume consistency terms and near-channel boundary conditions are added as soft regularization constraints to encourage locally plausible mass and momentum dynamics during autoregressive prediction.

3. It adopts scheduled sampling and pushforward rollout training to reduce the gap between training and recursive inference. These rollout-aware strategies expose the model to its own predictions during training, thereby improving long-horizon stability and reducing error accumulation.

The remainder of this paper is organized as follows. In Sect. 2, we describe the study area, triangular irregular network, rainfall scenarios, and hydrodynamic simulation data. Then, in Sect. 3, we present the proposed methodology, including problem formulation, graph construction, PI-STGCN architecture, and model setup. In Sect. 4, we present the experimental results, including prediction accuracy, temporal and spatial flood dynamics, ablation experiments, and computational efficiency. In Sect. 5, we discuss model accuracy, interpretability, generalization and computational advantages. Finally, conclusions are provided in Sect. 6.

2 Study Area and Data

2.1 Study Area

The study area is the Taoyukou watershed, located in Changping District, Beijing, China. The watershed lies in the mountainous region of northern Beijing and covers approximately 37.73 km² (Fig. 1). The watershed is characterized by complex terrain, pronounced elevation gradients, and a dendritic river network.

Topographic information was derived from the ALOS PALSAR radiometric terrain-corrected digital elevation model (DEM) with a spatial resolution of 30 m. Elevation within the watershed ranges from approximately 74.5 m to 823.8 m. Higher elevations are mainly distributed in the upstream mountainous areas, whereas lower elevations are concentrated along the downstream valley and river-channel zones. The elevation profile in Fig. 1 shows a clear upstream-to-downstream decline, indicating strong topographic control on runoff concentration and flood propagation.

The river network was extracted from the DEM using hydrological terrain analysis, including flow-direction and flow-accumulation calculations. The extracted channel network serves as the primary pathway for flood routing within the watershed. Owing to the steep terrain and short flow-concentration distances, rainfall-generated runoff can rapidly converge into the channel system, resulting in flash-flood processes with strong temporal variability (Liu et al., 2026; Ma et al., 2025). Therefore, accurate representation of terrain characteristics, channel connectivity, and local hydraulic exchange is essential for reliable flood forecasting and inundation simulation in this watershed.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

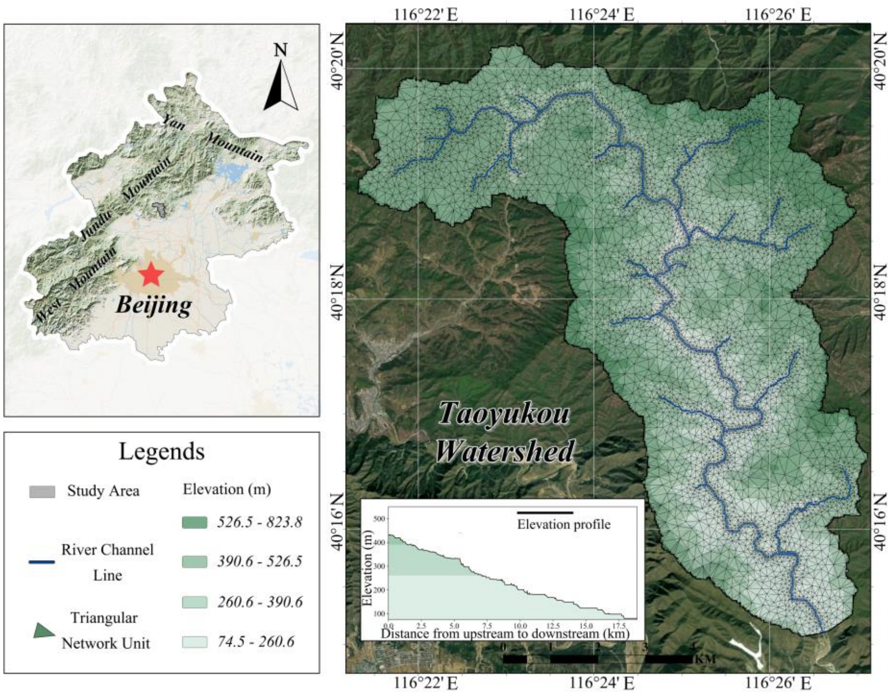


Figure 1. Location and topographic characteristics of the Taoyukou watershed. The map shows the watershed boundary, triangular irregular network, river-channel system, elevation distribution, and upstream-to-downstream elevation profile.

2.2 Triangular Irregular Network

The hydraulic computational domain was represented using a triangular irregular network (TIN). The TIN was generated by integrating the watershed boundary, DEM, land-use/land-cover data, and river network information. This mesh-generation strategy preserves key topographic and channel features while allowing flexible spatial discretization over complex terrain.

Mesh resolution was adaptively refined according to hydraulic relevance. Areas adjacent to the river network were discretized using finer triangular elements because flood routing, channel-floodplain exchange, and rapid water-depth variations are most pronounced in these regions. In contrast, regions farther from the river network were represented by coarser elements because hydraulic gradients and flow variability are generally weaker. This variable-resolution design improves computational efficiency while maintaining detailed hydraulic representation in dynamically important regions.

Each triangular element was treated as a control volume in the hydrodynamic simulation and later as a graph node in the PI-STGCN model. Shared interfaces between adjacent triangular elements define local hydraulic connectivity and are converted into graph edges. In this way, the same triangular control-volume structure provides both the numerical discretization for hydrodynamic simulation and the spatial graph foundation for data-driven prediction.

The final TIN consists of 13,055 triangular control volumes and 19,415 graph edges, providing a detailed representation of the watershed topography and hydraulic connectivity. The minimum and maximum triangular element areas are 28.51 m²

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

and 18,438.20 m², respectively, while the average element area is 3,065.78 m². The wide range of element sizes reflects the adaptive mesh-generation strategy, in which finer discretization is applied in hydraulically important regions, particularly near river channels, and coarser discretization is used in less dynamic areas. These mesh characteristics directly influence the computational cost of the hydrodynamic simulations, the size of the graph representation, and the inference efficiency of the PI-STGCN model.

2.3 Rainfall Scenarios

Rainfall forcing is the primary external driver of flash-flood generation in the Taoyukou watershed. In this study, both design storm events and a real flash-flood event were used to evaluate the proposed PI-STGCN model.

The design storm events were obtained from the water conservancy department and include four 24 h rainfall scenarios with different return periods: 10a24h, 20a24h, 50a24h, and 100a24h. These design storms represent rainfall processes with increasing intensity and provide controlled forcing conditions for model training, validation, and rainfall-magnitude extrapolation assessment. The 10a24h, 20a24h, and 50a24h events were used for model training, while the 100a24h event was reserved for validation. Because the validation event has a higher return period than all training events, this split evaluates extrapolation to a more extreme design rainfall scenario rather than interpolation within the training range.

In addition to the design storms, a real flash-flood rainfall event, denoted as Event 237, was used as an independent extrapolation case. Rainfall data for this event were obtained from the China Meteorological Administration Land Data Assimilation System dataset (Sun et al., 2020). The event lasted approximately 72 h and was divided into three consecutive 24 h periods: Event 237 (0-24 h), Event 237 (24-48 h), and Event 237 (48-72 h). This segmentation enables the real event to be evaluated under the same prediction framework as the 24 h design storms.

The use of both design storms and a real rainfall event allows the model to be assessed under two levels of generalization. The 100a24h validation event tests rainfall-magnitude extrapolation from lower- and medium-return-period design storms to a more extreme return-period event. Event 237 provides an additional event-type transfer test, in which the model trained on design storms is applied to a real rainfall-driven flash-flood process with more complex temporal variability.

The rainfall forcing data used in this study have a temporal resolution of 1 h and a spatial resolution of 500 m × 500 m. Rainfall values were spatially assigned to triangular control volumes according to the location of each control-volume centroid within the rainfall grid. The design storms exhibit progressively increasing rainfall severity, with maximum hourly intensities of 75.99, 92.15, 113.37, and 129.35 mm h^{-1} and total rainfall amounts of 215.85, 279.83, 367.16, and 434.72 mm for the 10a24h, 20a24h, 50a24h, and 100a24h events, respectively. For the real flash-flood event, the three consecutive 24 h periods have maximum hourly intensities of 41.68, 28.00, and 29.49 mm h^{-1} and total rainfall amounts of 183.80, 172.40, and 109.77 mm, respectively. The corresponding average rainfall intensities are 8.99, 11.66, 15.30, and 18.11 mm h^{-1} for the design storms and 9.19, 9.07, and 4.57 mm h^{-1} for the three Event 237 segments.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

2.4 Hydrodynamic Simulation Results

Based on the rainfall scenarios, a two-dimensional hydrodynamic model was used to generate reference flash-flood simulation data for the Taoyukou watershed. The hydrodynamic model simulated rainfall-driven flood propagation over the triangular computational mesh and provided spatially and temporally distributed hydraulic variables for model training, validation, and extrapolation evaluation.

For each rainfall scenario, the hydrodynamic model produced time-series outputs of water depth and depth-averaged velocity components for each triangular control volume. These simulation outputs were used as reference data for training and evaluating the proposed PI-STGCN model. Therefore, the accuracy reported in this study represents agreement between PI-STGCN predictions and reference hydrodynamic simulation results, rather than direct validation against observed inundation measurements.

The target variables in this study are water depth (h) and the two momentum components (q_x) and (q_y). The original hydrodynamic simulation outputs include water depth (h) and depth-averaged velocity components (u) and (v). To ensure consistency with the conservative form of SWE, the velocity components were converted into momentum variables as: $q_x = h{\cdot}u$, $q_y = h{\cdot}v$. These hydrodynamic simulation data serve two purposes. First, they provide supervised target variables for training the data-driven component of PI-STGCN. Second, they provide the hydraulic states required to compute the physics-informed SWE residuals, water-volume consistency terms and near-channel boundary conditions during model optimization.

3 Methodology

The objective of the proposed model is to learn flood inundation dynamics from historical hydraulic states and rainfall forcing, and to predict future water depth and momentum fields over an irregular triangular control-volume mesh. Figure 2 summarizes the overall workflow of this study, which consists of four components: data preparation, graph construction, the physics-informed ST-GCN architecture, and model setup.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

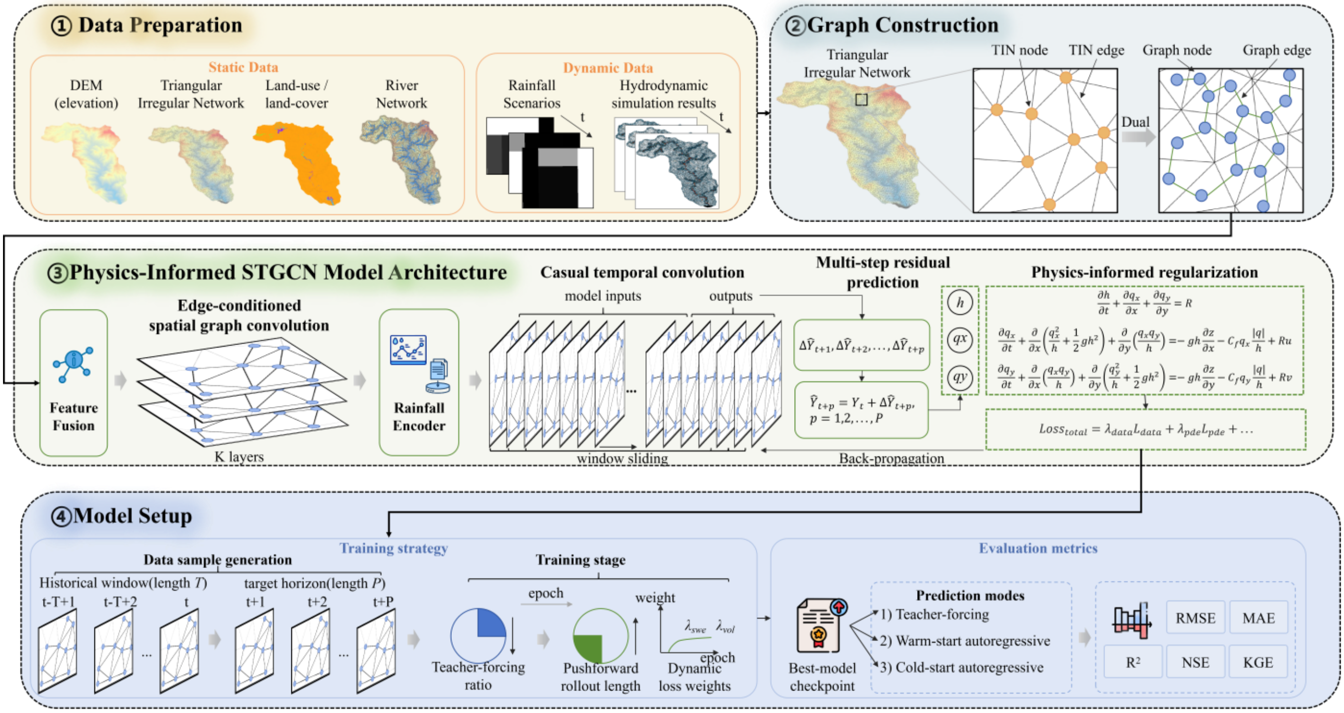


Figure 2. Workflow of this study. The workflow includes data preparation, graph construction, the physics-informed ST-GCN model architecture, and model setup.

3.1 Data Preparation

Flash flood forecasting and inundation simulation are formulated as a spatiotemporal forecasting problem over a triangular irregular control-volume mesh. Hydrodynamic simulation outputs are used as reference data, including water depth and momentum components at each triangular control volume and each simulation time step. Rainfall forcing is mapped to the same triangular control volumes to provide time-varying external input. Static catchment attributes are also assigned to each control volume to describe the spatial heterogeneity of terrain, land surface, roughness, and channel-related properties.

The input features are divided into node attributes and edge attributes. Node attributes describe information associated with individual triangular control volumes. Static node features include cell-center elevation, control-volume area, terrain slope, terrain aspect, Manning roughness coefficient, land-use/land-cover type, channel indicators, and distance-to-channel information, while dynamic node features include water depth, x-direction momentum, y-direction momentum, and rainfall intensity. Edge attributes describe the geometric and hydraulic relationship between neighboring triangular control volumes, including shared-interface length, mesh-edge length, normal-vector components, elevation difference, prior flow-direction indicators, channel-edge indicators, and the distance from the edge midpoint to the nearest channel. These attributes provide interface-level information for learning local flood exchange between adjacent control volumes and are also used in the physics-informed regularization terms. The complete feature set is summarized in Table 1.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Table 1. Input features used in the proposed PI-STGCN model.

Object	Attribute type	Attribute	Description	Role in model
Node	Static	Z	Cell-center elevation	Terrain feature
		Control_Area	Area of triangular control volume	Geometric feature; also used in volume constraint
		Slope	Terrain slope derived from DEM	Terrain feature
		Sin_Aspect	Sine component of terrain aspect	Terrain feature
		Cos_Aspect	Cosine component of terrain aspect	Terrain feature
		Roughness	Manning roughness coefficient	Hydraulic feature
		Lulc	Land-use / land-cover category	Land-surface feature
		Is_Ch	Channel cell indicator	Channel-related feature
		Dist2Ch	Distance from triangular cell to channel	Channel-related feature
		Ch_Len_in	Channel length inside the triangular cell	Channel-related feature
	Dynamic	h	Water depth	Hydraulic input variable
		q_x	X-direction momentum	Hydraulic input variable
		q_y	Y-direction momentum	Hydraulic input variable
		rain	Rainfall forcing at each node and time step	External forcing input
Edge	Static	Sh_Len	Shared-edge length between two adjacent triangular cells	Geometric feature; also used in SWE residual
		Edge_Len	Geometric length of the mesh edge	Geometric feature
		N_x	X-component of the edge normal vector	Directional edge feature; also used in SWE residual
		N_y	Y-component of the edge normal vector	Directional edge feature; also used in SWE residual
		Diff_Z	Elevation difference between adjacent triangular cells	Topographic relation
		Dir_Prior	Prior flow-direction indicator	Hydraulic direction prior
		Edge_Ch	Channel-edge indicator	Channel-related edge feature
		Mid2Ch	Distance from edge midpoint to channel	Channel-related edge feature

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

3.2 Graph Construction

The triangular computational domain is converted into a directed graph $G = (V, E)$, where V denotes the set of graph nodes and E denotes the set of directed graph edges. Each node $i \in V$ corresponds to one triangular control volume, and each edge $(i, j) \in E$ represents the hydraulic connectivity between two adjacent control volumes. This graph representation preserves the original unstructured mesh topology and allows flood propagation to be modeled through information exchange between physically connected cells.

For each shared interface between two adjacent triangular cells, two directed graph edges are generated, one in each direction. This bidirectional representation is required because hydraulic exchange is not fixed in a single direction and can vary over time with local water-surface gradients, bed slope, rainfall forcing, and channel connectivity. For reverse edges, orientation-dependent quantities, including normal-vector components and elevation difference, are assigned opposite signs. Therefore, the directed graph allows bidirectional message passing while preserving the geometric directionality required for hydraulic interaction and shallow-water-equation residual computation.

At each time step t, the dynamic input state at node i is defined as

$$x_{i,t} = [h_{i,t}, q_{x,i,t}, q_{y,i,t}, R_{i,t}], \tag{1}$$

where $h_{i,t}$ is the water depth, $q_{x,i,t}$ and $q_{y,i,t}$ are the momentum components in the x- and y-directions, respectively, and $R_{i,t}$ is the rainfall intensity. The corresponding prediction target is

$$y_{i,t} = [h_{i,t}, q_{x,i,t}, q_{y,i,t}], \tag{2}$$

For the whole graph, the dynamic input matrix at time t is denoted as $X_t \in \mathrm{R}^{N\mathrm{x}Fx}$, where N is the number of graph nodes and $F_x = 4$ is the number of dynamic input features. The prediction target is denoted as $Y_t \in \mathrm{R}^{N\mathrm{x}Fy}$, where $F_y = 3$ corresponds to water depth and two momentum components.

Given a historical sequence of length T, the forecasting task is expressed as

$$(X_{t-T+1}, X_{t-T+2}, \dots, X_t) \rightarrow (\hat{Y}_{t+1}, \hat{Y}_{t+2}, \dots, \hat{Y}_{t+P}), \tag{3}$$

where P is the prediction horizon. Static node attributes and directed edge attributes remain fixed during prediction, whereas hydraulic states and rainfall forcing vary over time. Together, these inputs enable the model to learn spatiotemporal flood propagation over the irregular triangular mesh.

3.3 PI-STGCN Architecture

The proposed PI-STGCN is designed to learn both spatial hydraulic interactions between neighboring control volumes and temporal flood evolution driven by rainfall forcing, terrain properties, and channel connectivity. The model combines edge-conditioned spatial graph convolution, rainfall encoding, causal temporal convolution, multi-step residual prediction, and physics-informed regularization.

For a batch of training samples, the dynamic input tensor is denoted as

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

$$X_{dyn} \in R^{BxTxNx4} \tag{4}$$

where B is the batch size, T is the historical sequence length, N is the number of graph nodes, and the four dynamic variables are water depth, x-direction momentum, y-direction momentum, and rainfall intensity. The model uses a historical dynamic sequence, static node attributes S, and directed edge features E to predict future hydraulic state increments:

$$\Delta\hat{Y}_{t+1:t+P} = f_\theta(X_{t-T+1:t}, S, E), \tag{5}$$

where f_θ denotes the proposed PI-STGCN model. Instead of directly predicting absolute future hydraulic states, the model predicts increments relative to the latest input state. The predicted state at prediction step p is reconstructed as:

$$\hat{\mathrm{Y}}_{\mathrm{t+p}} = \mathrm{Y}_{\mathrm{t}} + \Delta\hat{\mathrm{Y}}_{\mathrm{t+p}}, \tag{6}$$

This residual formulation reduces the learning difficulty because the model only needs to learn hydraulic-state changes rather than full future states. It is also compatible with autoregressive simulation, in which predicted states are recursively appended to the input sequence for long-horizon flood rollout.

The spatial module is based on edge-conditioned graph convolution. For each node, information from neighboring control volumes is aggregated through directed edges whose attributes describe interface geometry and hydraulic directionality. This allows the model to learn spatial flood propagation through physically meaningful mesh interfaces rather than through binary adjacency alone. Multiple spatial graph-convolution layers are stacked to capture higher-order hydraulic interactions.

Rainfall forcing is processed by a rainfall encoder. This design is used because rainfall intensity may have a different numerical scale from water depth and momentum after normalization. The rainfall encoder enables the model to learn rainfall accumulation and lag effects before combining rainfall information with hydraulic-state features. After spatial graph convolution and rainfall encoding, a causal temporal convolution module is used to learn temporal dependencies from the historical sequence. Causal convolution ensures that the prediction at each time step depends only on current and previous information, which is required for forecasting.

The physics-informed component is introduced as a soft regularization term in the training objective. It does not replace the hydrodynamic solver. Instead, it penalizes neural-network predictions that violate local shallow-water-equation residuals and water-volume consistency.

The conservative hydraulic variables are defined as

$$U = \left[h, q_x, q_y\right]^T, \tag{7}$$

The two-dimensional shallow-water equations are used to evaluate local mass and momentum residuals over triangular control volumes. In this study, the residual calculation uses finite-volume-related quantities, including control-volume area, shared-edge length, edge-normal direction, elevation difference, rainfall forcing, bed-slope source terms, and friction effects (Xia et al., 2017). For each node i, the finite-volume residual can be written compactly as

$$A_i \frac{U_i^{t+1} - U_i^t}{\Delta t} + \sum_{j \in N(i)} F_{ij} L_{ij} - S_i A_i = r_i, \tag{8}$$

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

where A_i is the area of control volume, Δt is the time interval, L_{ij} is the shared-edge length between nodes i and j, F_{ij} is the numerical flux through the shared interface, S_i is the source term, and r_i is the residual vector. The physics-informed loss is then defined as

$$L_{pde} = \frac{1}{N}\sum_{i=1}^{N}\left(r_{h,i}^2 + r_{qx,i}^2 + r_{qy,i}^2\right), \tag{9}$$

where $r_{h,i}$, $r_{qx,i}$, and $r_{qy,i}$ are the continuity, x-momentum, and y-momentum residuals, respectively. In addition to the shallow-water-equation residual loss, a water-volume consistency penalty is used to constrain total predicted water storage. The predicted and reference water volumes at prediction step p are calculated as

$$\begin{cases} \hat{V}_{t+p} = \sum_{i=1}^{N} \hat{h}_{i,t+p} A_i \,, \\ V_{t+p}^{true} = \sum_{i=1}^{N} h_{i,t+p}^{true} A_i \,, \\ \Delta V_R = \sum_{i=1}^{N} R_{i,t} A_i \,, \end{cases} \tag{10}$$

where $\hat{V}_{t+p}$ is the predicted water volume at prediction step p, V_{t+p}^{true} is the reference water volume, ΔV_R is the rainfall-induced water input, and A_i is the area of triangular control volume i, and N is the number of graph nodes. The water-volume consistency loss is defined as

$$L_{vol} = \frac{1}{P}\sum_{p=1}^{P}\left[\frac{\max(\hat{V}_{t+p} - V_t - \Delta V_R, 0)}{\max(V_t, 1)} + \frac{\max(V_{t+p}^{true} - \hat{V}_{t+p} - \Delta V_R, 0)}{\max(V_t, 1)}\right], \tag{11}$$

where V_t is the current water volume. The first term penalizes excessive predicted storage relative to the current storage plus rainfall input, whereas the second term penalizes excessive underestimation relative to the reference future storage. Thus, L_{vol} acts as a soft water-storage consistency constraint rather than a strict mass-conservation equation.

3.4 Model Setup

The model was trained using a strategy that combines supervised learning, near-channel depth correction, physics-informed regularization, water-volume consistency, and pushforward autoregressive rollout. For each training sample, the model predicts future hydraulic states over a prediction horizon P. The output variables are water depth and the two momentum components.

The supervised data loss is computed as a wet-node-weighted mean squared error. This weighting is introduced because flood inundation data are highly imbalanced: many cells remain dry or nearly dry during part of the simulation, whereas wet cells contain most of the hydraulic information relevant to inundation prediction. The data loss is expressed as

$$L_{data} = \frac{1}{BPN}\sum_{b=1}^{B}\sum_{p=1}^{P}\sum_{i=1}^{N} ||\hat{y}_{i,t+p}^{b} - y_{i,t+p}^{b}||_2^2, \tag{12}$$

where B is the batch size, P is the prediction horizon, N is the number of graph nodes, $y_{i,t+p}^{b}$ is the reference hydraulic state. In implementation, wet nodes are assigned larger weights to reduce the dominance of dry cells in the loss function.

A near-channel boundary-condition loss is also introduced to reduce underestimation of water depth around river channels and adjacent floodplain areas. A node is regarded as near-channel if it is a channel cell or if its distance to the channel is less than 50 m. The loss penalizes negative depth bias in near-channel wet regions:

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

$$L_{bc} = \frac{1}{|\Omega_c|} \sum_{i \in \Omega_c} \left[max\left(h_{i,t}^{true} - \hat{h}_{i,t}, 0\right) \right]^2 \tag{13}$$

where Ω_c denotes the set of wet nodes located in or near the channel region. This term penalizes negative depth bias in hydraulically important wet channel regions.

The physics-informed losses L_{pde} and L_{vol} defined in Sect. 3.3 penalize violations of the shallow-water-equation residuals and water-volume consistency residuals.

Pushforward rollout is introduced to reduce the mismatch between training and inference. Instead of optimizing only the first prediction window, the model is unrolled for multiple future steps. The rollout length increases gradually after epoch 40 and reaches a maximum of 4 steps. This curriculum-learning strategy allows the model to first learn stable short-horizon prediction and then adapt to recursive long-horizon forecasting. The rollout loss is calculated by comparing each autoregressive prediction step with its corresponding reference sequence:

$$L_{roll} = \frac{1}{BKN} \sum_{b=1}^{B} \sum_{k=1}^{K} \sum_{i=1}^{N} ||\hat{y}_{i,t+k}^{b} - y_{i,t+k}^{b}||_2^2, \tag{14}$$

where K is the number of rollout prediction steps. This term forces the model to remain accurate when its own previous predictions are recursively used as future inputs. The overall training objective is defined as

$$L_{total} = w_{data} L_{data} + w_{bc} L_{bc} + w_{pde} L_{pde} + w_{vol} L_{vol} + w_{roll} L_{roll}, \tag{15}$$

where w_{data}, w_{bc}, w_{pde}, w_{vol} and w_{roll} are the corresponding loss weights.

In this study, the model was optimized using the AdamW optimizer with weight decay (Tab. 2). The learning rate was adjusted using a ReduceLROnPlateau scheduler based on the validation loss, and gradient clipping was applied to prevent unstable parameter updates during rollout training. The implemented configuration used 300 training epochs, a batch size of 4, an initial learning rate of $1 \cdot 10^{-3}$, a weight decay of $1 \cdot 10^{-5}$, and a gradient-clipping threshold of 1.0. The historical sequence length was set to $T = 8$, and the prediction horizon was set to $P = 6$. The 10a24h, 20a24h, and 50a24h design storm events were used for training, while the 100a24h event was used for validation. This event split was designed to evaluate rainfall-magnitude extrapolation. Specifically, the model was trained using lower- and medium-magnitude rainfall scenarios and then tested on the more extreme 100a24h event, which lies outside the range of the training events. Compared with an interpolation setting, in which the validation event lies within the training-event range, this extrapolation setting provides a stricter assessment of model generalization.

Table 2. Model setup parameters used for PI-STGCN training.

Category	Parameter	Value
Input-output setting	Historical sequence length, T	8 time steps
	Prediction horizon, P	6 time steps
Model architecture	Batch size, B	4
	Number of spatial graph-convolution layers	3

Category	Parameter	Value
	Number of temporal-convolution layers	1
	Temporal kernel size	1
	Dropout rate	0.1
Training configuration	Number of training epochs	300
	Initial learning rate	1×10^{-3}
	Weight decay (AdamW)	1×10^{-5}
	Scheduler patience	15 epochs
	Scheduler factor	0.5
	Gradient clipping threshold	1
Scheduled sampling	Start epoch	40
	End epoch	100
	Rollout start epoch	40
	Rollout ramp-up duration	100 epochs
	Maximum rollout steps	4
Pushforward rollout	Data loss weight, w_{data}	10
	Boundary-condition loss weight, w_{bc}	1
	Initial PDE loss weight, w_{pde_init}	0.1
	Final PDE loss weight, w_{pde_final}	1.0
	Initial volume loss weight, w_{vol_init}	0.1
	Final volume loss weight, w_{vol_final}	1
	Rollout loss weight, w_{roll}	10
	PDE warm-up epoch	20
	PDE ramp-up duration	100 epochs
	Volume-loss warm-up epoch	20
	Volume-loss ramp-up duration	100 epochs

The trained model was evaluated from both temporal and spatial perspectives. Water depth was used as the primary evaluation variable because it directly determines flood inundation extent and hazard intensity (Rasmy et al., 2019).

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Although the model also predicts the two momentum components and these variables are used in the physics-informed loss, water depth is the main variable for assessing inundation accuracy.

The root mean square error (RMSE) is defined as

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{h}_m - h_m\right)^2}, \quad (16)$$

where h_m and $\hat{h}_m$ are the reference and predicted water depths, respectively, and M is the number of evaluated samples over selected nodes and time steps. RMSE measures the absolute magnitude of prediction error and is sensitive to large individual errors.

The mean absolute error (MAE) is defined as

$$\mathrm{MAE} = \frac{1}{\mathrm{M}}\sum_{\mathrm{m}=1}^{\mathrm{M}} |\hat{\mathrm{h}}_{\mathrm{m}} - \mathrm{h}_{\mathrm{m}}|, \quad (17)$$

MAE measures the average absolute prediction error and is less sensitive to large individual errors than RMSE.

The coefficient of determination is calculated as the squared Pearson correlation coefficient

$$R^2 = \left[\frac{\sum_{m=1}^{M}(h_m - \bar{h})\left(\hat{h}_m - \bar{\hat{h}}\right)}{\sqrt{\sum_{m=1}^{M}(h_m - \bar{h})^2}\sqrt{\sum_{m=1}^{M}\left(\hat{h}_m - \bar{\hat{h}}\right)^2}}\right]^2, \quad (18)$$

where $\bar{h}$ and $\bar{\hat{h}}$ are the mean reference and predicted water depths, respectively. The R^2 value measures the strength of linear agreement between predicted and reference water-depth variations.

The Nash-Sutcliffe efficiency (NSE) is defined as

$$NSE = 1 - \frac{\sum_{m=1}^{M}\left(h_m - \hat{h}_m\right)^2}{\sum_{m=1}^{M}(h_m - \bar{h})^2}, \quad (19)$$

NSE evaluates the predictive skill relative to using the mean reference value as the prediction.

The Kling-Gupta efficiency (KGE) is used to evaluate correlation, variability, and bias simultaneously (Gupta et al., 2009):

$$KGE = 1 - \sqrt{(\gamma - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}, \quad (20)$$

where γ is the Pearson correlation coefficient between predicted and reference water depths, $\alpha = \sigma_{\hat{h}}/\sigma_h$ is the variability ratio, and $\beta = \mu_{\hat{h}}/\mu_h$ is the bias ratio. Here, $\sigma_{\hat{h}}$ and σ_h denote the standard deviations of predicted and reference water depths, respectively, while $\mu_{\hat{h}}$ and μ_h denote their corresponding mean values.

The model was evaluated under three prediction modes. The first mode is teacher-forcing prediction, in which the model uses observed historical states to make short-horizon predictions. The second mode is warm-start autoregressive prediction, in which the model is initialized using an observed hydraulic sequence and then recursively predicts future states. The third mode is cold-start autoregressive prediction, in which the initial hydraulic state is set to zero and the model generates flood evolution using rainfall forcing and static graph information. These three modes were used to assess short-term accuracy, long-horizon autoregressive stability, and the ability to simulate flood development from dry initial conditions.

For visualization, predicted water-depth maps, reference water-depth maps, and spatial error maps were generated at selected time steps. Hydrographs at representative nodes were also compared to examine flood arrival time, peak depth, and recession

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

behavior. Together, the global metrics, flood-depth maps, spatial error distributions, and temporal hydrographs provide a comprehensive evaluation of the proposed PI-STGCN model.

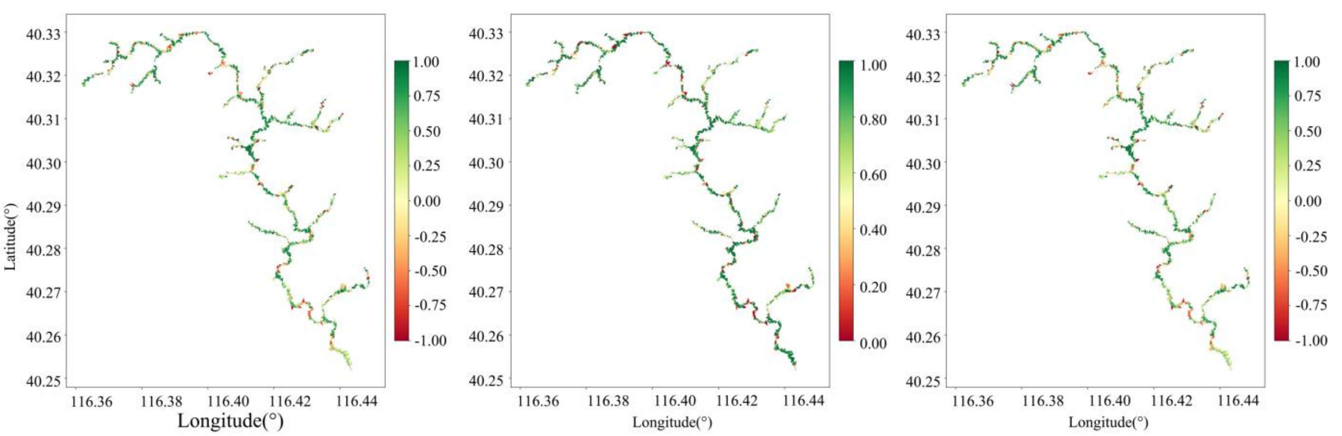
4 Results

4.1 Prediction Accuracy

We evaluated the overall prediction accuracy of the proposed PI-STGCN model using the 100a24h event as the validation case. The model was trained using the 10a24h, 20a24h, and 50a24h rainfall events, while the 100a24h event was excluded from training. This validation experiment therefore represents a rainfall-magnitude extrapolation test, in which the trained model is required to predict flood inundation dynamics under a rainfall scenario more extreme than those used for model calibration.

The spatial distribution of node-level prediction metrics for the 100a24h validation event shows that PI-STGCN maintains strong predictive skill over the main inundated channel network (Fig. 3). Averaged over all dynamically wet nodes, the model achieved an NSE of 0.9088, an R^2 of 0.9142, and a KGE of 0.8565. These values indicate strong linear agreement and high predictive skill relative to the reference hydrodynamic simulation. The KGE value further suggests that the model preserves not only correlation but also the variability and bias characteristics of the simulated water-depth field.

The spatial metric distributions reveal that predictive performance is not uniform across the domain. Higher R^2, NSE, and KGE values are mainly concentrated along the primary channel network and near-channel regions, where flood dynamics are persistent and governed by well-defined flow pathways. In many of these areas, node-level R^2 and NSE values approach 1.0, indicating close agreement between the PI-STGCN predictions and the hydrodynamic simulation. In contrast, lower metric values occur at scattered locations, particularly around channel branches, flood-front transition zones, and marginal inundation areas. These regions are more difficult to predict because small errors in flood arrival time or water depth can produce large changes in node-level metrics when the observed water-depth variation is limited.



(a)NSE (b)R^2 (c)KGE

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Figure 3. Spatial distribution of node-level prediction metrics for the 100a24h validation event. The spatial maps show prediction skill across dynamically wet nodes.

Because the river and floodplain are initially dry for the 100a24h event, the model can be evaluated under the cold-start autoregressive mode. In this mode, the initial hydraulic states, including water depth and momentum components, are set to zero, and the model recursively predicts subsequent flood evolution using rainfall forcing, static attributes, graph connectivity, and edge features.

To examine the ability of PI-STGCN to reproduce the temporal evolution of flood inundation, we compare the basin-wide maximum water depth and the spatial variability of water depth during the 24 h simulation period (Fig. 4). The upper panel shows the temporal evolution of the maximum water depth predicted by PI-STGCN and simulated by the hydrodynamic model. Both curves remain close to zero during the first 14 h, indicating limited runoff generation and inundation under relatively weak early rainfall. As rainfall intensifies around hours 15–16, both models capture the rapid onset of inundation and the sharp increase in maximum water depth. The maximum depth then rises quickly to approximately 19–20 m and enters a high-water-level stage after hour 18. PI-STGCN closely follows the hydrodynamic reference throughout this transition, although it slightly underestimates the maximum depth during parts of the rising and peak stages.

The lower panel shows rainfall intensity together with the interquartile range (Q1–Q3) of water depth across wet nodes. The water-depth range remains near zero before the flood onset and increases rapidly after hour 15, consistent with the rainfall peak and the expansion of inundated areas. The predicted and simulated box ranges show similar temporal evolution, indicating that PI-STGCN reproduces not only the timing and magnitude of maximum inundation but also the development of spatial variability in flood depths. The largest discrepancies occur during the rapid flood-expansion and peak stages, when the predicted box range is generally slightly smaller than that of the hydrodynamic simulation. This suggests a mild underestimation of the spatial spread of water depths, but the overall temporal pattern remains well captured.

These results demonstrate that PI-STGCN can simulate the complete flood-development process from dry initial conditions, including flood onset, rapid inundation growth, the high-water-level stage, and subsequent stabilization. The agreement between the predicted and simulated temporal trajectories indicates that the model has robust cold-start autoregressive prediction capability under the 100a24h extreme-rainfall scenario.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

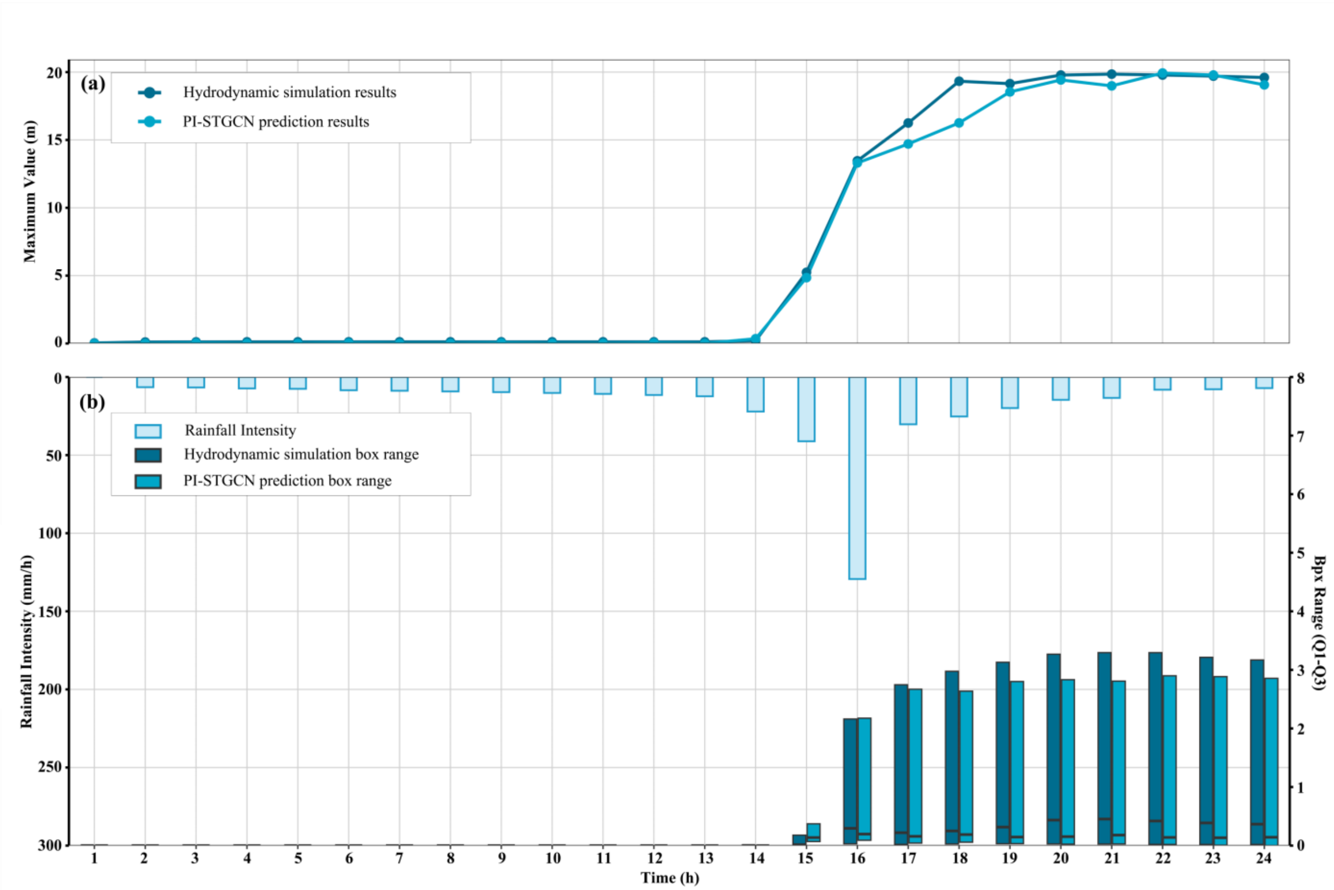


Figure 4. Temporal evolution of flood inundation for the 100a24h validation event under cold-start autoregressive prediction. (a) compares the maximum water depth predicted by PI-STGCN and simulated by the hydrodynamic model. (b) shows rainfall intensity together with the interquartile range (Q1–Q3) of water depth across wet nodes. The results demonstrate that PI-STGCN accurately reproduces both the temporal evolution and spatial variability of flood inundation under an unseen extreme-rainfall scenario.

4.2 Temporal and Spatial Flood Dynamics

To further evaluate the temporal prediction capability, we compared water-depth time series between the hydrodynamic simulation results and the PI-STGCN predictions. In addition to the 100a24h validation event, we used three 24 h segments, namely Event 237 (0−24 h), Event 237 (24−48 h), and Event 237 (48−72 h), to examine extrapolation performance under more realistic rainfall-runoff conditions.

The warm-start autoregressive mode requires an initial historical sequence of T = 8 time steps to characterize the current hydraulic state of the river system. In this study, these 8 time steps correspond to approximately 2 h of hydraulic evolution. Therefore, for the three Event 237 segments, the effective autoregressive prediction starts after approximately 2 h, 26 h, and 50 h, respectively. This setting allows the model to obtain the initial water-depth and momentum conditions before recursively predicting subsequent flood evolution.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

The spatial comparison between the hydrodynamic simulation and PI-STGCN-predicted flood inundation depths at selected time steps demonstrates that the model successfully captures the major inundation dynamics of the event (Fig. 5). The reference maps show the development of inundation along the river network as rainfall-driven runoff accumulates and propagates downstream. The PI-STGCN predictions reproduce the main spatial structure of the inundated channel system and capture the expansion of high-depth regions during the flood process. The predicted inundation pattern remains consistent with the hydrodynamic simulation throughout the autoregressive rollout, indicating that the model can learn flood-routing behavior over the triangular graph structure.

The spatial error distributions further show that most prediction errors are localized rather than widespread across the domain. Larger discrepancies mainly occur near channel branches, flood-front transition zones, and regions with rapid water-depth changes. These areas are inherently difficult to predict because small timing errors in flood-wave arrival can lead to visible spatial differences between the reference and predicted depth maps. Nevertheless, the overall channel-scale inundation pattern is preserved, suggesting that PI-STGCN can maintain spatial coherence during cold-start autoregressive prediction.

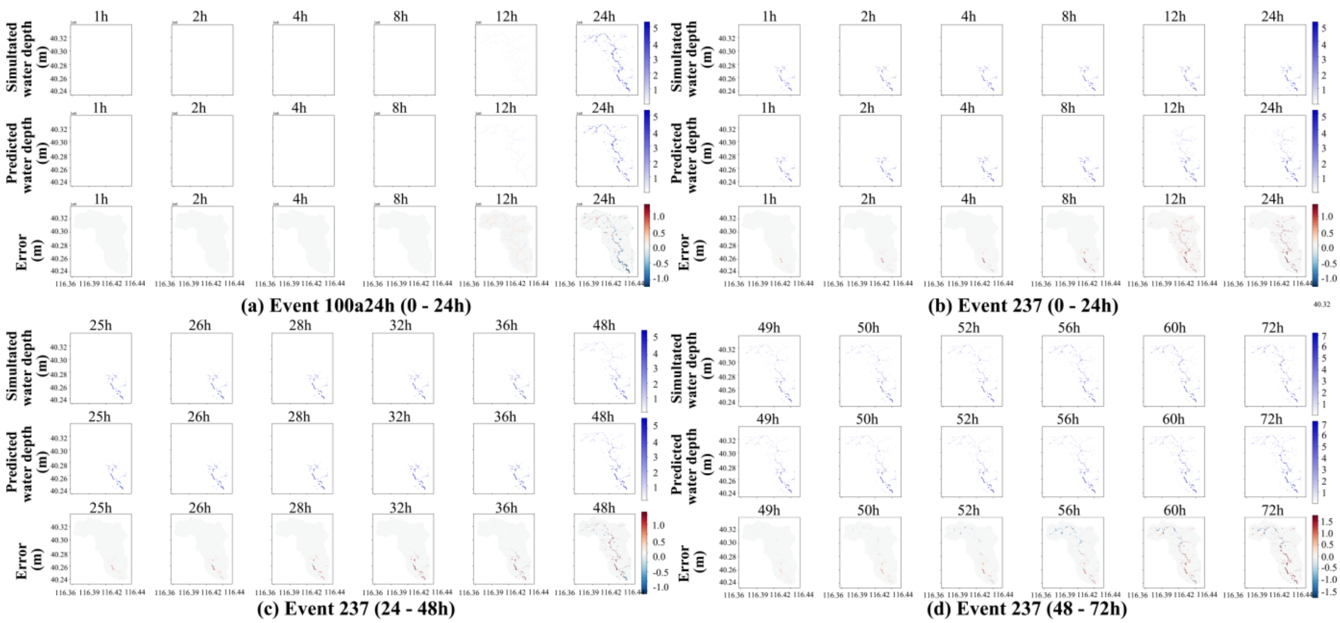


Figure 5. Spatial comparison of reference and predicted flood inundation depths. (a) The 100a24h validation event under cold-start autoregressive prediction. (b–d) The three Event 237 extrapolation segments under warm-start autoregressive prediction, corresponding to 237(0–24 h), 237(24–48 h), and 237(48–72 h), respectively.

The temporal water-depth series at representative triangular control volumes provide additional insight into the model's ability to reproduce local flood dynamics (Fig. 6). The selected nodes represent different hydraulic locations within the river network. Specifically, Node 1555 is located in the upstream reach of the main channel, Node 4444 in the midstream reach,

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Node 6404 in the downstream reach, and Node 12411 in the midstream section of a secondary branch. This selection allows model performance to be evaluated across different flow conditions and propagation stages within the watershed.

The predicted hydrographs generally follow the reference hydrodynamic simulation, including the dry initial period, rapid rising limb, peak water depth, and post-peak recession behavior. Quantitatively, the model shows strong agreement with the reference results at all four locations. For the upstream Node 1555, the prediction yields NSE = 0.972 and RMSE = 0.3327. For the midstream Node 4444, the model achieves NSE = 0.990 and RMSE = 0.8905. At the downstream Node 6404, NSE reaches 0.993 with RMSE = 0.7444. For Node 12411 in the secondary branch, NSE is 0.993 and RMSE is 0.1583. These high NSE values and relatively low RMSE values indicate that PI-STGCN can accurately reproduce local water-depth evolution across different parts of the river system.

Some discrepancies are observed during the rising and peak stages. At several nodes, PI-STGCN slightly advances or delays flood arrival, producing short-duration positive or negative errors. Peak-depth differences are also visible at some locations, particularly where the water depth increases rapidly within a short time. These deviations are expected in cold-start autoregressive prediction because small errors from earlier steps can propagate into later predictions. However, the overall hydrograph shapes remain close to the reference results, demonstrating the stability of the model during recursive rollout.

The temporal comparison further supports the effectiveness of the residual prediction strategy and the autoregressive training procedure. By predicting hydraulic-state increments and recursively updating the input sequence, the model can simulate continuous flood development from initially dry conditions. The strong agreement between predicted and reference hydrographs across upstream, midstream, downstream and tributary locations suggests that PI-STGCN has learned the dominant rainfall-runoff response and flood-propagation behavior for the 100a24h event.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

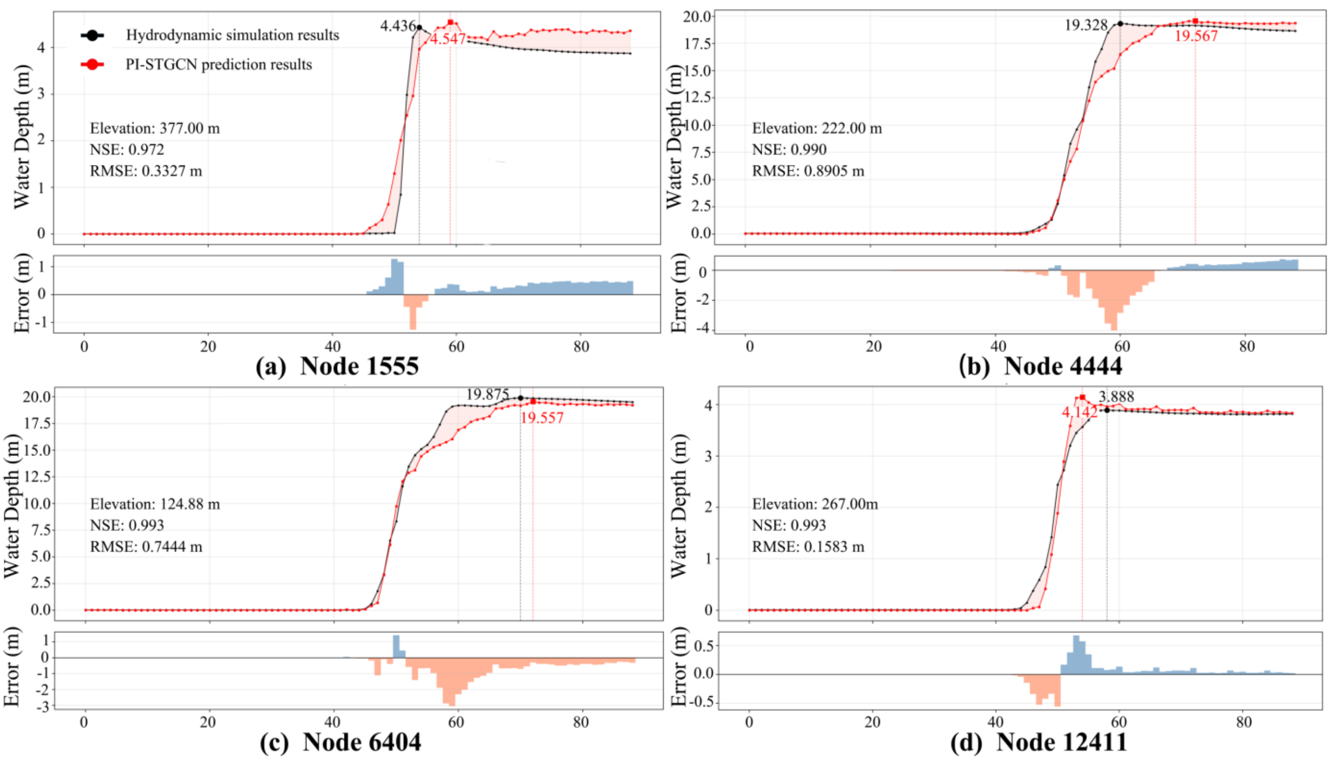


Figure 6. Temporal comparison of predicted and reference water depths at representative nodes. Black curves represent hydrodynamic simulation results, and red curves represent PI-STGCN predictions. The lower panels show prediction errors over time. The selected nodes represent different hydraulic locations within the river network: (a) Node 1555 in the upstream reach of the main channel, (b) Node 4444 in the midstream reach of the main channel, (c) Node 6404 in the downstream reach of the main channel, and (d) Node 12411 in the midstream section of a secondary branch.

4.3 Ablation Experiments and Physical Consistency Diagnostics

To examine the contribution of each major component in PI-STGCN, we conducted a series of ablation experiments. The full PI-STGCN model was used as the baseline, and five degraded variants were constructed by removing one key module at a time. A1 denotes the model without the edge-conditioned spatial graph convolution module, A2 denotes the model without the rainfall encoder module, A3 denotes the model without the causal temporal convolution module, A4 denotes the model without the physics-informed regularization module, and A5 denotes the model without the residual prediction module. These variants were used to evaluate the contribution of the main architectural and training components of the proposed framework.

The ablation results for the 100a24h validation event are summarized in Fig. 7. The baseline PI-STGCN achieved R^2 = 0.9142, NSE = 0.9088, KGE = 0.8565, MAE = 0.3390, and RMSE = 0.7892. These results provide the reference performance for assessing the degraded variants. Overall, the baseline model achieved the best or near-best performance

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

across all five metrics, indicating that the combined use of graph-based spatial learning, rainfall encoding, temporal convolution, physics-informed regularization, and residual prediction contributes to accurate flood inundation simulation.

Removing the edge-conditioned spatial graph convolution module caused the largest degradation among the normally plotted variants. In A1, R^2 decreased to 0.7889, NSE decreased to 0.7714, and KGE decreased to 0.7029, while MAE and RMSE increased to 0.5973 and 1.2497, respectively. This substantial deterioration indicates that graph-based spatial interactions among neighboring triangular control volumes are essential for representing flood propagation over the irregular mesh. Without this module, the model loses an important mechanism for learning hydraulic exchange between physically connected cells.

Removing the rainfall encoder also reduced model performance. A2 achieved R^2 = 0.8721, NSE = 0.8701, MAE = 0.4360, and RMSE = 0.9418, all worse than the baseline. This result indicates that dedicated rainfall-feature extraction improves the representation of rainfall–runoff response and flood generation. However, A2 produced a slightly higher KGE value of 0.8735 than the baseline, suggesting that removing the rainfall encoder may improve some aspects of hydrograph distribution matching while degrading the overall accuracy and error-based metrics. Therefore, the rainfall encoder is beneficial for balanced predictive performance across multiple evaluation criteria.

Eliminating the causal temporal convolution module produced a moderate decline in performance. A3 achieved R^2 = 0.9034, NSE = 0.9007, KGE = 0.8234, MAE = 0.3674, and RMSE = 0.8234. These values remain close to the baseline but are consistently worse across all metrics. This suggests that causal temporal convolution contributes to capturing short-term hydraulic memory and delayed flood responses, although the model retains considerable predictive capability through the remaining graph, rainfall, and residual-prediction components.

The variant without physics-informed regularization, A4, also showed slightly weaker performance than the full PI-STGCN. A4 achieved R^2 = 0.9140, NSE = 0.9029, KGE = 0.8126, MAE = 0.3412, and RMSE = 0.8146. The differences in R^2 and MAE were small, but the reductions in NSE and KGE and the increase in RMSE indicate that physics-informed regularization provides additional benefit even for the 100a24h validation event. This result suggests that the shallow-water-equation residuals and water-volume consistency terms help regularize the learned flood dynamics rather than merely adding complexity to the training objective.

The most severe degradation occurred when the residual prediction module was removed. A5 retained only R^2 = 0.4272 in the plotted results, while the remaining metrics became extreme and were omitted from the corresponding panels for readability. This failure indicates that direct prediction of future hydraulic states is unstable for long-horizon autoregressive flood simulation. By predicting hydraulic-state increments rather than absolute future states, the residual prediction strategy reduces the learning difficulty, improves temporal continuity, and limits error accumulation during rollout.

Overall, the ablation experiments show that the edge-conditioned spatial graph convolution and residual prediction modules are the most critical components of PI-STGCN. The rainfall encoder, causal temporal convolution, and physics-informed regularization further improve model robustness and balanced predictive skill. These results confirm that the final PI-

STGCN performance arises from the combined contribution of mesh-aware spatial learning, rainfall-response representation, temporal dependency modeling, physically informed regularization, and stable autoregressive state updating.

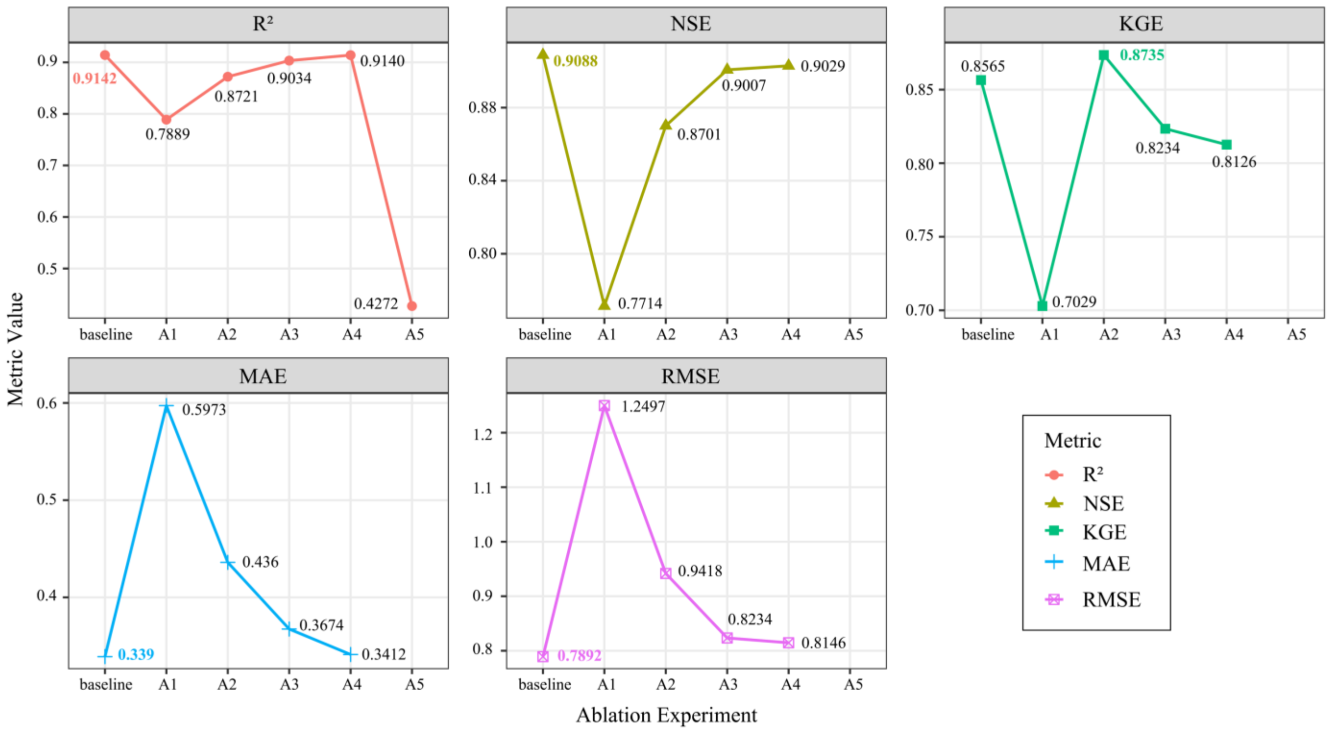


Figure 7. Ablation analysis of the proposed PI-STGCN components. The baseline denotes the full PI-STGCN model. A1: without spatial graph convolution; A2: without rainfall encoder; A3: without causal temporal convolution; A4: without physics-informed regularization; A5: without residual prediction. A5 is omitted from the plotted lines due to extreme values: NSE = -205.4734, KGE = -14.8044, MAE = 12.3953, RMSE = 37.5544.

To further evaluate the role of physics-informed regularization, we compared the full PI-STGCN model with A4, the variant without physics-informed constraints, using post-processed shallow-water-equation residual diagnostics. The comparison was conducted for the 100a24h validation event and the three Event 237 extrapolation segments. Unlike the predictive-skill metrics, these diagnostics directly assess whether the predicted hydraulic fields satisfy local continuity and momentum consistency (Fig. 8).

Figure 8a–c compares the mean residuals of the continuity equation, x-momentum equation, and y-momentum equation. The full PI-STGCN generally reduces the momentum-equation residuals relative to A4 across all evaluated events. This reduction is particularly evident for both x- and y-momentum residuals, indicating that physics-informed regularization improves the consistency of the predicted momentum evolution. In contrast, the continuity residual shows a more mixed response. The full model is comparable to or slightly better than A4 in some cases, but it does not uniformly reduce the continuity residual across all Event 237 segments. This suggests that the physics-informed constraints primarily improve local momentum consistency rather than enforcing uniformly lower mass residuals in every event.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Figure 8d–f further decomposes the residuals into their major component terms. For the continuity equation, the flux-divergence term dominates the residual magnitude, whereas the rainfall source term is several orders of magnitude smaller after conversion to SI units and averaging over space and time. For the momentum equations, the topographic source term is the dominant component, reflecting the strong influence of terrain gradients in the mountainous catchment. The friction and temporal terms are generally smaller, although they still contribute to the overall residual balance. The component-level comparison shows that the full PI-STGCN tends to reduce several dominant momentum-residual components compared with A4, which is consistent with the lower total momentum residuals observed in Fig. 8b and c.

These results indicate that physics-informed regularization improves the hydrodynamic plausibility of PI-STGCN predictions, especially in terms of momentum-equation consistency during autoregressive rollout. However, the improvement is not uniform for the continuity equation, and the model should therefore be interpreted as a physics-regularized surrogate rather than a strictly conservative numerical solver. Overall, the residual diagnostics support the use of shallow-water-equation residuals and water-volume consistency terms as soft constraints for improving local physical consistency under both design-rainfall validation and real-event extrapolation conditions.

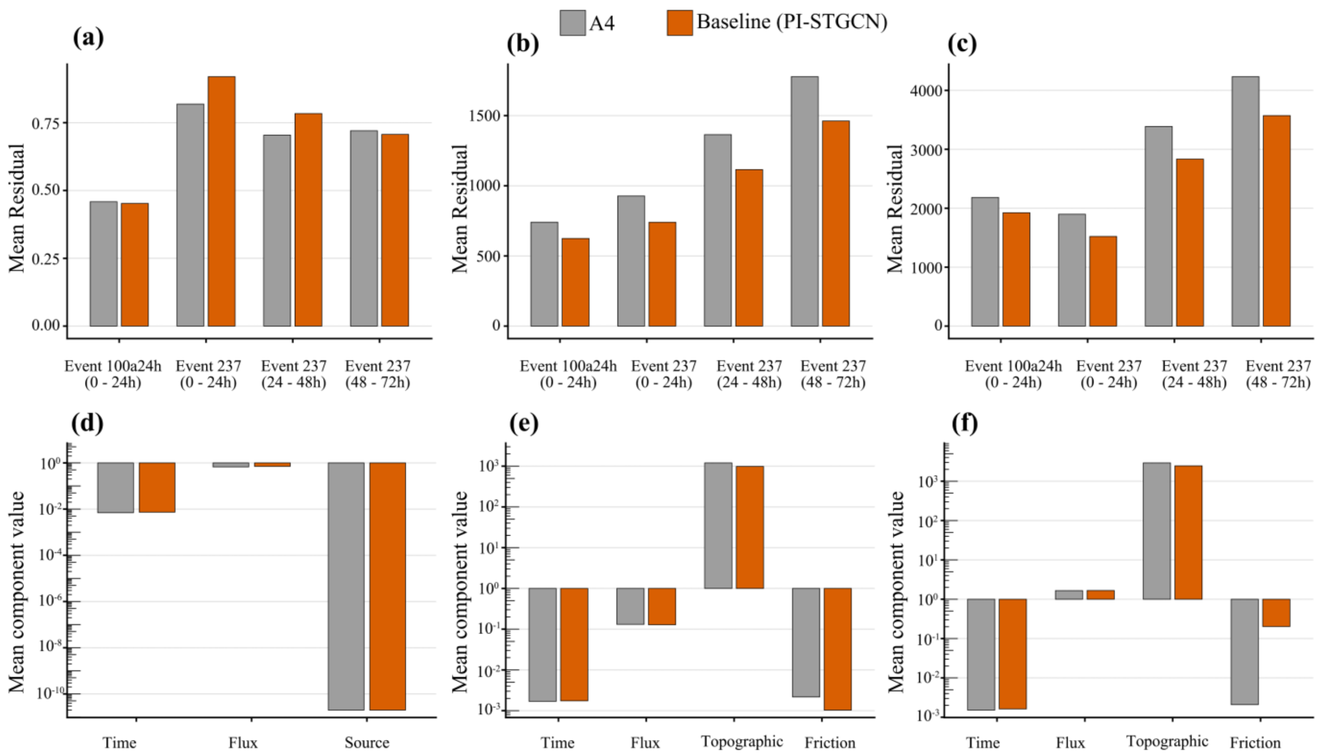


Figure 8. Physical consistency diagnostics of PI-STGCN predictions. (a–c) Mean residuals of the continuity equation, x-momentum equation, and y-momentum equation for the full PI-STGCN and A4, the variant without physics-informed regularization, across the 100a24h validation event and the three Event 237 extrapolation segments. (d–f) Decomposition of the corresponding residuals into major component terms, averaged over the evaluated events and shown on a logarithmic scale.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Overall, the ablation experiments show that each module contributes to model performance, while spatial graph convolution and residual prediction are particularly critical for predictive accuracy and autoregressive stability. The physics-informed component provides additional regularization by improving local shallow-water-equation consistency, especially for momentum evolution. Although the continuity residual is not uniformly reduced across all events, the lower momentum residuals indicate that the full PI-STGCN produces more hydrodynamically plausible predictions than the purely data-driven A4 variant during long-horizon rollout.

4.4 Computational Efficiency

A major motivation for developing PI-STGCN is to provide faster flood inundation prediction than conventional hydrodynamic simulation. To assess computational efficiency, we evaluated both training time and inference time. On a computer equipped with an Intel Xeon w5-2465X processor, 128 GB RAM, and an NVIDIA RTX 5880 Ada Generation GPU, the PI-STGCN model took 1 h 37 min 10 s to train. Once trained, the neural network does not solve the SWE iteratively at each time step. Instead, it directly predicts future hydraulic states through forward propagation across the graph convolution, temporal convolution, and output layers.

We compared the runtime of the hydrodynamic simulation and PI-STGCN prediction for the 100a24h design event and the three 24 h segments of Event 237 (Tab. 3). The hydrodynamic simulation required 63–105 s depending on the event segment, whereas PI-STGCN required only approximately 1.06–1.08 s. The corresponding speed-up ratios ranged from 58.33 to 98.13. For the 100a24h event, the hydrodynamic simulation required 1 min 3 s, while PI-STGCN completed the prediction in 1.08 s, corresponding to a speed-up ratio of 58.33. For Event 237, the acceleration was more pronounced, with speed-up ratios of 98.13, 87.74, and 79.63 for the three consecutive 24 h segments.

These results demonstrate that PI-STGCN provides a substantial computational advantage after training. This efficiency is particularly important for operational flood forecasting, where multiple rainfall scenarios or ensemble forecasts may need to be evaluated within a short time. Compared with a full hydrodynamic solver, the trained PI-STGCN avoids repeated numerical iterations and rapidly generates flood inundation predictions while preserving the main spatiotemporal flood dynamics. The reported speed-up ratios reflect the implemented runtime under the specified hardware configuration and should be interpreted as an operational inference-time comparison.

Table 3. Computational time comparison between hydrodynamic simulation and PI-STGCN prediction.

Event	Hydrodynamic simulation time	PI-STGCN prediction time	Speed-up ratio
100a24h	1 min 3 s	1.08 s	58.33
237 (0–24 h)	1 min 45 s	1.07 s	98.13
237 (24–48 h)	1 min 33 s	1.06 s	87.74
237 (48–72 h)	1 min 26 s	1.08 s	79.63

5 Discussion

5.1 Accuracy of PI-STGCN in Reproducing Flash-Flood Inundation Dynamics

PI-STGCN accurately reproduces the main spatial and temporal characteristics of flash-flood inundation in the Taoyukou watershed. The 100a24h validation event represents a rainfall-magnitude extrapolation test, because the model was trained using the 10a24h, 20a24h, and 50a24h design storms and evaluated on a more extreme unseen rainfall scenario. For this event, PI-STGCN achieved $R^2 = 0.9142$, NSE = 0.9088, KGE = 0.8565, MAE = 0.3390, and RMSE = 0.7892, indicating strong agreement with the reference hydrodynamic simulation. Spatially, the model performs best along the main river network and near-channel regions, where inundation is persistent and hydraulic pathways are well defined. Lower metric values mainly occur near channel branches, marginal inundation zones, and flood-front transition areas, where small errors in flood-arrival timing or local water-depth variation can strongly affect node-level metrics. The error pattern therefore suggests that the main prediction difficulty lies in wetting–drying transitions, terrain-controlled flow exchange, and channel–floodplain interactions, rather than in domain-wide prediction failure.

The cold-start autoregressive results further show that PI-STGCN can simulate flood development from initially dry conditions. Using rainfall forcing, static graph attributes, edge connectivity, and learned hydraulic dynamics, the model captures the dry initial stage, rapid inundation growth, high-water-level period, and subsequent stabilization. Although it slightly underestimates the spatial spread of water depths during parts of the rapid expansion and peak stages, the basin-wide maximum depth, water-depth variability, and representative-node hydrographs remain close to the hydrodynamic reference. The ablation experiments explain the source of this accuracy: removing edge-conditioned spatial graph convolution substantially degrades performance, confirming the importance of hydraulic information exchange among neighboring triangular control volumes; removing the rainfall encoder or causal temporal convolution also weakens predictive skill; and removing residual prediction leads to unstable long-horizon forecasting. These results indicate that PI-STGCN's accuracy arises from the combined effects of mesh-aware spatial learning, rainfall-response representation, temporal dependency modeling, and residual state updating.

5.2 Interpretable Learning Through Graph Representation and Physical Constraints

PI-STGCN is more physically interpretable than a purely black-box neural network because its learning structure is tied to the triangular hydrodynamic mesh. Each graph node corresponds to a triangular control volume, and each directed edge corresponds to a shared hydraulic interface between adjacent cells. Therefore, message passing in the edge-conditioned graph convolution can be interpreted as information exchange between hydraulically connected control volumes. Unlike binary adjacency, the directed edge attributes encode shared-interface length, edge geometry, normal-vector components, elevation difference, channel proximity, and directional flow priors, allowing the model to distinguish different hydraulic connections. The independent rainfall encoder further improves interpretability by treating rainfall as an external forcing

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

rather than simply concatenating it with hydraulic states. This enables the model to learn both instantaneous rainfall effects and antecedent rainfall information, which is consistent with the rainfall–runoff response of flash floods.

The physics-informed terms provide an additional interpretive link between neural prediction and hydrodynamic principles. In this study, shallow-water-equation residuals and water-volume consistency are used as soft regularization terms, not as strict conservative numerical constraints. The residual diagnostics support this interpretation: compared with A4, the full model generally reduces x- and y-momentum residuals across the evaluated events, indicating improved local momentum consistency during autoregressive rollout. However, the continuity residual is event-dependent and is not uniformly reduced, so the physical benefit should not be overstated as strict mass conservation. The residual decomposition further shows that continuity imbalance is mainly controlled by flux divergence, whereas momentum residuals are dominated by topographic source terms, reflecting the strong terrain control of the mountainous catchment. These results suggest that the dominant residual patterns are physically interpretable and related to routing, terrain forcing, and local hydraulic gradients rather than arbitrary numerical error.

5.3 Generalization Beyond Training Rainfall Scenarios

Model generalization was evaluated through two extrapolation settings. The first was rainfall-magnitude extrapolation, in which PI-STGCN was trained on the 10a24h, 20a24h, and 50a24h design storms and evaluated on the unseen 100a24h design storm. The strong performance for this event, together with the spatial metric distributions, maximum-depth evolution, inundation maps, and representative-node hydrographs, indicates that the model learned transferable relationships among rainfall forcing, terrain attributes, channel connectivity, and hydraulic response within the design-storm setting. The second setting was event-type transfer from design rainfall scenarios to the real rainfall-driven Event 237. This event lasted approximately 72 h and was divided into three 24 h segments. In each segment, the first 8 time steps, corresponding to approximately 2 h, were used as the warm-start hydraulic sequence before autoregressive prediction. Compared with the 100a24h design event, Event 237 represents a more demanding test because real rainfall events contain more irregular temporal structure, different antecedent hydraulic states, and wetting–drying behavior that are not fully represented by the design-storm training set.

Event 237 provides a stress test rather than definitive evidence of broad generalization. Its results show that PI-STGCN maintains reasonable transferability under real rainfall forcing, while the role of physics-informed regularization is more clearly reflected in physical consistency than in conventional predictive metrics. Residual diagnostics indicate that the full PI-STGCN consistently reduces both x- and y-momentum residuals relative to A4 across all evaluated events, demonstrating improved consistency of momentum evolution during autoregressive rollout. In contrast, the continuity residual exhibits a more mixed response and is not uniformly reduced across all Event 237 segments. Residual decomposition further shows that continuity imbalance is dominated by flux divergence, whereas momentum residuals are primarily controlled by topographic source terms, reflecting strong terrain forcing in the mountainous catchment. The full model tends to reduce several dominant momentum-residual components compared with A4, which is consistent with the lower total momentum

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

residuals. These results indicate that physics-informed regularization improves hydrodynamic plausibility, particularly local momentum-equation consistency, but does not enforce uniformly lower mass residuals. Therefore, PI-STGCN should be interpreted as a physics-regularized surrogate with reasonable real-event transferability, rather than a strictly conservative numerical solver.

5.4 Computational Advantages for Operational Applications

PI-STGCN provides a substantial inference-efficiency advantage after training. The offline training process required 1 h 37 min 10 s on the specified hardware platform, including back-propagation, scheduled sampling, pushforward rollout, and physics-informed residual calculation. This cost is incurred only during model development. Once trained, PI-STGCN predicts future hydraulic states through neural-network forward propagation rather than repeatedly solving the shallow-water equations. For the evaluated events, the reference hydrodynamic simulations required 63–105 s, whereas PI-STGCN required only 1.06–1.08 s, corresponding to speed-up ratios of 58.33–98.13. These results show that the trained model can substantially reduce runtime while preserving the main spatiotemporal flood dynamics represented by the reference simulations.

This computational advantage is relevant for near-real-time flood forecasting support, rapid scenario screening, and ensemble-based risk assessment, where multiple rainfall scenarios may need to be evaluated within a limited decision window. However, the reported speed-up ratios should be interpreted as runtime comparisons under the specified hardware, preprocessing, and implementation settings, rather than hardware-independent algorithmic benchmarks. PI-STGCN does not replace full hydrodynamic simulation for applications requiring process diagnosis, model calibration, uncertainty attribution, or strictly conservative numerical solutions. Its practical value lies in providing a fast physics-regularized surrogate that balances predictive accuracy, local physical consistency, and computational efficiency for rapid inundation prediction in mountainous catchments.

6 Conclusions

This study proposed PI-STGCN, a physics-informed spatiotemporal graph convolutional network for rapid flash-flood forecasting and inundation simulation on irregular triangular control-volume meshes. The model represents the hydrodynamic domain as a directed graph and integrates terrain, land-surface, channel, rainfall, hydraulic-state, and edge-connectivity information. Shallow-water-equation residuals, water-volume consistency, and near-channel boundary constraints are incorporated as soft regularization terms to improve hydrodynamic plausibility. Results for the unseen 100a24h design storm demonstrate strong rainfall-magnitude extrapolation performance. PI-STGCN achieved $R^2 = 0.9142$, NSE = 0.9088, KGE = 0.8565, MAE = 0.3390, and RMSE = 0.7892. Cold-start autoregressive experiments show that the model can reproduce flood onset, rapid inundation growth, peak-stage dynamics, and spatial propagation from initially dry conditions. Ablation experiments further indicate that spatial graph convolution and residual prediction are critical for

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

predictive accuracy and rollout stability. Physics-informed regularization improves local hydrodynamic plausibility, particularly momentum-equation consistency. However, its effect on conventional predictive metrics is event- and metric-dependent. The trained model required only 1.06–1.08 s for inference, compared with 63–105 s for the reference hydrodynamic simulations. These results indicate that PI-STGCN provides an efficient physics-regularized surrogate for rapid flood prediction. Future work should include more real flood events, observational validation, uncertainty analysis, and multi-watershed testing.

Appendix A: Training Diagnostics

We examined the training, validation, data-driven, physics-informed, and rollout losses over 300 epochs to assess the convergence behavior of PI-STGCN (Fig. A1). The total training loss decreased rapidly during the early supervised stage and reached its minimum at epoch 58, with a loss value of 0.82514. After curriculum-learning components were activated, the training loss increased and fluctuated before gradually decreasing again. This non-monotonic behavior is expected because scheduled sampling, pushforward rollout training, and physics-informed constraints progressively make the optimization task more difficult. Therefore, the training loss should not be interpreted as a conventional single-objective loss with strictly monotonic convergence.

In contrast, the validation loss shows stable convergence. It decreases steadily during training and reaches its minimum at epoch 238, with a validation loss of 0.04048. After this point, the validation curve remains nearly unchanged, indicating that the model does not show obvious overfitting. The checkpoint corresponding to the best validation epoch was therefore selected for subsequent evaluation. The data loss also decreases smoothly throughout training, demonstrating that the model progressively improves its fit to the reference hydrodynamic simulation outputs.

Figure A1c shows the evolution of individual physics-informed loss components on a logarithmic scale. The volume-consistency loss decreases substantially during training, while the near-channel boundary-condition loss also declines steadily, indicating improved hydraulic-state prediction in channel-dominated regions. The SWE residual loss exhibits intermittent spikes, which are expected because residual evaluation is sensitive to local wetting–drying transitions, momentum changes, and rollout-generated states. Nevertheless, its overall magnitude decreases and stabilizes as training proceeds. These trends suggest that the model improves both data-fitting accuracy and local physical consistency during training.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

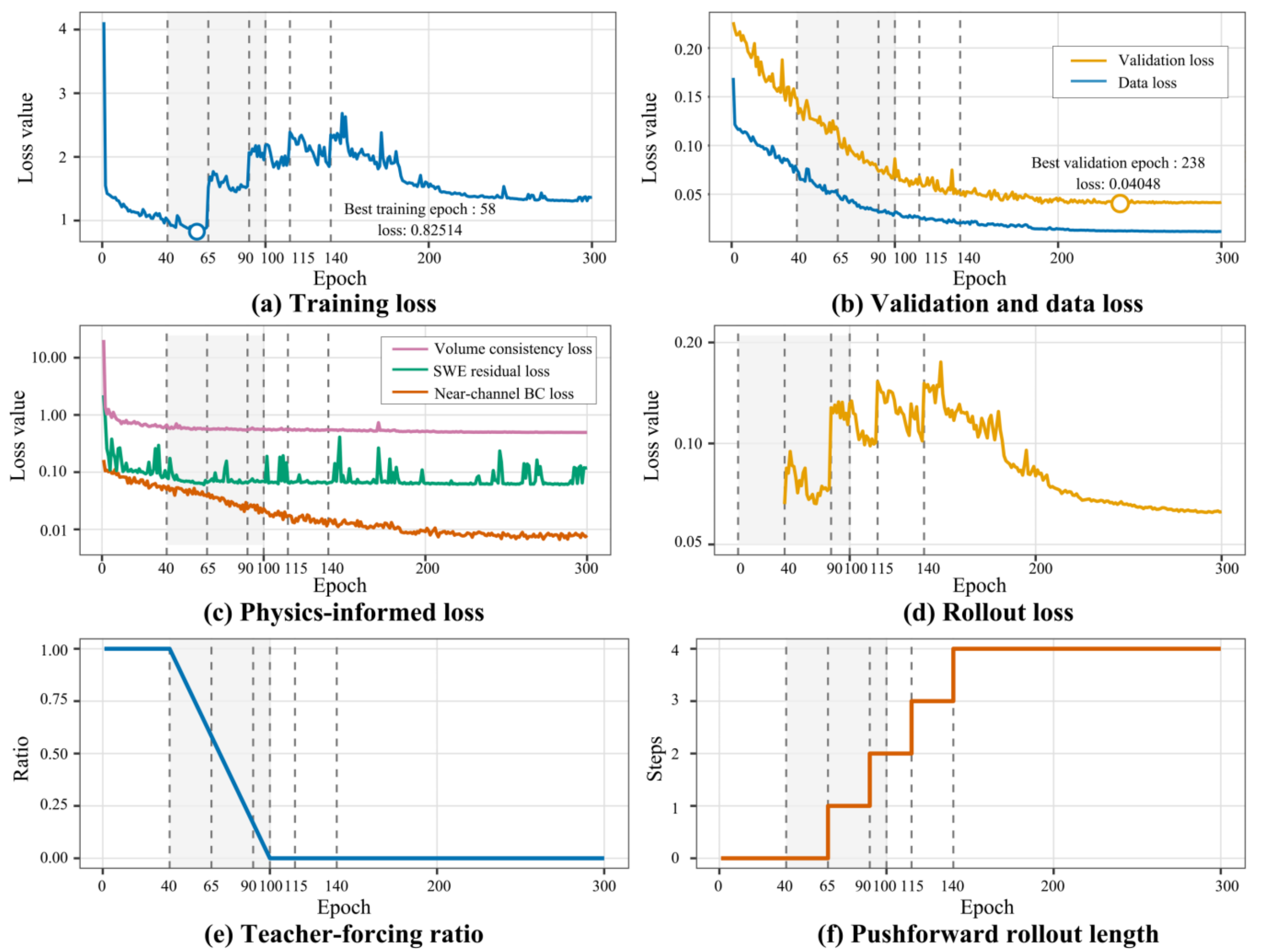


Figure A1. Overall convergence during PI-STGCN training. (a) Total training loss. (b) Validation loss and data loss. (c) Physics-informed loss components, including volume-consistency loss, shallow-water-equation residual loss, and near-channel boundary-condition loss, shown on a logarithmic scale. (d) Rollout loss during pushforward training. (e) Teacher-forcing ratio. (f) Pushforward rollout length.

The dynamically scheduled weights for physics-informed loss terms are defined as

$$w(e) = \begin{cases} w_{init}, & e < e_w \\ w_{init} + \frac{e-e_w}{e_r}(w_{final} - w_{init}), & e_w \le e < e_w + e_r \\ w_{final}, & e \ge e_w + e_r \end{cases} \tag{A1}$$

where e is the current epoch, e_w is the warm-up epoch, e_r is the ramp-up duration, and w_{init} and w_{final} are the initial and final weights, respectively. This schedule prevents the physics-informed terms from dominating early training before the model has learned a reasonable hydraulic-state approximation.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Scheduled sampling is used to reduce the mismatch between training and autoregressive inference. In the early stage, the model uses teacher forcing, in which reference hydraulic states are used to construct the next input window. The teacher-forcing ratio remains 1.0 before epoch 40 and then decreases linearly to 0 by epoch 100:

$$\rho(e) = \begin{cases} 1, & e < e_s \\ 1 - \frac{e-e_s}{e_e-e_s}, & e_s \le e < e_e \\ 0, & e \ge e_e \end{cases} \tag{A2}$$

where e_s and e_e denote the start and end epochs of scheduled sampling, respectively. After epoch 100, the model operates fully in autoregressive mode during training.

Pushforward rollout training is further introduced to improve long-horizon stability. Instead of optimizing only one-step predictions, the model is unrolled for multiple future steps, so that prediction errors can be propagated and corrected during training. In this study, the rollout length is increased stepwise after the warm-up stage and reaches the maximum value of four steps at epoch 140:

$$K(e) = \left\lceil K_{max} \cdot min\left(1, \frac{e-e_r}{E_r}\right)\right\rceil, \tag{A3}$$

where $K(e)$ is the rollout length at epoch e, K_{max} is the maximum rollout length, e_r is the epoch at which rollout training starts, and E_r is the rollout ramp-up period. The rollout loss becomes active after rollout training begins. Its temporary increase during the middle stage reflects the increasing rollout difficulty as the model is exposed to longer autoregressive sequences. After the rollout length reaches its maximum, the rollout loss gradually decreases and stabilizes, indicating improved recursive prediction capability.

Overall, the training diagnostics demonstrate stable convergence under a progressive curriculum-learning strategy. Although the total training loss becomes non-monotonic after scheduled sampling and rollout training are activated, the validation loss, data loss, near-channel boundary-condition loss, and volume-consistency loss all show clear convergence. The best validation checkpoint occurs late in training, after the model has adapted to fully autoregressive prediction and multi-step rollout. These results support the stability of the training procedure and the effectiveness of combining data-driven fitting, physics-informed regularization, scheduled sampling, and pushforward rollout training.

Appendix B: Detailed network operations

For a directed edge $j \rightarrow i$, the message from source node j to target node i is defined as

$$m_{j\rightarrow i} = \phi_m([h_j, e_{ji}]), \tag{B1}$$

where h_j is the hidden representation of node j, e_{ji} is the directed edge feature vector, $[\cdot,\cdot]$ denotes concatenation, and $\phi_m(\cdot)$ is a multilayer perceptron. The incoming messages are aggregated by summation:

$$a_i = \sum_{j \in N(i)} m_{j\rightarrow i}, \tag{B2}$$

where $N(i)$ is the neighbor set of node i. The target-node representation is then updated by combining the original node representation with the aggregated message:

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

$$h'_i = \emptyset_u([h_i, a_i]), \tag{B3}$$

where $\phi_u(\cdot)$ is another multilayer perceptron.

This operation enables the model to learn spatial flood propagation through physically meaningful mesh interfaces. Multiple edge-conditioned graph convolution layers are stacked to capture higher-order spatial interactions, and residual connections are used to improve gradient propagation and preserve local node information. For each historical time step, the spatial graph convolution can be written as

$$H_t^s = SConv(H_t^f, A_{edge}, E), \tag{B4}$$

where H_t^f is the fused node representation at time t, A_{edge} is the directed edge index, and E is the edge feature matrix.

Rainfall is the primary external forcing of flash-flood generation (Sikorska et al., 2015). However, after unit conversion and normalization, rainfall intensity may have a numerical scale different from water depth and momentum variables. If rainfall is directly concatenated with hydraulic states, its contribution may be weakened during feature learning. To address this issue, PI-STGCN uses an independent rainfall encoder.

The rainfall input is first transformed by logarithmic amplification:

$$\tilde{R}_{i,t} = log(1 + \alpha R_{i,t}), \tag{B5}$$

where α is a scaling factor and $R_{i,t}$ is the rainfall intensity at node i at time t. This transformation enhances small rainfall signals while limiting excessive amplification of large rainfall values. The transformed rainfall is then passed through a pointwise multilayer perceptron:

$$r_{i,t} = \emptyset_r(\tilde{R}_{i,t}), \tag{B6}$$

To capture rainfall accumulation and lag effects, the rainfall embeddings are further processed by temporal convolution along the historical sequence:

$$H^r = RainEnc(R_{t-T+1:t}), \tag{B7}$$

where H^r denotes the rainfall hidden representation. This design allows the model to learn both instantaneous rainfall forcing and antecedent rainfall effects, which are important for simulating rainfall-runoff response in mountainous catchments.

After spatial graph convolution is applied to each historical time step, the model obtains a sequence of spatially informed node representations:

$$H_{t-T+1:t}^s = [H_{t-T+1}^s, H_{t-T+2}^s, \dots, H_t^s], \tag{B8}$$

The temporal evolution of these representations is modeled using causal temporal convolution. Causal convolution ensures that the hidden representation at a given time step depends only on the current and previous time steps, not on future information. This property is essential for flood forecasting because future hydraulic states must not be used when predicting subsequent inundation.

For each node, the spatial representation sequence is rearranged along the temporal dimension, and the temporal convolution module produces

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

$$H_i^t = TConv(H_i^s),\tag{B9}$$

The temporal module uses one-dimensional causal convolution to learn short-term hydraulic evolution from the historical sequence. This enables the model to represent delayed rainfall-runoff response, flood-wave propagation, and temporal dependence in water-depth and momentum evolution.

Code and data availability

The processed data and code supporting the findings of this study are available from the corresponding author upon reasonable request.

Author contributions

FW and ZQZ were involved in the conceptualization and methodology of the project. ZQZ was responsible for model development, with guidance from FW, QZ, FBS and NLG. ZQZ also ran the model simulations and analyzed the results under the supervision of SKF, ZHC and MFZ. SKF and ZHC provided the original model data. The original draft of the paper was prepared by ZQZ. All authors were involved in reviewing and editing the manuscript.

Competing interests

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

References

Albano, R., Limongi, C., Dal Sasso, S. F., Mancusi, L., and Adamowski, J.: Flood scenario spatio-temporal mapping via hydrological and hydrodynamic modelling and a remote sensing dataset: A case study of the Basento river (Southern Italy), International Journal of Disaster Risk Reduction, 111, 104758, https://doi.org/10.1016/j.ijdrr.2024.104758, 2024.

Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R.: Deep learning methods for flood mapping: a review of existing applications and future research directions, Hydrol. Earth Syst. Sci., 26, 4345–4378, https://doi.org/10.5194/hess-26-4345-2022, 2022.

Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R.: Rapid spatio-temporal flood modelling via hydraulics-based graph neural networks, Hydrol. Earth Syst. Sci., 27, 4227–4246, https://doi.org/10.5194/hess-27-4227-2023, 2023.

Brandstetter, J., Worrall, D., and Welling, M.: Message Passing Neural PDE Solvers, https://doi.org/10.48550/arXiv.2202.03376, 20 March 2023.

Fraehr, N., Wang, Q. J., Wu, W., and Nathan, R.: Supercharging hydrodynamic inundation models for instant flood insight, Nat Water, 1, 835–843, https://doi.org/10.1038/s44221-023-00132-2, 2023.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Gao, H., Sun, L., and Wang, J.-X.: PhyGeoNet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain, Journal of Computational Physics, 428, 110079, https://doi.org/10.1016/j.jcp.2020.110079, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

He, M., Jiang, S., Ren, L., Cui, H., Du, S., Zhu, Y., Qin, T., Yang, X., Fang, X., and Xu, C.-Y.: Exploring the performance and interpretability of hybrid hydrologic model coupling physical mechanisms and deep learning, Journal of Hydrology, 649, 132440, https://doi.org/10.1016/j.jhydrol.2024.132440, 2025.

Hu, X. and Song, L.: Hydrodynamic modeling of flash flood in mountain watersheds based on high-performance GPU computing, Nat Hazards, 91, 567–586, https://doi.org/10.1007/s11069-017-3141-7, 2018.

Kumar, V., Sharma, K., Caloiero, T., Mehta, D., and Singh, K.: Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances, Hydrology, 10, 141, https://doi.org/10.3390/hydrology10070141, 2023.

Li, W., Liu, C., Xu, Y., Niu, C., Li, R., Li, M., Hu, C., and Tian, L.: An interpretable hybrid deep learning model for flood forecasting based on Transformer and LSTM, Journal of Hydrology: Regional Studies, 54, 101873, https://doi.org/10.1016/j.ejrh.2024.101873, 2024.

Liao, Y., Wang, Z., Chen, X., and Lai, C.: Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model, Journal of Hydrology, 624, 129945, https://doi.org/10.1016/j.jhydrol.2023.129945, 2023.

Liu, J., Zhao, G., Xiong, J., and Kinouchi, T.: A Graph-Based Deep Learning Approach for Daily Flash Flood Susceptibility Modeling in China, Water Resources Research, 62, e2025WR041360, https://doi.org/10.1029/2025WR041360, 2026.

Lu, L., Zou, Y., Wang, J., Zou, S., Zhang, L., and Deng, X.: Unsupervised learning with physics informed graph networks for partial differential equations, Appl Intell, 55, 617, https://doi.org/10.1007/s10489-025-06479-1, 2025.

Ma, J., Gao, H., Xu, C., and Qi, S.: Characterization and formation mechanism of the catastrophic flash flood-debris flow hazard triggered by the July 2023 extreme rainstorm in Hantai Gully of Beijing, China, Landslides, 22, 877–893, https://doi.org/10.1007/s10346-024-02433-3, 2025.

Maity, R., Srivastava, A., Sarkar, S., and Khan, M. I.: Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning, Applied Computing and Geosciences, 24, 100206, https://doi.org/10.1016/j.acags.2024.100206, 2024.

Moghim, S., Gharehtoragh, M. A., and Safaie, A.: Performance of the flood models in different topographies, Journal of Hydrology, 620, 129446, https://doi.org/10.1016/j.jhydrol.2023.129446, 2023.

Ni, L., Wang, D., Singh, V. P., Wu, J., Wang, Y., Tao, Y., and Zhang, J.: Streamflow and rainfall forecasting by two long short-term memory-based models, Journal of Hydrology, 583, 124296, https://doi.org/10.1016/j.jhydrol.2019.124296, 2020.

Peng, J.-Z., Hua, Y., Aubry, N., Chen, Z.-H., Mei, M., and Wu, W.-T.: Data and physics-driven modeling for fluid flow with a physics-informed graph convolutional neural network, Ocean Engineering, 301, 117551, https://doi.org/10.1016/j.oceaneng.2024.117551, 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.

Rasmy, M., Sayama, T., and Koike, T.: Development of water and energy Budget-based Rainfall-Runoff-Inundation model (WEB-RRI) and its verification in the Kalu and Mundeni River Basins, Sri Lanka, Journal of Hydrology, 579, 124163, https://doi.org/10.1016/j.jhydrol.2019.124163, 2019.

Shao, P., Feng, J., Lu, J., and Tang, Z.: Data-driven and knowledge-guided denoising diffusion probabilistic model for runoff uncertainty prediction, Journal of Hydrology, 638, 131556, https://doi.org/10.1016/j.jhydrol.2024.131556, 2024.

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., Liu, X., Ren, W., Zheng, Y., Harman, C. J., Clark, M., Farthing, M., Feng, D., Kumar, P., Aboelyazeed, D., Rahmani, F., Song, Y., Beck, H. E., Bindas, T., Dwivedi, D., Fang, K., Höge, M., Rackauckas, C., Mohanty, B., Roy, T., Xu, C., and Lawson, K.: Differentiable modelling to unify machine learning and physical models for geosciences, Nat Rev Earth Environ, 4, 552–567, https://doi.org/10.1038/s43017-023-00450-9, 2023.

Sikorska, A. E., Viviroli, D., and Seibert, J.: Flood-type classification in mountainous catchments using crisp and fuzzy decision trees, Water Resources Research, 51, 7959–7976, https://doi.org/10.1002/2015WR017326, 2015.

Sun, A. Y., Jiang, P., Yang, Z.-L., Xie, Y., and Chen, X.: A graph neural network (GNN) approach to basin-scale river network learning: the role of physics-based connectivity and data fusion, Hydrol. Earth Syst. Sci., 26, 5163–5184, https://doi.org/10.5194/hess-26-5163-2022, 2022.

Sun, S., Shi, C., Pan, Y., Bai, L., Xu, B., Zhang, T., Han, S., and Jiang, L.: Applicability Assessment of the 1998–2018 CLDAS Multi-Source Precipitation Fusion Dataset over China, J Meteorol Res, 34, 879–892, https://doi.org/10.1007/s13351-020-9101-2, 2020.

Tam, W. C., Fu, E. Y., Li, J., Huang, X., Chen, J., and Huang, M. X.: A spatial temporal graph neural network model for predicting flashover in arbitrary building floorplans, Engineering Applications of Artificial Intelligence, 115, 105258, https://doi.org/10.1016/j.engappai.2022.105258, 2022.

Wasko, C., Nathan, R., Stein, L., and O'Shea, D.: Evidence of shorter more extreme rainfalls and increased flood variability under climate change, Journal of Hydrology, 603, 126994, https://doi.org/10.1016/j.jhydrol.2021.126994, 2021.

Xia, X., Liang, Q., Ming, X., and Hou, J.: An efficient and stable hydrodynamic model with novel source term discretization schemes for overland flow and flood simulations, Water Resources Research, 53, 3730–3759, https://doi.org/10.1002/2016WR020055, 2017.

Yin, H., Guo, Z., Zhang, X., Chen, J., and Zhang, Y.: RR-Former: Rainfall-runoff modeling based on Transformer, Journal of Hydrology, 609, 127781, https://doi.org/10.1016/j.jhydrol.2022.127781, 2022.

https://doi.org/10.5194/egusphere-2026-4038
Preprint. Discussion started: 21 July 2026

Yoganand, V., B, S. R., Kattukota, N., Balakrishna, E., and Meer, M. S.: Rainfall-based flood prediction by hybrid deep architecture with entropy and statistical feature set, International Journal of Image and Data Fusion, 16, 2427212, https://doi.org/10.1080/19479832.2024.2427212, 2025.

Yuan, X., Zhang, X., and Tian, F.: Research and application of an intelligent networking model for flood forecasting in the arid mountainous basins, J Flood Risk Management, 13, e12638, https://doi.org/10.1111/jfr3.12638, 2020.

Zhang, J., Yang, X., Fan, G., Li, H., and Zhou, J.: Physical and numerical modeling of a landslide dam breach and flood routing process, Journal of Hydrology, 628, 130552, https://doi.org/10.1016/j.jhydrol.2023.130552, 2024.

Zhang, L., Qin, H., Mao, J., Cao, X., and Fu, G.: High temporal resolution urban flood prediction using attention-based LSTM models, Journal of Hydrology, 620, 129499, https://doi.org/10.1016/j.jhydrol.2023.129499, 2023.

Zhang, S., Zhang, D., Huang, G., Wan, J., Kang, Y., Jiang, D., Xia, B., Zhao, Z., and Liu, R.: A novel framework for multi-step water level predicting by spatial–temporal deep learning models based on integrated physical models, Journal of Hydrology, 661, 133683, https://doi.org/10.1016/j.jhydrol.2025.133683, 2025.

Zhou, G., Cui, M., Wan, J., and Zhang, S.: A Review on Snowmelt Models: Progress and Prospect, Sustainability, 13, 11485, https://doi.org/10.3390/su132011485, 2021.

Zhou, Y., Wu, W., Nathan, R., and Wang, Q. J.: Deep Learning-Based Rapid Flood Inundation Modeling for Flat Floodplains With Complex Flow Paths, Water Resources Research, 58, e2022WR033214, https://doi.org/10.1029/2022WR033214, 2022.