Integrating Euclidean and non-Euclidean spatial information for deep learning-based spatiotemporal hydrological simulation

Liangkun Deng[1,2,3], Xiang Zhang[1,2,*], Louise J. Slater[3], Haoyuan Liu[1,2], Shiyong Tao[1,2]

[1] State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China

[2] Hubei Key Laboratory of Water System Science for Sponge City Construction, Wuhan University, Wuhan 430072, China

[3] School of Geography and the Environment, University of Oxford, South Parks Road, Oxford OX1 3QY, United Kingdom

Abstract

Spatiotemporal deep learning (DL) has emerged as a promising paradigm for hydrological simulation compared with lumped models using basin-averaged inputs. However, existing research primarily focuses on either Euclidean data, characterized by regular structures such as grid-like meteorological forcing, or non-Euclidean data, which features irregular topological connectivity such as downstream information transfer. Here, a novel spatiotemporal DL-based approach is developed to integrate both types of spatial information for streamflow modeling. Our proposed model utilizes convolutional neural network (CNN)- and graph neural network (GNN)-based modules to extract the spatial features of Euclidean and non-Euclidean data, respectively.

* Corresponding author.

E-mail address: zhangxiang@whu.edu.cn (X. Zhang)

Spatial-attention mechanisms are incorporated into these modules to automatically focus on important features. Long Short-Term Memory (LSTM) networks are then employed to capture temporal dependencies. The adaptability and response of the spatiotemporal DL-based model to different Euclidean and non-Euclidean inputs are investigated. The proposed model is applied to simulate multi-step and multi-gauge streamflow in the mid-lower Hanjiang River basin (ML-HRB), one of the largest sub-basins of the Yangtze River. Results show that when integrating Euclidean and non-Euclidean spatial data, the proposed model outperforms comparative models temporally, spatially and spatiotemporally. The average model improvements across all steps and gauges are of 17.75% for the Nash-Sutcliffe Efficiency coefficient (NSE), 9.02% for the correlation coefficient (CC), and 14.79% for the Kling–Gupta efficiency metric (KGE), compared with the commonly-used lumped LSTM model. The simulations of the tributary stream gauges are particularly sensitive to model refinements and spatial input selection. This study sheds light on the potential of integrating different kinds of spatial information for spatiotemporal DL-based simulation.

Keywords

Spatiotemporal deep learning; Hydrological simulation; Euclidean and non-Euclidean spatial data; Spatial information integration; The mid-lower Hanjiang River basin

1. Introduction

Hydrological simulation plays a pivotal role in managing water resources and safeguarding downstream ecosystems (Coopersmith et al., 2012; Ghotbi et al., 2020). Achieving accurate streamflow modeling remains a long-standing challenge owing to its inherent complexity and non-stationarities in the data (Lin et al., 2021; Yan et al., 2021). A multitude of methodologies have been developed to tackle this challenge, spanning from process-based to data-driven approaches (Zhang et al., 2015). Process-based models, including lumped models such as the Xin'anjiang model (Zhao, 1992) and distributed models such as the Soil and Water Assessment Tool (Gassman et al., 2007), strive to emulate the water cycle by leveraging the physics governing hydrological processes. However, these models are constrained by their physical inputs, simplified assumptions and time-consuming calibration, thus limiting their practical utility (Demir et al., 2015; Yaseen et al., 2019; Zhang et al., 2018). Data-driven models, especially deep learning (DL) techniques, provide a powerful alternative to traditional hydrological models (Hussain et al., 2020; Marçais and de Dreuzy, 2017; Van et al., 2020). Notably, Long Short-Term Memory (LSTM) networks have demonstrated superior performance over process-based hydrological models at the regional (Xie et al., 2021) and temporal scale (Gauch et al., 2021; Kratzert et al., 2019; Lees et al., 2021). However, current DL-based models often adopt a lumped approach with basin-averaged forcing data as inputs (Gauch and Lin, 2020). Yet, the hydrological response at the outlet is influenced by the spatial dependencies of meteorological and

geophysical characteristics across the entire basin (Hall et al., 2014; Jiang et al., 2018; Yang et al., 2020), and DL-based models largely overlook this spatial heterogeneity in contrast to process-based distributed hydrological models (Anderson and Radić, 2022; Marsh et al., 2020). Nowadays, many spatially distributed datasets, such as satellite estimates and reanalysis products, offer insights into such heterogeneity and can help achieve more accurate hydrological simulation. Therefore, there exists a critical avenue to further exploration of DL-based models to effectively utilize spatial information and advance hydrological modeling capabilities.

At least three types of approaches are commonly employed to conduct spatiotemporal streamflow simulation using DL-based models. The first type is associated with the field of large-sample hydrology, leveraging recurrent neural networks (RNNs) like LSTM to construct regional models by simultaneously training on a multitude of basins (Nearing et al., 2020; Xie et al., 2021). This method offers the advantage of enhancing the model's generalization across basins exhibiting hydrological similarity, enabling simulation even in ungauged basins (Kratzert et al., 2019). However, this approach falls short in representing explicit spatial variability due to the lack of spatially-variable inputs such as gridded precipitation, and the lack of topological connectivity (reflecting the transfer of information between upstream and downstream). The second type of approach integrates convolutional neural networks (CNNs) and RNN modules within a serial DL model. While CNNs are adept at extracting spatial features, RNNs effectively manage temporal dependencies (Sun et al., 2021). Originating from the field of

computer vision, CNNs demonstrate proficiency in processing various forms of Euclidean data (Pan et al., 2019; Weyn et al., 2019; Wu et al., 2020). These data, often derived from gauge interpolation, satellite remote sensing and reanalysis products, are structured in grid-like formations to represent the spatial variability of meteorological and hydrological features. RNN modules may occasionally be substituted with 1D-CNNs (Wunsch et al., 2022) or stacked time-varying inputs (Sun et al., 2021). In addition to the models with serial structures, a closely coupled model, named ConvLSTM, has gained in popularity in hydroclimatic prediction (Chen et al., 2022; Zhu et al., 2023). By replacing fully connected operators with convolutional operators, ConvLSTM can directly harness distributed meteorological data (Shi et al., 2015). The third type of approach centers on another kind of spatial information, i.e. topological connectivity, which falls under the category of non-Euclidean data. Unlike the organization of regular grid-like meteorological forcing data, the structure of topological connectivity depends on actual physical representations and adopts an irregular format. The streamflow at upstream gauges impacts the streamflow at downstream gauges as water moves through the river channel (Chen et al., 2022; Zhao et al., 2020). Topological connectivity offers an explicit depiction of these contextual dependencies among the stream gauges (Sun et al., 2021; Wu et al., 2019). Recently, graph neural networks (GNNs) have emerged as a promising tool for extracting non-Euclidean topological dependencies, which CNNs fail to process due to their assumptions of spatial regularity and locality. GNNs aggregate spatial information from

target gauges and their neighborhood and then update their status to simulate streamflow across all gauges within a river network. One of the advantages of GNNs lies in their incorporation of the physical realism of a river network, allowing for controlling information transferring at both node and edge levels—a feature absent from many other machine learning algorithms dealing with spatiotemporal characteristics (Sun et al., 2022). The spatiotemporal realization of this type of approach can either combine GNNs with a temporal module (Deng et al., 2023; Zanfei et al., 2022) or structure them in a recurrent manner (Chen et al., 2021; Liu et al., 2023). However, current research has predominantly focused on leveraging either CNNs or GNNs, neglecting the potential benefits of integrating multiple types of spatial inputs. Given the distinct value of separate spatial inputs in streamflow simulation, the combination of regular grid-like meteorological forcing data and irregular topology-like connectivity is hypothesized to provide more advantageous features to delineate spatial variability and achieve better hydrological performance.

Proper input selection holds considerable importance and exerts a critical influence on simulation (Kim et al., 2023; Wang et al., 2022). Currently, model developers have access to various options. For instance, considering meteorological forcing data, advancements in remote sensing and numerical simulation techniques have facilitated the creation of diverse open-access gridded meteorological datasets derived from satellite and reanalysis products. These datasets offer finer spatiotemporal resolution, easier access, and greater spatial continuity than gauge precipitation (Hussain et al.,

2018). Due to the disparities in data sources and retrieval algorithms, they exhibit varying quality and characteristics (Satgé et al., 2021; Wang et al., 2021; Yuan et al., 2017). A significant number of studies have assessed the hydrological applications of these products using process-based hydrological models, revealing broad variations in predictive results (Gu et al., 2023; Muñoz et al., 2023; Wang et al., 2023; Xu et al., 2022). In contrast, the exploration of spatiotemporal DL-based models' response to these datasets remains relatively uncharted (Speight et al., 2021). Zhu et al. (2023) compared the hydrological performance of three precipitation products using DL-based models. However, their selection of precipitation products was limited and the comparisons were not comprehensive enough to demonstrate the adaptability and response of spatiotemporal DL-based approaches. They only considered the effect of spatial resolution without examining the effects of product types (satellite estimates, reanalysis products and merging products) and improvement techniques like bias correction and merging. Therefore, further investigation is warranted to assess the adaptability of spatiotemporal DL-based models to gridded meteorological forcing data of varying quality. Moreover, the hydrological responses of spatiotemporal DL-based and process-based models need to be compared to increase our understanding of the utility of various meteorological products.

Different graph representations also have implications for predictive results, as they are the foundation of GNNs used to describe topological connectivity and determine learning processes. However, whether factors such as direction and weight are taken

into account during the graph generation relies on the hydrological perspectives of model developers, with comprehensive analysis often overlooked in prior studies. Moreover, the incorporation of a complete river network into a graph raises questions about the degree of physical realism necessary for effective graph learning. Sun et al. (2022) compared the performance of graph-based river network models consisting of complete and reduced nodes to find out whether simplified graphs would still yield precise predictions. However, their study area comprises dense nodes with observed data for training. It is unclear about the generalizability of their findings to many other regions, where historical streamflow observations are not publicly available and only sparsely distributed nodes left in a coarse graph meet data requirements for simulation. This study aims to address the above research gaps by developing a novel spatiotemporal DL-based model to integrate spatial information from regular grid-like meteorological forcing data (Euclidean spatial data) and irregular topology-like connectivity (non-Euclidean spatial data). The proposed model utilizes CNN-based and GNN-based modules to deal with different kinds of spatial inputs and is referred to as the spatiotemporal CNN- and GNN-based DL model (SCGDL). The model is used to conduct multi-step streamflow simulation at multiple stream gauges simultaneously in the mid-lower Hanjiang River basin (ML-HRB). The objectives of this work are four-fold: (1) develop a novel spatiotemporal DL-based model that integrates Euclidean and non-Euclidian spatial information; (2) evaluate the temporal, spatial and spatiotemporal performance of the proposed model; (3) investigate the adaptability of the

spatiotemporal DL-based model to gridded meteorological forcing data varying in quality and characteristics; and (4) explore the response of the spatiotemporal DL-based model to topological connectivity with different graph representations. The principal contribution of this work is to unravel the potential of integrating varied kinds of spatial information into spatiotemporal DL-based models and to investigate the hydrological performance feedback on different spatial inputs.

The remainder of the paper is organized as follows. Section 2 introduces the study area and data collection. The details of the proposed SCGDL are described in Section 3. Section 4 presents the experimental results, and the discussion is provided in Section 5. Section 6 summarizes this study and suggests areas for future research.

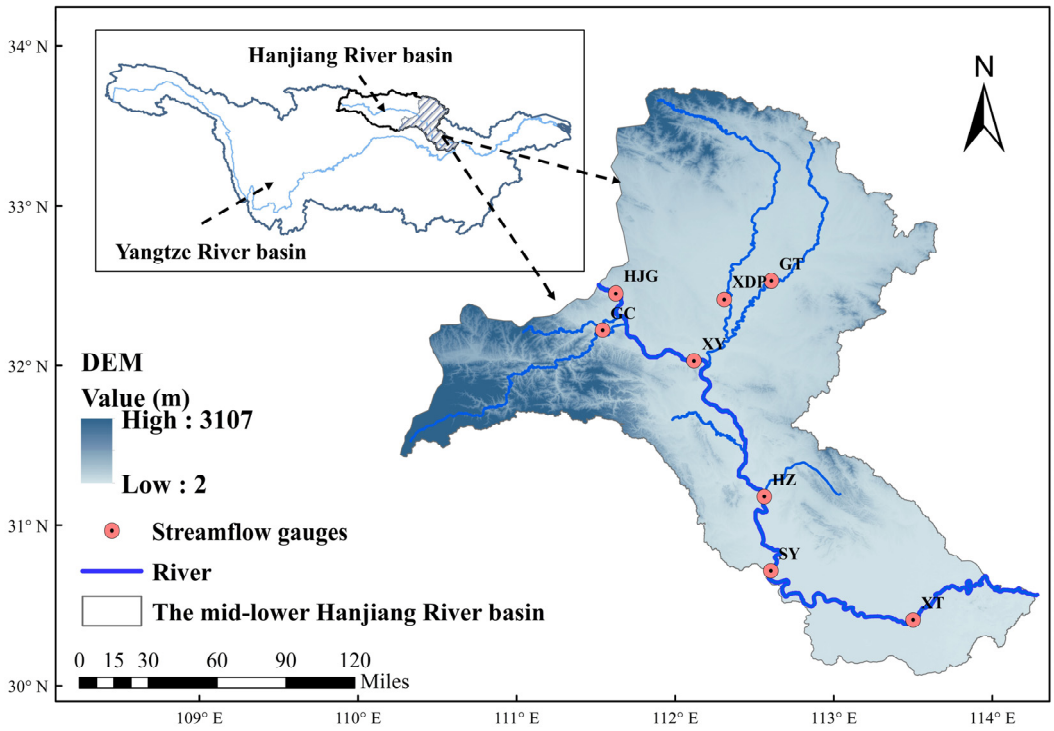
2. Study area and materials

2.1 Study area

The ML-HRB is assigned as the study area shown in Fig. 1. The Hanjiang River basin (HRB), located in central China, is one of the largest sub-basins of the Yangtze River basin. The basin is separated into an upper and a mid-lower part by the Danjiangkou reservoir. The mid-lower section of the basin extends from the reservoir to Wuhan City, where it joins the Yangtze River. Encompassing an area of 63,800 km^2, with a river length of approximately 652 km, the ML-HRB experiences subtropical monsoons with an average annual temperature of 15-17°C. The hydrometeorological characteristics of the basin exhibit strong spatiotemporal variability. While the southern plains receive an annual precipitation of around 1200mm, the northern hilly regions receive between 700-800mm. Moreover, over 70% of the annual precipitation falls between May and September (Cheng et al., 2019). These characteristics result in the complex and non-linear streamflow generation.

A total of eight stream gauges with close topological connectivity in this river network are selected, named Huangjiagang (HJG), Xiangyang (XY), Huangzhuang (HZ), Shayang (SY), Xiantao (XT), Gucheng (GC), Xindianpu (XDP) and Guotan (GT) respectively. They are classified into two groups: the first five gauges are distributed in the mainstream arranged from upstream to downstream and the last three are located in the tributaries. The average streamflow of the first group is approximately 18 times larger than that of the second. The spatiotemporal variability across the basin

196 complicates the task of accurately simulating streamflow.



197

198 **Fig. 1.** Locations of the study area ML-HRB and the eight stream gauges.

199 **2.2 Data collection**

200 The data used in our research include meteorological forcing data, historical streamflow
201 observations and topological connectivity among the stream gauges. The time series
202 are all aggregated at the daily scale and the period is from 2001/1/1 to 2015/12/31 due
203 to the limited data accessibility. The meteorological forcing data include daily fields of
204 precipitation and potential evaporation. Precipitation and potential evaporation are
205 obtained from the ERA5-Land reanalysis dataset, downloaded from the Copernicus
206 Climate Data Store. ERA5-Land is a replay of the land component of the ERA5
207 reanalysis at a finer spatial resolution of 0.1° over the global land surface (Hersbach et
208 al., 2020; Muñoz-Sabater et al., 2021). The reference precipitation observations derived
209 from 10 meteorological stations and 11 rain gauges within the study area are

downloaded from the China Meteorological Data Service Center (http://data.cma.cn/) and provided by the Hubei Institute of Water Resources Survey and Design in China. The daily historical streamflow observations of the eight stream gauges are provided by the Hubei Institute of Water Resources Survey and Design in China. We obtained a total of nine gridded precipitation products to investigate the adaptability of spatiotemporal DL-based models to different precipitation inputs, including ERA5 reanalysis (Hersbach et al., 2020), two spatial resolutions of the Climate Hazards Group InfraRed Precipitation with Station data (hereafter, CHIRPS005 and CHIRPS025 respectively; 0.05° × 0.05° and 0.25° × 0.25°) (Funk et al., 2015), Integrated Multi-satellite Retrievals for GPM Early Run and Final Run (hereafter, IMERG-E and IMERG-F respectively) (Huffman et al., 2019a, 2019b), the Climate Data Record version and the Cloud Classification System-Climate Data Record version of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (hereafter, PCDR and PCCSCDR respectively) (Ashouri et al., 2015; Sadeghi et al., 2021), ERA5-Land reanalysis and a merging product. The merging product (TCMerge for short) comes from the ERA5 reanalysis, PCDR and CHIRPS025 using the triple collocation (Li et al., 2018). These satellite estimates and reanalysis products are publicly available on their websites.

3. Methodology

Here, we introduce the SCGDL, a novel spatiotemporal DL-based model designed to leverage diverse spatial information for multi-step streamflow simulation across multiple gauges simultaneously. As illustrated in Fig. 2, the SCGDL is fed with regularly gridded meteorological forcing data (precipitation and potential evaporation), irregular topological connectivity, and historical streamflow observations. The model architecture consists of two key phases: a spatial dependency extraction phase followed by a temporal feature capture phase. This kind of sequential structure has been proven to be powerful in many fields, such as traffic forecasts (Tao et al., 2023), air pollution predictions (Yan et al., 2021), groundwater level estimates (Bai and Tahmasebi, 2023) and streamflow forecasts (Anderson and Radić, 2022; Wunsch et al., 2022).

The first phase adopts the CNN-based modules to process Euclidean grid-like meteorological forcing data and the GNN-based modules to handle non-Euclidean topological connectivity. Subsequently, the temporal phase utilizes the LSTM module to capture latent time characteristics of the spatial representation. The detailed descriptions of the CNN-based and GNN-based modules are provided below. The details of LSTM can be found in Kratzert et al. (2018). Our previous work only considered one kind of spatial information, i.e., non-Euclidean topological connectivity, using a GNN-based model, and without incorporating meteorological forcing data (Deng et al., 2023).

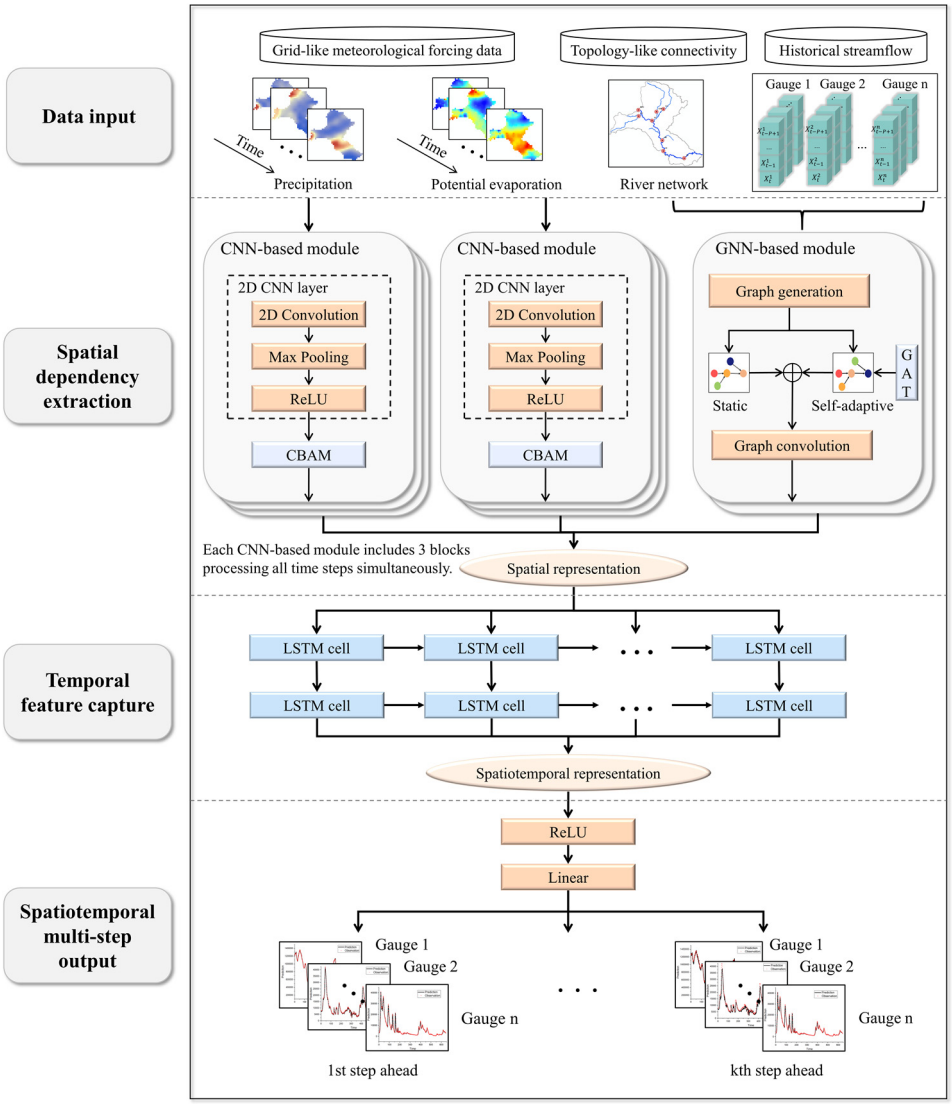


Fig. 2. Framework of the proposed model SCGDL.

3.1 CNN-based modules for extracting grid-like spatial dependencies

CNNs, first introduced by LeCun et al. (1998), have been broadly applied to learn features of regularly structured data in several domains, such as video prediction (Oprea et al., 2020), semantic segmentation (Frizzi et al., 2021) and fault recognition (Wang et al., 2019). The structure of the CNN-based module is established by the repetition of blocks, each composed of one 2D CNN layer and one convolutional block attention module (CBAM). A 2D CNN layer is made of a 2D convolution, a pooling, and a

257 rectified linear unit (ReLU) activation function in sequence. The 2D convolution first
258 uses local receptive fields and shared kernels to produce abstract feature maps (Tu et
259 al., 2021). Next, the produced feature maps are down-sampled and compressed by the
260 pooling operation, which helps further extract main features and reduces the workload
261 of computational complexity (LeCun et al., 2015). Here we use a max pooling operation.
262 After the convolution and pooling operation, the ReLU function is adopted to enhance
263 the non-linear properties of the linear convolved feature maps. The incorporation of
264 spatial-attention mechanisms helps improve model performance (Bello et al., 2019; Hu
265 et al., 2018; Zhang et al., 2018). CBAM, one of the most popular spatial-attention
266 mechanisms, can amplify desired meteorological features and suppress undesired ones
267 (Trebing et al., 2021; Woo et al., 2018). It directs the CNN-based modules to pay more
268 attention to essential signals and contributes to higher accuracy. CBAM employs
269 attention in sequence to the channel and spatial dimensions of inputs. The output is a
270 weighted feature map that highlights the essential spatial information. The details are
271 described in Woo et al. (2018).
272 Mathematically, the whole process of a block is expressed as follows:

273 $$h_j^l = CBAM(ReLU(Maxpooling(\sum_{i \in M} W_{ij}^l * X_i^{l-1} + b_j^l))) \quad (1)$$

274 where i, j and l are the ith input neuron, the jth output neuron and the lth block,
275 respectively, M denotes the number of input channels of maps, X_i^{l-1} represents the
276 input from the last block, W_{ij}^l denotes the convolution kernel, which contains the
277 learnable parameters, $*$ is the convolutional operation, b_j^l represents the learnable

bias for the jth output in the lth block, and h_j^l denotes the jth output in the lth block after a series of operations.

3.2 GNN-based modules for extracting topology-like spatial dependencies

Downstream gauges receive streamflow from upstream gauges, establishing close topological dependencies that are expressed as contextual relationships within the river network (Zhao et al., 2020). Here, we introduce the GNN-based modules to extract irregular topological dependencies among stream gauges. A river network can be generalized into a graph of $G = (V, E, W)$, where V is a set of n stream gauges on the graph, also referred to as nodes, and E is a set of edges representing the topological connectivity between node pairs (v_i, v_j) for $v_i \in V$ and $v_j \in V$. $W \in \mathbb{R}^{n \times n}$ denotes the learnable parameters of the graph. Each node in the graph is characterized by its attributes and the neighboring nodes to which it is connected. The attribute matrix of the nodes in the graph is denoted as $X \in \mathbb{R}^{n \times d}$, where d is the attribute number of each node. The node attribute in our study is historical streamflow observation. Topological connectivity can be specified by a static adjacency matrix $\mathrm{A}_{st} = (a_{ij})_{n \times n}$, of which its element $a_{ij} = 1$ in the case of unweighted graph or a decimal number presenting similarity in the case of a weighted graph if nodes v_i and v_j are connected, and $a_{ij} = 0$ otherwise. The self-loops are also considered and the neighborhood is described as $N(i) = \{v_j \in V | e(v_i, v_j) \in E\} \cup \{v_i\}$. The graph can be either undirected or directed, which depends on whether connective edges are bidirected. In our study, we assume that the influence between upstream and downstream gauges follows a one-

299 way direction of flow path, and we consider the graph as directed. We also consider the
300 graph as unweighted.
301 A GNN-based module comprises two stages, namely graph generation and graph
302 convolution, shown in Fig. 3. In the first stage, two types of graphs are generated to
303 describe static and self-adaptive characteristics of topological connectivity. The static
304 graph is pre-defined based on the physical topological relationships among stream
305 gauges. Since the gauge locations are stable, it remain constant over time. Hence, the
306 corresponding adjacency matrix and graph are referred to as "static". However, the
307 static graph derived from physical connectivity measurements may not fully capture
308 latent dependencies between each pair of gauges (Bai and Tahmasebi, 2023; Sarkar et
309 al., 2015). It is necessary to add an extra spatial-attention mechanism for graph learning
310 to capture the underlying changes adaptively. Specifically, a self-adaptive adjacency
311 matrix is built with its elements generated automatically by Graph Attention Networks
312 (GAT) (Veličković et al., 2017). Both the static and self-adaptive graphs share the same
313 nodes, but have different adjacency matrices. The self-adaptive adjacency matrix
314 $A_{sa} = (\alpha_{ij})_{n\times n}$ is constructed with a shared operator $\vec{a}$, which is a feed-forward fully
315 connected layer:
316 $$e_{ij} = LeakyReLU(\vec{a}^T[W_{sa}\vec{h}_i || W_{sa}\vec{h}_j]) \quad (2)$$
317 where e_{ij} denotes the attention coefficient depicting the importance level of the node
318 v_i to v_j, W_{sa} is the learnable parameter matrix, $\vec{h}$ represents the input features, and
319 $.^T$ and $||$ are the transposition and concatenation operations, respectively. The

equation is calculated when $v_i \in N(j)$. The attention coefficients need to be normalized across all nodes using the softmax function given below:

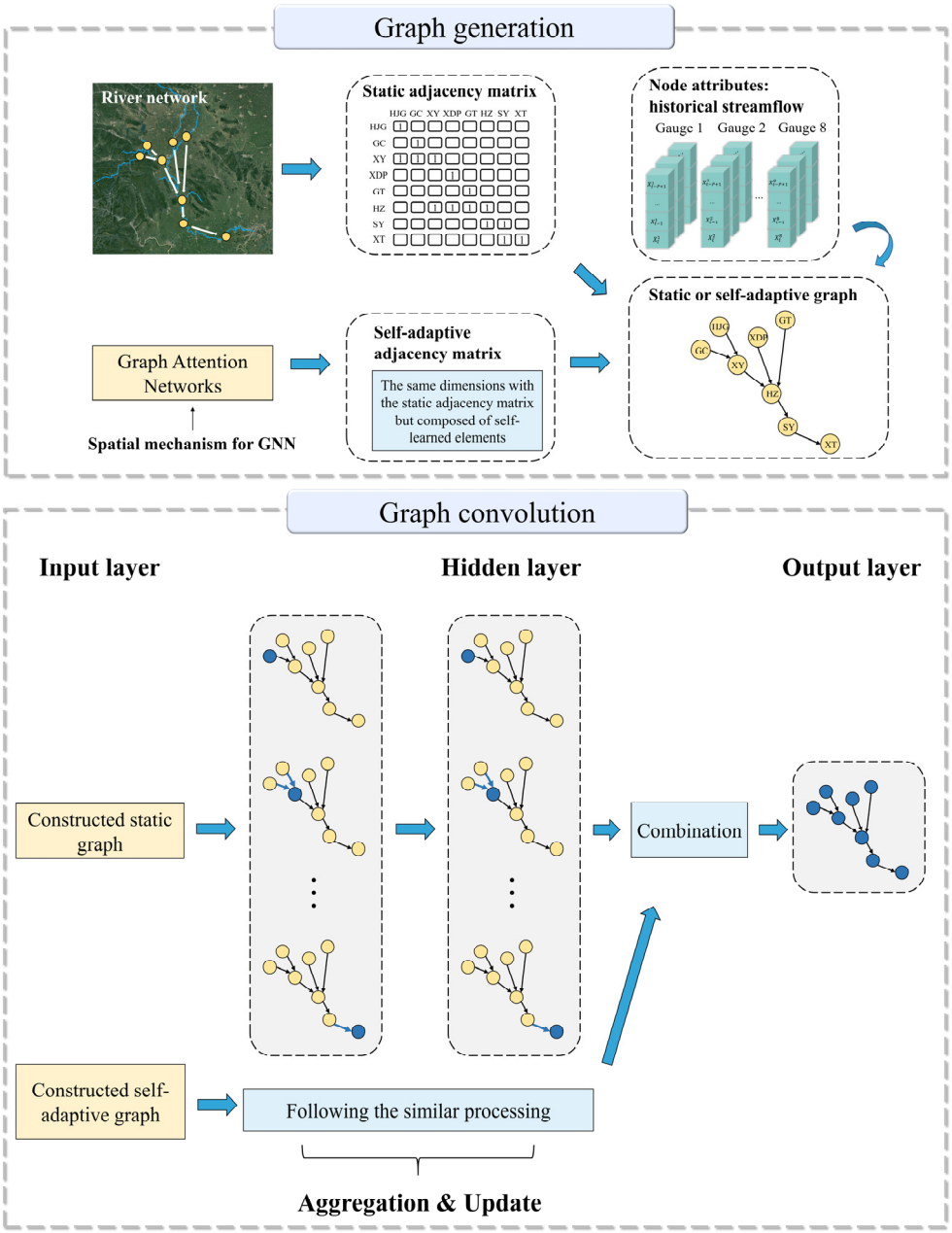
$$\alpha_{ij} = softmax_i(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{i \in N(j)} \exp(e_{ij})} \quad (3)$$

The generated static and self-adaptive graphs are then sent into the second stage, i.e. graph convolution. The objective of graph convolution is to consider static and self-adaptive topological properties and aggregate the information of the neighborhood $N(j)$ to update the hidden state of target nodes. The graph convolution operation is specified as a diffusion process with Q hops such that nodal information is diffused to q-hop neighborhoods (Li et al., 2017; Wu et al., 2019), which indicates a node can aggregate information up to qth-order neighbors. It can be expressed as :

$$\mathrm{Z} = \sum_{q=0}^{Q} L^q X W^q \quad (4)$$

where, L^q is the normalized adjacency matrix at the qth hop and $L = A/rowsum(A)$. The combination of the static graph and the self-adaptive graph convolution is defined as:

$$\mathrm{Z} = \sum_{q=0}^{Q} L_{st}^q X W_{st}^q + L_{sa}^q X W_{sa}^q \quad (5)$$



335

336 **Fig. 3.** Illustration of the GNN-based modules. The GNN-based module is separated into two stages,

337 namely graph generation and graph convolution. In the first stage, static and self-adaptive graphs

338 are generated based on the physical topological relationships among stream gauges and the GAT

339 respectively. In the second stage, the established graphs are sent into the graph convolution to

340 aggregate the information of neighborhood to update the hidden state of target nodes. Two

341 convolution layers are employed to aggregate information from farther neighbors.

3.3 Framework of the proposed model

Procedure 1: data preprocessing

The input data are gridded precipitation and potential evaporation over the study area, along with the historical streamflow observations and the topological connectivity of the gauges in the river network. The meteorological inputs and streamflow observations are preprocessed, including data cleaning, data normalization and window sliding process (separating the input data into overlapping chunks for training). The static graph is constructed based on the physical locations of the gauges.

Procedure 2: spatiotemporal modeling

After data preprocessing, the input data is sent into the spatial modules to extract the dependencies in space. The gridded precipitation and potential evaporation are fed into two independent CNN-based modules respectively. The self-adaptive graph is constructed by using the GAT. The generated static and self-adaptive graphs are provided to the graph convolutional layer in the GNN-based modules. Subsequently, three outputs with spatial awareness are derived and concatenated together to explore their temporal correlations through the LSTM module. This module processes the extracted features across all the gauges. Finally, the ReLU activation function and dense layer are followed to receive the spatiotemporal representation and output the multi-step streamflow simulation.

3.4 Experimental design

This study aims to assess the potential of integrating Euclidean and non-Euclidean

spatial information, and to investigate the response of the spatiotemporal DL-based model to spatial inputs of different quality and graph representations. Fig. 4 shows the flowchart of this research. Three experiments are designed accordingly. In experiment I, the SCGDL is developed to address the first two objectives introduced in Section 1. The gridded meteorological forcing data come from ERA5-Land, while the graphs are built using all the gauges representing complete physical realism. The comparison with the baseline model LSTM and the ablation experiments are then conducted to quantify the overall superiority and the effectiveness of each spatial component, respectively. The evaluation metrics include the Nash-Sutcliffe Efficiency coefficient (NSE), the correlation coefficient (CC) and the Kling–Gupta efficiency metric (KGE), shown in Table 1. The metrics for each gauge are calculated as the ensemble averages after training each model 10 times using different random seeds. In experiment II, the nine precipitation products of varied quality sourced from satellite estimates, reanalysis products and a merging product are independently sent into the SCGDL to train parameters to realize the third objective. In experiment III, the influences of weight, direction and degree of physical realism in different graph representations are investigated to address the fourth objective. As the unweighted and directed graph is adopted in the SCGDL, the weighted and undirected graphs are constructed respectively and are put into the model independently. In addition, simplified graphs with reduced nodes are created to assess the degree of physical realism required for river network learning. This aims to explore whether simplified graphs would still make

accurate simulation in a river network with sparsely distributed nodes.

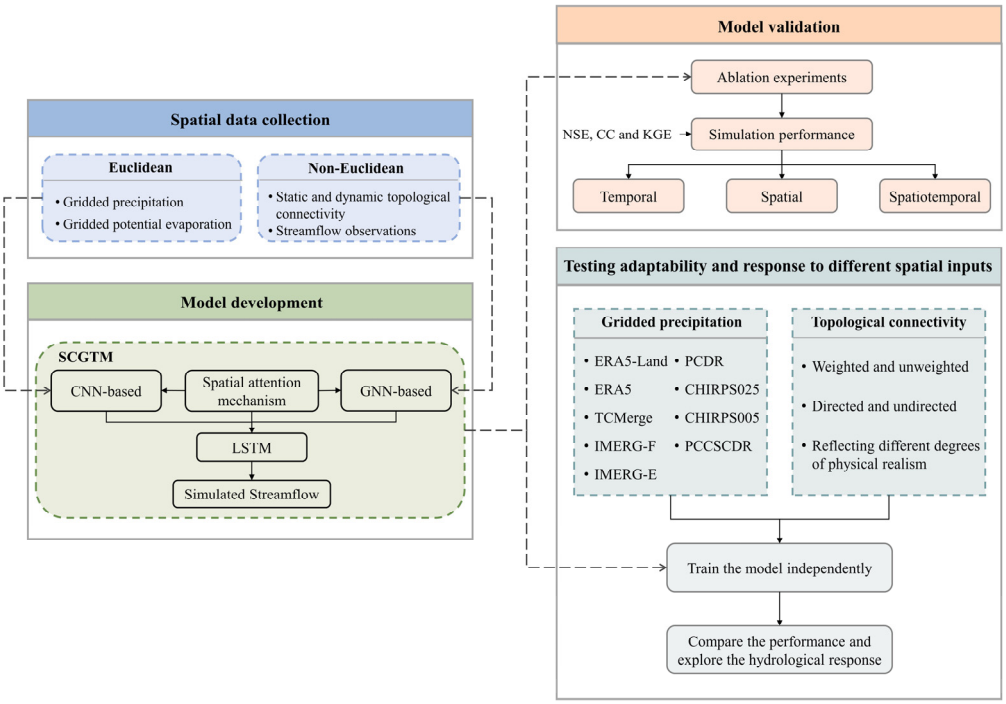


Fig. 4. Flowchart of this study.

Considering the computational cost of employing gridded meteorological forcing data, a past-30-day window is used to establish the training instances. The target is set as the streamflow in the next 5 days to test the multi-step simulation performance. The collected data ranges from 2001 to 2015. The splitting ratio of the whole sequences for training, validation and testing is 6:2:2. The maximum training epoch is set to 200. For model training, different sets of hyperparameters, such as the CNN blocks (1,2,3), the kernels (32, 64, 128), the diffusion hops (1, 2, 3), the LSTM hidden size (64, 128, 256), the LSTM hidden layers (1, 2, 3), and the dropout rates (0.2, 0.3, 0.4) are set up to identify the optimal combination. After experimenting with these possible values, the final hyperparameter values have been settled. For the CNN-based modules, the number of the blocks is 3, and the numbers of the 2D CNN kernels in the blocks are 32, 64, and

64 respectively. For the GNN-based module, the diffusion hop is 2 and the dropout rate is 0.3. For the LSTM module, the number of hidden layers is 2, and therefore the LSTM has two rows in Fig. 2 (while the columns indicate the processing steps in a hidden layer). The layers both have a hidden size of 256. The updated hidden states in the first layer are passed as inputs to the second layer. The LSTM module has a dropout rate of 0.4 as well. Table S1 in the supporting information presents the descriptions and dimensions of each module. The loss function adopts Mean Absolute Error and the optimizer is Adam with the initial learning rate of $1e^{-3}$. All the models are implemented on a Linux system (CPU: 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz, GPU: NVIDIA GeForce RTX 3080; deep learning library: PyTorch 1.10 framework on Python 3.8).

Table 1

Summary of the statistical metrics used for model evaluation.

Metrics	Equation	Range
Nash-Sutcliffe Efficiency coefficient (NSE)	$NSE = 1 - \frac{\sum_{t=1}^{T}(S_t - O_t)^2}{\sum_{t=1}^{T}(O_t - \bar{O})^2}$	(-∞,1], close to 1 is better
Correlation coefficient (CC)	$CC = \frac{\sum_{t=1}^{T}(S_t - \bar{S})(O_t - \bar{O})}{\sqrt{\sum_{t=1}^{T}(S_t - \bar{S})^2}\sqrt{\sum_{t=1}^{T}(O_t - \bar{O})^2}}$	[-1,1], close to 1 is better
Kling–Gupta efficiency (KGE)	$KGE = 1 - \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$	(-∞,1], close to 1 is better

411 Note: S and O are the simulations and observations, r denotes the Pearson's correlation
412 coefficient, β represents the ratio of simulated to observed mean values, and γ refers to the ratio
413 of simulated to observed coefficients of variation.

4. Results

4.1 Experiment I: evaluating the integration of multiple kinds of spatial information

The LSTM serves as the baseline model in our study. The LSTM is separately established for each stream gauge, using the basin-averaged precipitation and potential evaporation directly contributing to the target gauge, along with the historical streamflow of that gauge. The acronyms SCDL, SGDL and SCGDL-NoAtt denote models where GNN-based modules, CNN-based modules and spatial-attention mechanisms are removed from the SCGDL respectively. In contrast to the LSTM, the other models are input with spatial information and can be considered as spatiotemporal models. Table 2 provides the comparative model settings.

Table 2

Summary of the comparative model settings.

Model	Inputs			
	Gridded meteorological forcing data	Basin-averaged meteorological forcing data	Topological connectivity	Spatial-attention mechanisms
SCGDL	√		√	√
SCGDL-NoAtt	√		√	
SCDL	√			√
SGDL		√	√	√

LSTM	√				

Table 3 reveals the performance rankings of the models, with SCGDL leading, followed by SCGDL-NoAtt, SCDL, SGDL, and LSTM. After incorporating the spatial meteorological data, topological connectivity and the spatial-attention mechanisms, the proposed SCGDL model obtains a NSE of 0.889, a CC of 0.943 and a KGE of 0.908. Relative to the LSTM, the SCGDL showcases enhancements with percentage improvements of 17.75% for NSE, 9.02% for CC, and 14.79% for KGE. The spatiotemporal models all outperform the benchmark model, which suggests that improvements can be attained no matter what kinds of spatial information are incorporated. The performance comparison between the SCDL and SGDL offers valuable insights into the effectiveness of various types of spatial information. It indicates that the gridded meteorological information plays a more important role in spatiotemporal modeling. This observation is logical, considering that meteorological variables, particularly precipitation, are widely recognized as pivotal factors influencing streamflow simulation. (Bárdossy and Anwar, 2023; Nijssen and Lettenmaier, 2004). After integrating the Euclidean and non-Euclidean spatial information, adding the spatial-attention mechanisms further helps the model reach higher precision. By comparing the performance between the LSTM, SCGDL-NoAtt and SCGDL models, the improvement of employing the spatial-attention mechanisms is smaller than that of integrating different kinds of spatial inputs. It indicates that the spatial-attention mechanisms have an secondary effect while the spatial inputs are the

447 primary reason why the predictive performance is greatly enhanced in our study.

448 **Table 3**

449 Summary of the average evaluation metrics alongside the percentage improvements of the SCGDL
450 relative to each comparative model for the testing period. The percentage improvements are used to
451 quantify the superiority of the proposed SCGDL model. The best metrics are bolded.

Model	NSE	NSE-Improvement	CC	CC-Improvement	KGE	KGE-Improvement
SCGDL	**0.889**	-	**0.943**	-	**0.908**	-
SCGDL-NoAtt	0.851	4.47%	0.925	1.95%	0.878	3.42%
SCDL	0.839	5.96%	0.921	2.39%	0.860	5.58%
SGDL	0.789	12.67%	0.889	6.07%	0.834	8.87%
LSTM	0.755	17.75%	0.865	9.02%	0.791	14.79%

452 Fig. 5 shows the temporal variation of the spatial-averaged metrics and the percentage
453 improvements at each lead time. The performance comparison consistently
454 demonstrates the superior performance of SCGDL, with the baseline LSTM exhibiting
455 the poorest performance. The NSE of the SCGDL experiences a moderate decrease of
456 15.40% from 0.948 to 0.802 between lead time of 1 to 5 days, while LSTM's NSE
457 exhibits a more pronounced decline of 21.68% from 0.844 to 0.661 over the same
458 period. The percentage improvements of the SCGDL over the LSTM and SGDL
459 increase rather than decrease as the lead time extends. This phenomenon is similarly

observed in CC and KGE comparisons. Such trends suggest that after integrating different kinds of spatial information, the proposed SCGDL demonstrates the superiority in accurate and robust multi-step simulation.

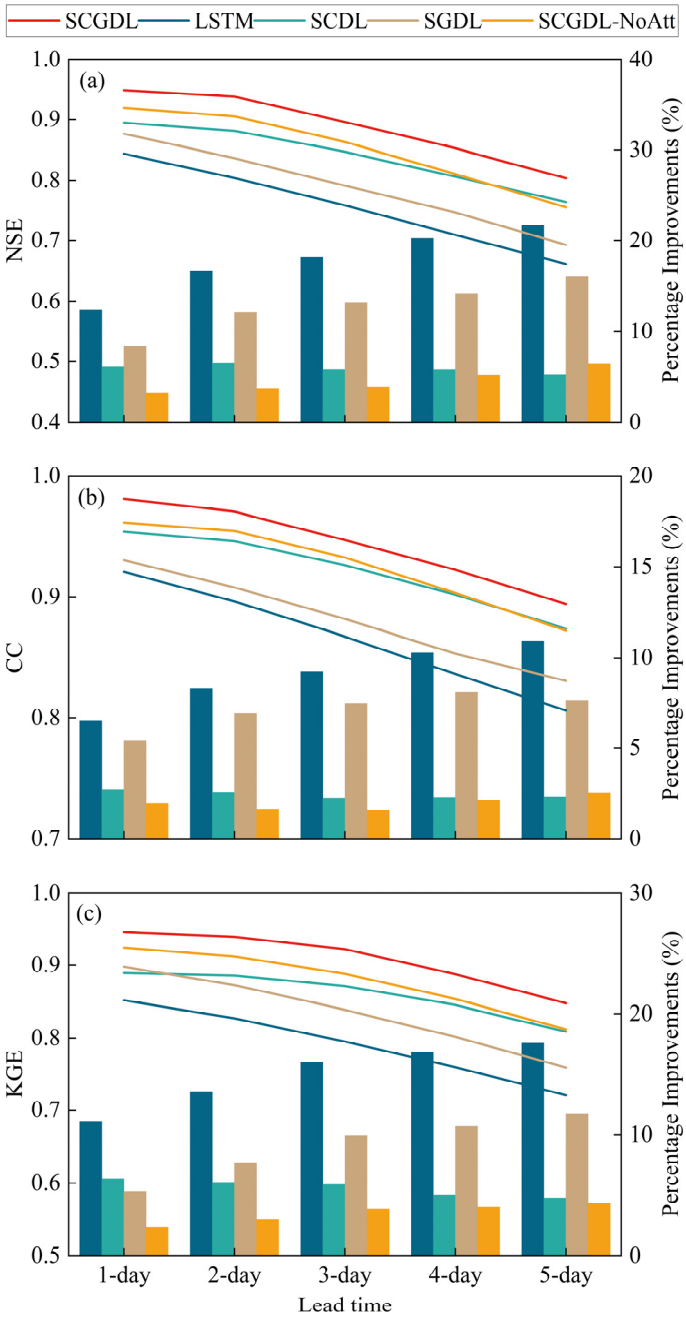


Fig. 5. Temporal variation of the spatial-averaged metrics including NSE, CC, and KGE for each model (represented by lines), along with the percentage improvements of the proposed SCGDL model relative to each comparative model (depicted by bars) at each lead time. In the line charts,

each color in the legend corresponds to the performance of a specific model. In the bar charts, each color signifies the improvement percentage over the corresponding model being compared to the SCGDL. Specifically, the SCGDL is compared with each comparative model, including the LSTM, SCDL, SGDL, and SCGDL-NoAtt, respectively. This comparison aims to assess the improvement of the SCGDL over the baseline (SCGDL vs. LSTM) and the effectiveness of incorporating Euclidean (SCGDL vs. SGDL), non-Euclidean spatial data (SCGDL vs. SCDL) and spatial-attention mechanisms (SCGDL vs SCGDL-NoAtt).

The eight gauges fall into two groups: the mainstream gauges, namely HJG, XY, HZ, SY and XT, characterized by significantly higher streamflow, and the tributary gauges, namely GC, XDP and GT, characterized by notably lower streamflow. We further investigate the spatial variation of timestep-averaged performance over 5 simulation steps of each stream gauge. On one hand, the gauges sited in the mainstream tend to attain higher performance and each model demonstrates a satisfactory ability to model the streamflow at the XY, HZ, SY and XT gauges (Fig. 6(a)). The metrics obtained by SCGDL and LSTM for the four mainstream gauges are remarkably similar. Fig. 6(b) shows the percentage improvements are under 5.6%, particularly for the SY and XT gauges, where the improvements are even under 2.5%, indicating relatively marginal accuracy gains. The HJG gauge, located at the outlet of the Danjiangkou reservoir, faces vulnerability to reservoir outflow, leading to less stable predictive performance compared to the other mainstream gauges. However, the incorporation of spatial information helps mitigate the impact of reservoir outflow and achieve a higher

accuracy. On the other hand, the LSTM predictions for the tributary gauges, i.e. GC, XDP, and GT, are comparatively inferior, with the lowest NSE no exceeding 0.5. Notably, more pronounced improvements are observed at the tributary gauges, as depicted in Fig. 6(b), with NSE percentage improvements exceeding 50% at the GC and XDP gauges. This highlights the primary contribution of SCGDL's superior predictive capacity in modeling smaller tributary streamflow.

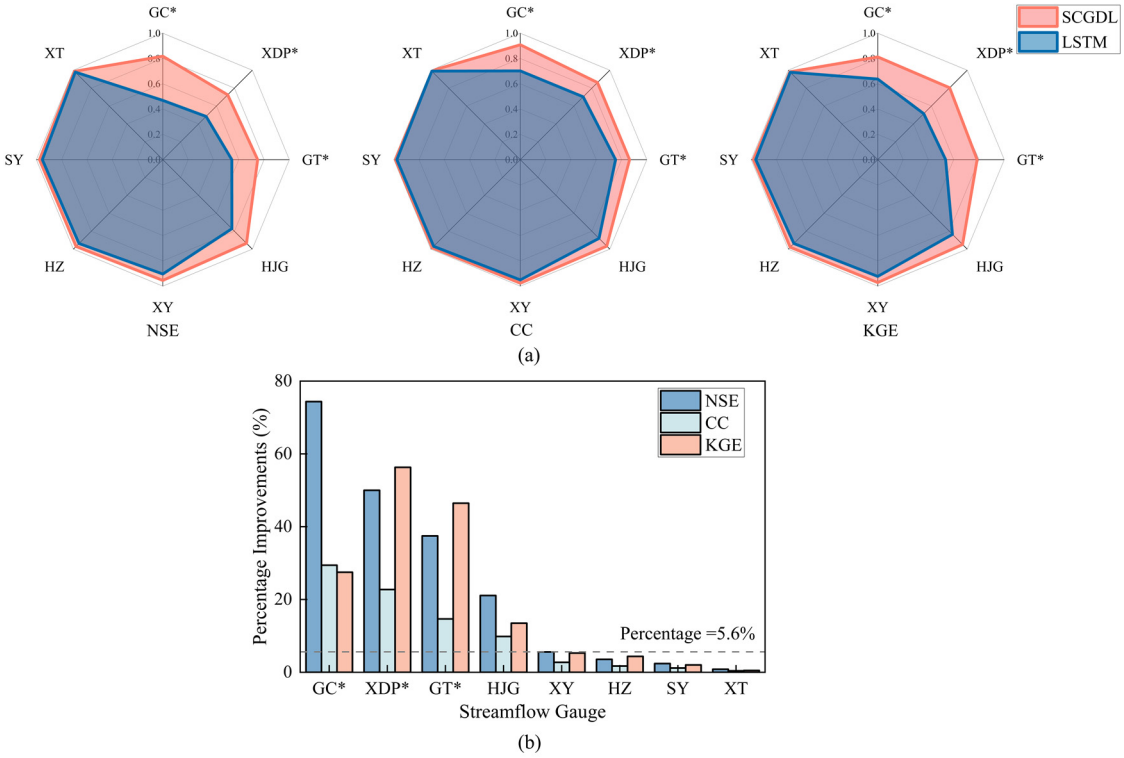


Fig. 6. Comparisons of (a) the timestep-averaged NSE, CC and KGE over 5 simulation steps of each stream gauge, and (b) the corresponding percentage improvements of the proposed SCGDL model relative to the baseline LSTM model at each gauge. The tributary gauges are marked with asterisks, while the mainstream gauges are arranged (a) clockwise or (b) sequentially from upstream to downstream.

To assess the model's ability in reproducing high flow events, the predictive results of

the top 5% of flow simulated by the SCGDL and the other comparative models are represented in Fig. 7(a). The SCGDL performs the best with the smallest average and median absolute relative error at each lead time. We investigate the most severe floods observed at two representative gauges located in the mainstream and tributary respectively, i.e. the XT and XDP gauges. Their peak discharges are in the top 1% and 2.5% of their long-term historical record, respectively. Their simulations at the 3-day lead time are compared in Fig. 7(b). Regarding the XT gauge simulation, both SCGDL and LSTM results both show satisfactory agreement with observations. However, for the simulation of the XDP gauge, the SCGDL's streamflow prediction better captures the peak discharge. Therefore, we find that the proposed SCGDL model exhibits superiority not only in modeling overall trends but also in reproducing extreme floods.

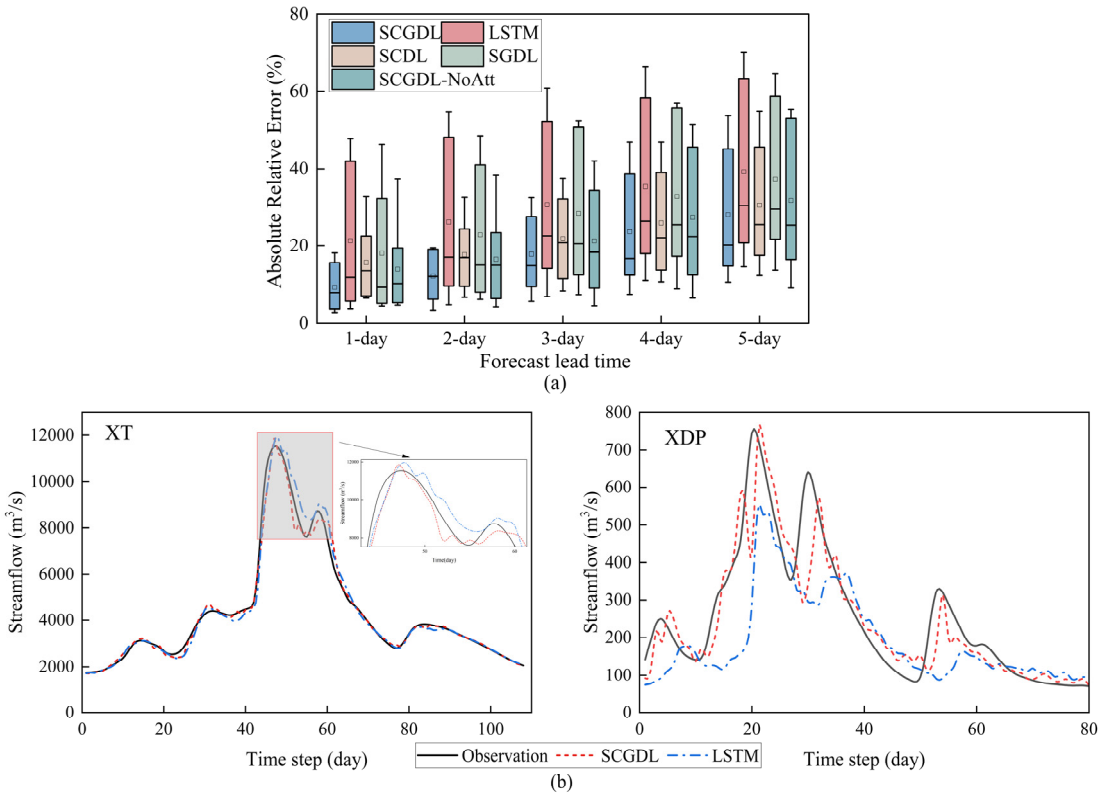


Fig. 7. Performance of high flow simulation. (a) Boxplot of the absolute relative error in simulating

514 the top 5% of flow. (b) Comparison of the SCGDL and the LSTM flood simulation.

515 **4.2 Experiment II: adaptability to spatial precipitation inputs of varied quality**
516 **and with different characteristics**

517 Table 4 shows regardless of the quality and characteristics of the precipitation inputs,
518 most NSEs are approximately or greater than 0.8. These findings indicate the strong
519 adaptability of the SCGDL model. The precipitation from ERA5-Land reanalysis
520 contributes to the best accuracy. The TCMerge precipitation, which blends ERA5,
521 PCDR and CHIRPS025, leverages the strengths of its source products, resulting in
522 higher accuracy in streamflow simulation. An inconsistency is observed between the
523 accuracy of precipitation products and the resulting streamflow simulation. For
524 example, although ERA 5 precipitation exhibits the worst accuracy in CC, KGE and
525 RMSE (Fig. 8), it still outperforms more than half of the selected products in the
526 streamflow simulation. The relationship between the precipitation product accuracy and
527 the performance in DL-based hydrological modeling is not straightforward, as
528 evidenced by insignificant correlations depicted in Fig. 8. Furthermore, comparing the
529 products with different spatial resolutions reveals that higher resolution does not
530 necessarily equate to better hydrological performance. For eample, the PCCSCDR with
531 the highest resolution, ranks at the bottom instead. Among precipitation inputs sourced
532 from the same origin but featuring different spatial resolutions, only the fine-grained
533 ERA5-land outperforms the less refined ERA5. Consequently, spatial resolution is a
534 non-critical factor influencing simulation outcomes.

Table 4

Summary of the streamflow simulation based on different precipitation products for the testing period.

Type	Product	Spatial resolution	Involved grid number	NSE	Rank
Reanalysis	ERA5-Land	0.1°	39*42	0.889	1
	ERA5	0.25°	16*17	0.848	3
Merging	TCMerge	0.25°	16*17	0.867	2
Satellite	IMERG-F	0.1°	39*42	0.847	4
	PCDR	0.25°	16*17	0.831	5
	CHIRPS025	0.25°	16*17	0.824	6
	CHIRPS005	0.05°	76*82	0.820	7
	IMERG-E	0.1°	39*42	0.809	8
	PCCSCDR	0.04°	96*102	0.798	9

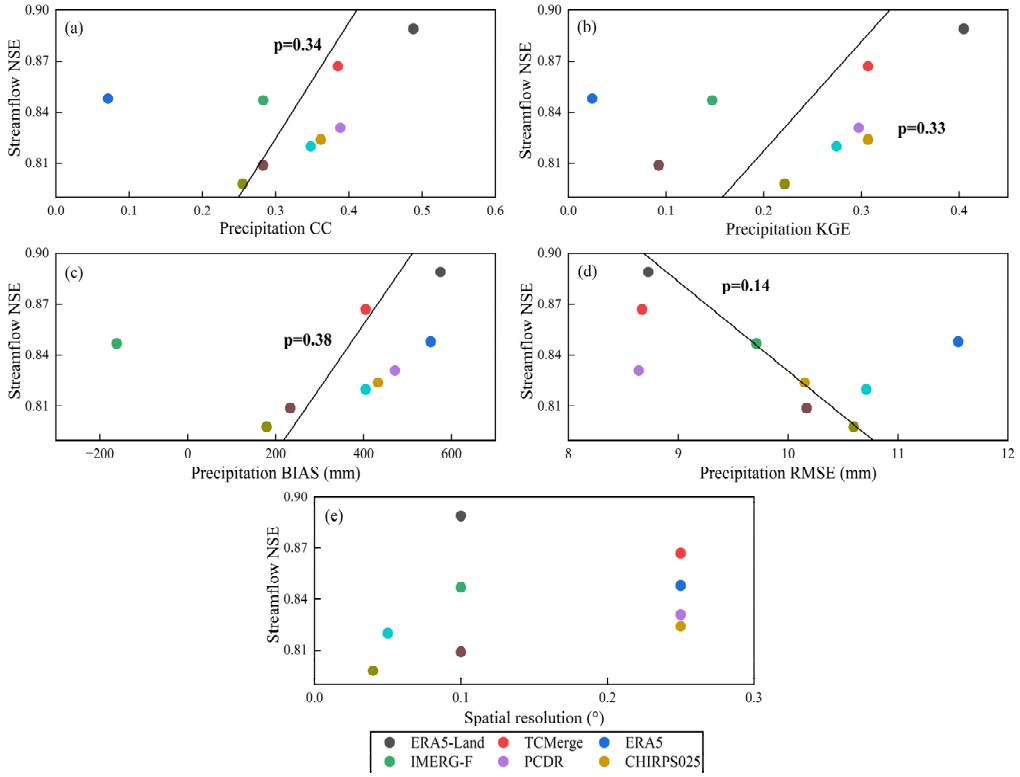


Fig. 8. Relationship between the precipitation (a) CC, (b) KGE, (c) RMSE, (d) BIAS, (e) spatial resolution, and the resulting simulated streamflow performance (NSE). p denotes the significance level of linear regressions. The accuracy is evaluated between the reference precipitation observations and each precipitation product.

Fig. 9 shows the spatial variation of timestep-averaged NSE over 5 simulation steps at each stream gauge. For the gauges in the mainstream, i.e. the HJG, XY, HZ, SY and XT gauges, their NSEs are comparatively similar regardless of the precipitation products used as inputs. Conversely, the NSE distribution of the tributary gauges, i.e. the GC, XDP and GT gauges, displays a wider range, with the highest value reaching around 0.8 and the lowest falling below 0.5. This suggests that tributary gauges are more susceptible to variations in precipitation inputs.

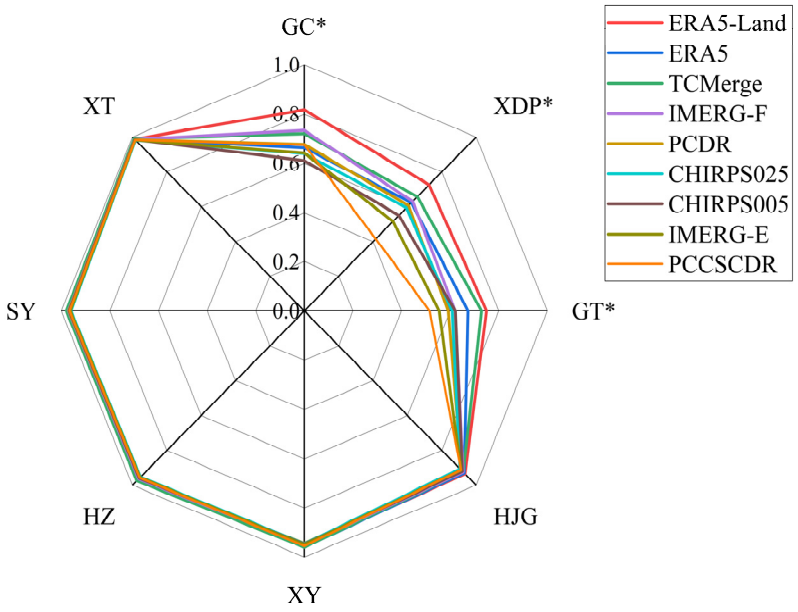


Fig. 9. Spatial variation of the timestep-averaged NSE over 5 simulation steps of each stream gauge obtained by the models using different precipitation products as inputs. The tributary gauges are marked with asterisks. The mainstream gauges are arranged clockwise from upstream to

554 downstream.

555 **4.3 Experiment III: response to topological connectivity with different graph**
556 **representations**

557 To further uncover the influence of the topological connectivity, different graph
558 representations are fed into the proposed SCGDL model independently. We first assess
559 whether the direction and weight of adjacency matrices would affect the simulation. To
560 this end, two additional comparative models are established, each utilizing different
561 adjacency matrices. SCGDL-UD denotes a model using an undirected and unweighted
562 adjacency matrix and SCGDL-W denotes a model using a directed and weighted
563 adjacent matrix. The undirected adjacency matrix considers dual upstream-downstream
564 relationships among stream gauges and is formed symmetrically. The weighted adjacent
565 matrix is generated using the physical distance among the gauges, with a thresholded
566 Gaussian kernel (Li et al., 2017; Shuman et al., 2013). The comparison presented in
567 Table 5 reveals minimal differences among the three models, suggesting that the
568 direction and weight of graphs exert limited influence on performance in our case. Next,
569 the influence of the topological connectivity with different degrees of physical realism
570 is investigated by generating two additional graphs. The first graph excludes the
571 tributary gauges and only contains five gauges in the mainstream. For countries such
572 as China with limited access to streamflow observations, collecting data from tributary
573 gauges can be more challenging than from mainstream gauges. Therefore, it is worth
574 investigating to what extent the model could interpolate the tributary streamflow using

575 the trained models without the inclusion of tributary historical streamflow data. The
576 second graph is more simplified by removing the SY gauge which does not require
577 aggregation of tributary information, resulting in a configuration consisting of merely
578 four gauges. This progressive simplification diminishes the physical realism of the
579 graphs continuously. The corresponding comparative models are denoted as SCGDL-5
580 and SCGDL-4 respectively, with predictive targets remaining consistent with SCGDL.
581 Streamflow simulation based on the simplified graphs is basically an interpolation for
582 the gauges assumed to be unobservable. The observable mainstream gauges are
583 expected to possess dominant information due to confluence processes and thereby
584 facilitate interpolation for unobservable gauges. Our findings reveal that as the
585 agreement of the graph structure with physical realism weakens, the model's
586 performance deteriorates. Fig. 10 shows the gauge-level comparisons of the NSE.
587 Irrespective of the graph simplification level, the mainstream gauges exhibit high
588 predictability with simulation accuracy nearly identical across comparative models. In
589 contrast, the simulation of the tributary gauges experiences performance deterioration
590 attributed to information loss during the interpolation process. This decline in
591 evaluation metrics predominantly accounts for the overall poorer performance of the
592 SCGDL-5 and SCGDL-4. It indicates the sensitivity of tributary gauge modeling to the
593 physical realism of graphs.

594 **Table 5**

595 Summary of the evaluation metrics of the models utilizing different graph representations of

topological connectivity.

	SCGDL	SCGDL-UD	SCGDL-W	SCGDL-5	SCGDL-4
NSE	0.889	0.882	0.886	0.874	0.863
CC	0.943	0.941	0.941	0.935	0.929
KGE	0.908	0.904	0.910	0.894	0.882

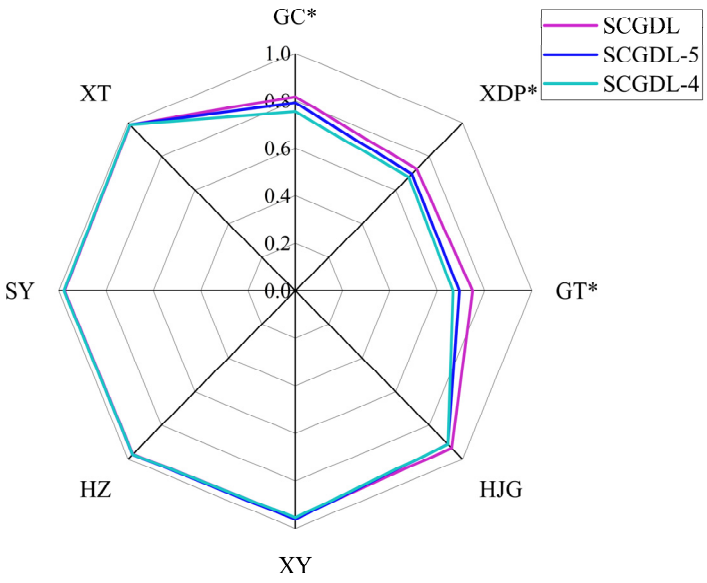


Fig. 10. Comparisons of the models employing different graph representations based on the NSE at each stream gauge. The SCGDL uses complete gauges to generate graphs. The graph of SCGDL-5 excludes the tributary gauges and contains only five gauges in the mainstream. The graph of SCGDL-4 further removes the SY gauge which does not require aggregation of tributary information resulting in a configuration of four gauges These graphs are gradually simplified, leading to a continuous decline in physical realism. The tributary gauges are marked with asterisks, while the mainstream gauges are arranged clockwise from upstream to downstream.

5. Discussion

Our model achieves superior performance acroos the temporal, spatial and spatiotemporal domains. In the temporal domain, performance degradation with increasing lead time is a long-standing issue (Liang et al., 2020). As a result, there is limited potential for model enhancement in this aspect. However, the SCGDL exhibits slower performance degradation and increasing percentage improvements with lead time compared with the baseline model. In other words, it yields more stable and robust multi-step streamflow simulation. Regarding the spatial domain, the mainstream gauges in the lower reach, i.e. the XY, HZ, SY and XT gauges, exhibit high predictability and even the baseline model performs well. As a result, there is limited potential for model enhancement in this aspect. This observation aligns with the findings from other research. In the experiments conducted by Liu et al. (2022) and Liu et al. (2023), their focus lies on gauges located in the mainstream of the Yangtze River, known for its substantial streamflow volumes. Although their proposed models showed effectiveness, the performance enhancements were relatively less pronounced due to the limited space for accuracy improvement. On the contrary, the integration of different kinds of spatial information proves more beneficial for gauges posing greater predictive challenges, such as those located at the outlet of the reservoir and in the tributaries in the study area. Modeling tributary streamflow presents challenges due to its typically lower flow magnitude and heightened susceptibility to natural and anthropogenic influences (Tamaddun et al., 2019; Tongal and Sivakumar, 2019). Previous research

has shown that larger river basins with greater streamflow abundance tend to achieve higher performance using process-based hydrological models (Zhang et al., 2021). The performance disparities observed between the mainstream and tributary gauges in our study suggest the presence of an implicit spatial scale effect in spatiotemporal DL-based simulation. Previous research has demonstrated that the inclusion of basin and river attributes can enhance hydrological simulation performance in both natural and regulated basins (Kratzert et al., 2019; Mangukiya et al., 2023; Tursun et al., 2024). Static attributes allow DL models to better distinguish between hydrological processes across various basins. However, contrary to previous multi-catchment LSTM studies, our study employs spatial meteorological forcing inputs, with multiple gridded values in each timestep. In the spatial case, repeating attributes for each grid and timestep may cause data redundancy and deteriorate simulation performance. Future research could potentially combine 1d historical streamflow data and static attributes into node attributes and use GNN-based modules to process them (Sun et al., 2021).

While existing research has extensively evaluated the hydrological applicability of precipitation products from diverse sources using process-based hydrological models, there has been limited investigation into how spatiotemporal DL-based models respond to these products. Despite the distinct mechanisms of process-based and DL-based models in simulating streamflow, we find similarities in their behavior when employing gridded precipitation products of varied quality as inputs.. First, we observe no significant relationship between the spatial resolution of products and the

corresponding hydrological performance (Satgé et al., 2021; Wei et al., 2020; Zhang et al., 2022). While finer spatial resolution might enhance the representation of precipitation heterogeneity within a river basin, it is probably not the critical factor to achieve better hydrological performance compared to better dependencies, as DL-based models typically achieve satisfactory performance by extracting highly dependent characteristics between input and output variables (Du et al., 2021; Xu and Liang, 2021). To verify this, normalized mutual information (NMI) indices of the coarser-resolution ERA5 (0.25°), IMERG-F (0.1°) and the finer-resolution PCCSCDR (0.04°) to the streamflow observations were quantified to measure the dependencies for the testing period (Fig. 11) (Dionisio et al., 2004; Jiang et al., 2023; Shin et al., 2023). The NMI of each product is computed by averaging the NMIs between the precipitation at each contributing grid and the streamflow at the gauge. We find that the ERA5 and IMERG-F, despite their coarser resolution, exhibit stronger dependencies on the observed streamflow of each gauge, leading to higher NSEs compared to the PCCSCDR. Therefore, dependency is a more critical factor in DL-based streamflow simulation, emphasizing the importance of evaluating dependencies prior to employing fine-resolution products of precipitation and other remotely sensed variables. Second, the discrepancy in streamflow prediction between precipitation products tends to be reduced over large-scale basins (Ji et al., 2024). This indicates the potential presence of a spatial scale effect with different precipitation products. Third, the inconsistent performance between precipitation analysis and hydrological modeling is observed.

While it may seem intuitive that precipitation products of better quality generally lead to better streamflow simulation, it is not always the case when using process-based models due to differences in the study area and models (Ang et al., 2022; Le et al., 2024; Qi et al., 2016; Wu et al., 2023). More accurate precipitation products do not always lead to better hydrological performance. When simulating streamflow using the DL-based model in our case, we encounter similar phenomena. This inconsistency may arise from the interaction between models and data products, suggesting that streamflow simulation performance relies not only on the quality of precipitation products but also on their synergy with the models. For example, the streamflow performance would be influenced by the model's adaptability to compensate for precipitation errors through calibrating or training parameters. Wang et al. (2023) investigated the parameters of a Xinanjiang model to explore why the model adapted to inaccurate precipitation inputs. However, the inherent black-box mechanism in DL-based models poses challenges in directly tracking the parameters to uncover how these models compensate for errors. In future work, explainable artificial intelligence methods might provide a promising alternative to explore the adaptability of DL-based modeling. It is noted that this work considered potential evaporation as a control variable and identified the influence of precipitation independently. If different potential evaporation products are utilized as well, their independent influence and joint impact with precipitation on hydrological performance need to be investigated (Wang et al., 2023).

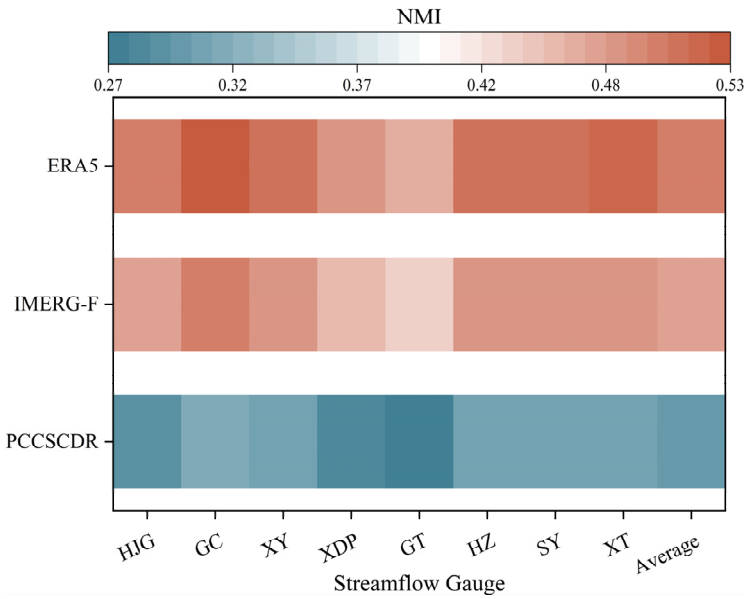


Fig. 11. NMI of the coarser-resolution ERA5 (0.25°), IMERG-F (0.1°) and the finer-resolution PCCSCDR (0.04°) to the streamflow observations, at each gauge. The closer the NMI is to 1, the more dependencies they have. The NMI of each product is computed by averaging the NMIs between the precipitation values at each contributing grid cell and the streamflow at the gauge. The last column shows the average across the gauges for each precipitation product.

Graph generation often relies on the hydrological understanding of model developers and neglects careful consideration of weight and direction. Some developers assumed adjacent gauges shared similar hydrological processes, resulting in the creation of undirected graphs (Sun et al., 2022; Topp et al., 2023). Conversely, others thought information transferring in river networks should follow a one-way flow path from upstream to downstream, leading to the generation of directed graphs (Chen et al., 2021; Moshe et al., 2020). Our findings suggest the inclusion or exclusion of graph weight and direction fail to yield significant differences in performance, which might be attributed to the strong ability of GNN-based modules to learn non-linear features,

particularly the adaptability of learnable parameters W in Eq. (5) to compensate for the differences in adjacency matrices. It's worth noting that this result is case-dependent as we only tested the performance in the ML-HRB, which presents a preliminary reference for a broader sensitivity study involving larger samples. As we vary the degree of physical realism, the mainstream gauges consistently yield robust results, while the tributary gauges prove more susceptible to variations in the graph structure. This finding enhances the existing understanding of the degree of physical realism needed for river network modeling, which was made by Sun et al. (2022) in a different kind of study area with abundant data for densely distributed nodes. However, due to data limitations, a small number of streamflow nodes are sparsely distributed in our study area, as is common in many other regions. Our work suggests that regardless of whether a river network contains densely or sparsely distributed nodes, achieving an accurate representation of complete physical realism is essential when graph learning particularly emphasizes tributaries, such as mountain torrents. In cases of limited access to streamflow observations in subbasins, our findings offers an opportunity to interpolate these subbasins using the trained models with simplified graphs and without the need for their streamflow observations as inputs, because the accuracy of interpolated streamflow does not substantially decline. Therefore, it highlights the varying degree of physical realism needed for different objectives.

6. Conclusion

This study introduces a novel spatiotemporal DL model, named SCGDL, that integrates both Euclidean and non-Euclidean spatial features, namely regularly gridded meteorological forcing data and irregular topological connectivity, respectively. The proposed SCGDL model is applied in the ML-HRB to produce multi-step streamflow simulation up to 5 days ahead at multiple stream gauges simultaneously.

Demonstrating superiority over comparative models, the SCGDL excels in temporal, spatial, and spatiotemporal performance, aligning with our initial hypothesis regarding the benefits of integrating Euclidean and non-Euclidean spatial data for hydrological modeling. Temporally, the SCGDL simulation exhibits greater stability with extended lead time. Spatially, the primary contribution to the performance of the SCGDL lies in its superior ability to predict smaller rates of streamflow at tributary gauges. We further explore the adaptability and response of the spatiotemporal DL-based model to gridded precipitation inputs of varied quality and topological connectivity with different graph representations. We find the simulations of the tributary gauges are more sensitive to changes in spatial inputs, whereas those of the mainstem gauges tend to be more stable, suggesting the presence of an implicit spatial scale effect in DL-based streamflow simulation with different model structures and spatial inputs.

Although process-based and DL-based models employ distinct mechanisms for streamflow simulation, we observe shared hydrological responses to precipitation products. First, spatial resolution of precipitation products is not a critical factor

affecting predictive skill. Second, inconsistencies in performance occur between precipitation analysis and hydrological simulation. This highlights the significance of evaluating hydrological performance of different precipitation products before their application in process-based and DL-based hydrological simulations. It indicates the performance of streamflow simulation depends not only on product quality but also on the interaction between models and products. Note that if different potential evaporation products are also incorporated, further attention should be paid to their independent influence and joint impact with precipitation on hydrological performance. Our investigation reveals that direction and weight of graphs have limited influence on simuation, while degree of physical realism required for graph learning varies for different objectives. In cases of limited access to streamflow observations in subbasins, leveraging trained models with simplified graphs presents an opportunity for interpolation. While for simulation particularly concentrating on tributaries like mountain torrents within small basins, achieving an accurate representation that embodys complete physical realism is essential.

The proposed model provides valuable insight into improving performance of spatiotemporal DL-based simulation for other hydrological variables, such as water level and stream temperature, which similarly rely on Euclidean and non-Euclidean spatial information. Meanwhile, some limitations require further investigation, such as inspecting whether information learned by spatial-attention mechanisms aligns with hydrological behaviors and quantifying influence of various spatial inputs, model

766 structures and their interactions on uncertainties of streamflow simulation.

767

CRediT authorship contribution statement

Liangkun Deng: Conceptualization, Methodology, Writing – original draft. **Xiang Zhang:** Supervision, Validation, Writing – review & editing. **Louise J. Slater:** Writing – review & editing. **Haoyuan Liu:** Investigation. **Shiyong Tao:** Data curation.

Data availability

The ERA5 and the ERA5-Land dataset can be freely downloaded from the C3S Climate Data Store (https://cds.climate.copernicus.eu/#!/home; accessed on 30 March 2023). The CHIRPS family of products is publicly available (at https://www.chc.ucsb.edu/data/chirps; accessed on 30 March 2023). The IMERG-E and IMERG-F datasets can be freely downloaded from the NASA Goddard Earth Sciences (GES) Data and Information Services Center (https://disc.gsfc.nasa.gov/; accessed on 30 March 2023). The PCDR and PCCSCDR datasets are publicly available (at http://chrsdata.eng.uci.edu/; accessed on 30 March 2023). Streamflow observations can be made available on request.

Declaration of Competing Interest

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

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China, grant number: No.U21A2002. We also appreciate the editors and anonymous reviewers for enhancing our paper.

References

Anderson, S., Radić, V., 2022. Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. Hydrol. Earth Syst. Sci. 26 (3), 795-825.

Ang, R., Kinouchi, T., Zhao, W., 2022. Evaluation of daily gridded meteorological datasets for hydrological modeling in data-sparse basins of the largest lake in southeast Asia. Journal of Hydrology: Regional Studies 42, 101135.

Ashouri, H., Hsu, K., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D., Nelson, B.R., Prat, O.P., 2015. PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull. Amer. Meteorol. Soc. 96 (1), 69-83.

Bai, T., Tahmasebi, P., 2023. Graph neural network for groundwater level forecasting. J. Hydrol. 616, 128792.

Bárdossy, A., Anwar, F., 2023. Why do our rainfall–runoff models keep underestimating the peak flows? Hydrol. Earth Syst. Sci. 27 (10), 1987-2000.

Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q.V., 2019. Attention augmented convolutional networks, in, Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3286-3295.

Chen, C., Jiang, J., Liao, Z., Zhou, Y., Wang, H., Pei, Q., 2022. A short-term flood prediction based on spatial deep learning network: a case study for xi county, china. J. Hydrol. 607, 127535.

Chen, S., Appling, A., Oliver, S., Corson-Dosch, H., Read, J., Sadler, J., Zwart, J., Jia, X., 2021. Heterogeneous stream-reservoir graph networks with data assimilation, in, 2021 IEEE International Conference on Data Mining (ICDM). IEEE, pp. 1024-1029.

Cheng, B., Xia, R., Zhang, Y., Yang, Z., Hu, S., Guo, F., Ma, S., 2019. Characterization and causes

811 analysis for algae blooms in large river system. Sust. Cities Soc. 51, 101707.

812 Coopersmith, E., Yaeger, M.A., Ye, S., Cheng, L., Sivapalan, M., 2012. Exploring the physical

813 controls of regional patterns of flow duration curves–part 3: a catchment classification system based

814 on regime curve indicators. Hydrol. Earth Syst. Sci. 16 (11), 4467-4482.

815 Demir, I., Conover, H., Krajewski, W.F., Seo, B., Goska, R., He, Y., Mceniry, M.F., Graves, S.J.,

816 Petersen, W., 2015. Data-enabled field experiment planning, management, and research using

817 cyberinfrastructure. J. Hydrometeorol. 16 (3), 1155-1170.

818 Deng, L., Zhang, X., Tao, S., Zhao, Y., Wu, K., Liu, J., 2023. A spatiotemporal graph convolution-

819 based model for daily runoff prediction in a river network with non-Euclidean topological structure.

820 Stoch. Environ. Res. Risk Assess. 37 (4), 1457-1478.

821 Dionisio, A., Menezes, R., Mendes, D.A., 2004. Mutual information: a measure of dependency for

822 nonlinear time series. Physica A: Statistical Mechanics and its Applications 344 (1-2), 326-329.

823 Du, B., Zhou, Q., Guo, J., Guo, S., Wang, L., 2021. Deep learning with long short-term memory

824 neural networks combining wavelet transform and principal component analysis for daily urban

825 water demand forecasting. Expert Syst. Appl. 171, 114571.

826 Frizzi, S., Bouchouicha, M., Ginoux, J.M., Moreau, E., Sayadi, M., 2021. Convolutional neural

827 network for smoke and fire semantic segmentation. IET Image Process. 15 (3), 634-647.

828 Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J.,

829 Harrison, L., Hoell, A., 2015. The climate hazards infrared precipitation with stations—a new

830 environmental record for monitoring extremes. Sci. Data 2 (1), 1-21.

831 Gassman, P.W., Reyes, M.R., Green, C.H., Arnold, J.G., 2007. The soil and water assessment tool:

historical development, applications, and future research directions. Trans. ASABE 50 (4), 1211-1250.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J., Hochreiter, S., 2021. Rainfall–runoff prediction at multiple timescales with a single long short-term memory network. Hydrol. Earth Syst. Sci. 25 (4), 2045-2062.

Gauch, M., Lin, J., 2020. A data scientist's guide to streamflow prediction. arXiv preprint arXiv:2006.12975

Ghotbi, S., Wang, D., Singh, A., Mayo, T., Sivapalan, M., 2020. Climate and landscape controls of regional patterns of flow duration curves across the continental united states: statistical approach. Water Resour. Res. 56 (11), e2020WR028041.

Gu, L., Yin, J., Wang, S., Chen, J., Qin, H., Yan, X., He, S., Zhao, T., 2023. How well do the multi-satellite and atmospheric reanalysis products perform in hydrological modelling. J. Hydrol. 617, 128920.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T.R., Kriaučiūnienė, J., Kundzewicz, Z.W., Lang, M., 2014. Understanding flood regime changes in europe: a state-of-the-art assessment. Hydrol. Earth Syst. Sci. 18 (7), 2735-2772.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., 2020. The era5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999-2049.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132-7141.

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019a. GPM IMERG Early precipitation L3 1 day 0.1 degree x 0.1 degree v06. In: Andrey Savtchenko, Greenbelt, M.D. (Eds.), Goddard Earth Sciences Data and Information Services Center (GES DISC)

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Tan, J., 2019b. GPM IMERG Final precipitation L3 1 day 0.1 degree x 0.1 degree v06. In: Andrey Savtchenko, Greenbelt, M.D. (Eds.), Goddard Earth Sciences Data and Information Services Center (GES DISC)

Hussain, D., Hussain, T., Khan, A.A., Naqvi, S.A.A., Jamil, A., 2020. A deep learning approach for hydrological time-series prediction: a case study of Gilgit river basin. Earth Sci. Inform. 13, 915-927.

Hussain, Y., Satgé, F., Hussain, M.B., Martinez-Carvajal, H., Bonnet, M., Cárdenas-Soto, M., Roig, H.L., Akhter, G., 2018. Performance of CMORPH, TMPA, and PERSIANN rainfall datasets over plain, mountainous, and glacial regions of Pakistan. Theor. Appl. Climatol. 131, 1119-1132.

Ji, P., Yuan, X., Jiao, Y., Zhang, M., 2024. On the reliability of 12 high-resolution precipitation products for process-based hydrological modeling in china. J. Hydrol. 628, 130598.

Jiang, P., Shuai, P., Sun, A., Mudunuru, M.K., Chen, X., 2023. Knowledge-informed deep learning for hydrological model calibration: an application to coal creek watershed in colorado. Hydrol. Earth Syst. Sci. 27 (14), 2621-2644.

Jiang, S., Zheng, Y., Babovic, V., Tian, Y., Han, F., 2018. A computer vision-based approach to fusing spatiotemporal data for hydrological modeling. J. Hydrol. 567, 25-40.

Kim, T., Shin, J., Cha, Y., 2023. Incorporation of feature engineering and attention mechanisms into deep learning models to develop an early warning system for harmful algal blooms. J. Clean Prod.,

137564.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22 (11), 6005-6022.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., Nearing, G.S., 2019. Toward improved predictions in ungauged basins: exploiting the power of machine learning. Water Resour. Res. 55 (12), 11344-11354.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23 (12), 5089-5110.

Le, X., Kim, Y., Van Binh, D., Jung, S., Nguyen, D.H., Lee, G., 2024. Improving rainfall-runoff modeling in the mekong river basin using bias-corrected satellite precipitation products by convolutional neural networks. J. Hydrol., 130762.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436-444.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278-2324.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., Dadson, S.J., 2021. Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. Hydrol. Earth Syst. Sci. 25 (10), 5517-5534.

Li, C., Tang, G., Hong, Y., 2018. Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: applicability of the triple collocation method across mainland china. J.

Hydrol. 562, 71-83.

Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv preprint arXiv:1707.01926

Liang, Z., Zou, R., Chen, X., Ren, T., Su, H., Liu, Y., 2020. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. J. Hydrol. 581, 124432.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., Dai, Y., 2021. A hybrid deep learning algorithm and its application to streamflow prediction. J. Hydrol. 601, 126636.

Liu, G., Ouyang, S., Qin, H., Liu, S., Shen, Q., Qu, Y., Zheng, Z., Sun, H., Zhou, J., 2023. Assessing spatial connectivity effects on daily streamflow forecasting using Bayesian-based graph neural network. Sci. Total Environ. 855, 158968.

Liu, Y., Hou, G., Huang, F., Qin, H., Wang, B., Yi, L., 2022. Directed graph deep neural network for multi-step daily streamflow forecasting. J. Hydrol. 607, 127515.

Mangukiya, N.K., Sharma, A., Shen, C., 2023. How to enhance hydrological predictions in hydrologically distinct watersheds of the Indian subcontinent? Hydrol. Process. 37 (7), e14936.

Marçais, J., de Dreuzy, J.R., 2017. Prospective interest of deep learning for hydrological inference. Groundwater 55 (5), 688-692.

Marsh, C.B., Pomeroy, J.W., Wheater, H.S., 2020. The Canadian Hydrological Model (CHM) v1. 0: a multi-scale, multi-extent, variable-complexity hydrological model-design and overview. Geosci. Model Dev. 13 (1), 225-247.

916 Moshe, Z., Metzger, A., Elidan, G., Kratzert, F., Nevo, S., El-Yaniv, R., 2020. Hydronets:
917 leveraging river structure for hydrologic modeling. arXiv preprint arXiv:2007.00595
918 Muñoz, P., Corzo, G., Solomatine, D., Feyen, J., Célleri, R., 2023. Near-real-time satellite
919 precipitation data ingestion into peak runoff forecasting models. Environ. Modell. Softw. 160,
920 105582.
921 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G.,
922 Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., 2021. Era5-land: a state-of-the-art global
923 reanalysis dataset for land applications. Earth Syst. Sci. Data 13 (9), 4349-4383.
924 Nearing, G.S., Ruddell, B.L., Bennett, A.R., Prieto, C., Gupta, H.V., 2020. Does information theory
925 provide a new paradigm for earth science? Hypothesis testing. Water Resour. Res. 56 (2),
926 e2019WR024918.
927 Nijssen, B., Lettenmaier, D.P., 2004. Effect of precipitation sampling error on simulated
928 hydrological fluxes and states: anticipating the global precipitation measurement satellites. Journal
929 of geophysical research: atmospheres 109 (D2)
930 Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J.A., Orts-Escolano, S.,
931 Garcia-Rodriguez, J., Argyros, A., 2020. A review on deep learning techniques for video prediction.
932 IEEE Trans. Pattern Anal. Mach. Intell. 44 (6), 2806-2826.
933 Pan, B., Hsu, K., Aghakouchak, A., Sorooshian, S., 2019. Improving precipitation estimation using
934 convolutional neural network. Water Resour. Res. 55 (3), 2301-2321.
935 Qi, W., Zhang, C., Fu, G., Sweetapple, C., Zhou, H., 2016. Evaluation of global fine-resolution
936 precipitation products and their uncertainty quantification in ensemble discharge simulations.

Hydrol. Earth Syst. Sci. 20 (2), 903-920.

Sadeghi, M., Nguyen, P., Naeini, M.R., Hsu, K., Braithwaite, D., Sorooshian, S., 2021. PERSIANN-CCS-CDR, a 3-hourly 0.04 global precipitation climate data record for heavy precipitation studies. Sci. Data 8 (1), 157.

Sarkar, R., Dutta, S., Dubey, A.K., 2015. An insight into the runoff generation processes in wet sub-tropics: field evidences from a vegetated hillslope plot. Catena 128, 31-43.

Satgé, F., Pillot, B., Roig, H., Bonnet, M., 2021. Are gridded precipitation datasets a good option for streamflow simulation across the Juruá river basin, amazon? J. Hydrol. 602, 126773.

Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., Woo, W., 2015. Convolutional lstm network: a machine learning approach for precipitation nowcasting. Advances in neural information processing systems 28

Shin, C., Dirmeyer, P.A., Huang, B., 2023. A joint approach combining correlation and mutual information to study land and ocean drivers of us droughts: methodology. J. Clim. 36 (9), 2795-2814.

Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P., 2013. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process. Mag. 30 (3), 83-98.

Speight, L.J., Cranston, M.D., White, C.J., Kelly, L., 2021. Operational and emerging capabilities for surface water flood forecasting. Wiley Interdisciplinary Reviews: Water 8 (3), e1517.

Sun, A.Y., Jiang, P., Mudunuru, M.K., Chen, X., 2021. Explore spatio-temporal learning of large sample hydrology using graph neural networks. Water Resour. Res. 57 (12), e2021WR030394.

Sun, A.Y., Jiang, P., Yang, Z., Xie, Y., Chen, X., 2022. 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 (19), 5163-5184.

Tamaddun, K.A., Kalra, A., Ahmad, S., 2019. Spatiotemporal variation in the continental us streamflow in association with large-scale climate signals across multiple spectral bands. Water Resour. Manag. 33, 1947-1968.

Tao, S., Zhang, H., Yang, F., Wu, Y., Li, C., 2023. Multiple information spatial–temporal attention based graph convolution network for traffic prediction. Appl. Soft. Comput. 136, 110052.

Tongal, H., Sivakumar, B., 2019. Entropy analysis for spatiotemporal variability of seasonal, low, and high streamflows. Stoch. Environ. Res. Risk Assess. 33, 303-320.

Topp, S.N., Barclay, J., Diaz, J., Sun, A.Y., Jia, X., Lu, D., Sadler, J.M., Appling, A.P., 2023. Stream temperature prediction in a shifting environment: explaining the influence of deep learning architecture. Water Resour. Res. 59 (4), e2022WR033880.

Trebing, K., Stańczyk, T., Mehrkanoon, S., 2021. SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. Pattern Recognit. Lett. 145, 178-186.

Tu, T., Ishida, K., Ercan, A., Kiyama, M., Amagasaki, M., Zhao, T., 2021. Hybrid precipitation downscaling over coastal watersheds in japan using WRF and CNN. Journal of Hydrology: Regional Studies 37, 100921.

Tursun, A., Xie, X., Wang, Y., Liu, Y., Peng, D., Zheng, B., 2024. Enhancing streamflow simulation in large and human-regulated basins: long short-term memory with multiscale attributes. J. Hydrol. 630, 130771.

Van, S.P., Le, H.M., Thanh, D.V., Dang, T.D., Loc, H.H., Anh, D.T., 2020. Deep learning convolutional neural network in rainfall–runoff modelling. J. Hydroinform. 22 (3), 541-561.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., 2017. Graph attention networks. arXiv preprint arXiv:1710.10903

Wang, C., Li, X., Li, H., 2022. Role of input features in developing data-driven models for building thermal demand forecast. Energy Build., 112593.

Wang, H., Li, S., Song, L., Cui, L., 2019. A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. Comput. Ind. 105, 182-190.

Wang, J., Zhuo, L., Han, D., Liu, Y., Rico Ramirez, M.A., 2023. Hydrological model adaptability to rainfall inputs of varied quality. Water Resour. Res. 59 (2), e2022WR032484.

Wang, J., Zhuo, L., Rico Ramirez, M.A., Abdelhalim, A., Han, D., 2023. Interacting effects of precipitation and potential evapotranspiration biases on hydrological modeling. Water Resour. Res. 59 (3), e2022WR033323.

Wang, Q., Xia, J., She, D., Zhang, X., Liu, J., Zhang, Y., 2021. Assessment of four latest long-term satellite-based precipitation products in capturing the extreme precipitation and streamflow across a humid region of southern china. Atmos. Res. 257, 105554.

Wei, Z., He, X., Zhang, Y., Pan, M., Sheffield, J., Peng, L., Yamazaki, D., Moiz, A., Liu, Y., Ikeuchi, K., 2020. Identification of uncertainty sources in quasi-global discharge and inundation simulations using satellite-based precipitation products. J. Hydrol. 589, 125180.

Weyn, J.A., Durran, D.R., Caruana, R., 2019. Can machines learn to predict weather? Using deep learning to predict gridded 500-hpa geopotential height from historical weather data. J. Adv. Model.

Earth Syst. 11 (8), 2680-2693.

Woo, S., Park, J., Lee, J., Kweon, I.S., 2018. Cbam: convolutional block attention module, in, Proceedings of the European conference on computer vision (ECCV), pp. 3-19.

Wu, H., Yang, Q., Liu, J., Wang, G., 2020. A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. J. Hydrol. 584, 124664.

Wu, X., Su, J., Ren, W., Lü, H., Yuan, F., 2023. Statistical comparison and hydrological utility evaluation of ERA5-Land and IMERG precipitation products on the Tibetan Plateau. J. Hydrol. 620, 129384.

Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C., 2019. Graph wavenet for deep spatial-temporal graph modeling. arXiv preprint arXiv:1906.00121

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., Goldscheider, N., 2022. Karst spring discharge modeling based on deep learning using spatially distributed input data. Hydrol. Earth Syst. Sci. 26 (9), 2405-2430.

Xie, K., Liu, P., Zhang, J., Han, D., Wang, G., Shen, C., 2021. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. J. Hydrol. 603, 127043.

Xu, T., Liang, F., 2021. Machine learning for hydrologic sciences: an introductory overview. Wiley Interdisciplinary Reviews: Water 8 (5), e1533.

Xu, X., Frey, S.K., Ma, D., 2022. Hydrological performance of era5 and merra-2 precipitation products over the great lakes basin. Journal of Hydrology: Regional Studies 39, 100982.

Yan, L., Feng, J., Hang, T., Zhu, Y., 2021. Flow interval prediction based on deep residual network

and lower and upper boundary estimation method. Appl. Soft. Comput. 104, 107228.

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., Li, F., 2021. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. Expert Syst. Appl. 169, 114513.

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W., Zhao, B., 2020. A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. J. Hydrol. 590, 125206.

Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K., 2019. An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387-408.

Yuan, F., Zhang, L., Win, K.W.W., Ren, L., Zhao, C., Zhu, Y., Jiang, S., Liu, Y., 2017. Assessment of GPM and TRMM multi-satellite precipitation products in streamflow simulations in a data-sparse mountainous watershed in Myanmar. Remote Sens. 9 (3), 302.

Zanfei, A., Brentan, B.M., Menapace, A., Righetti, M., Herrera, M., 2022. Graph convolutional recurrent neural networks for water demand forecasting. Water Resour. Res. 58 (7), e2022WR032299.

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137-152.

Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: an extremely efficient convolutional neural network for mobile devices, in, Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition, pp. 6848-6856.

Zhang, Y., Wu, C., Yeh, P.J., Li, J., Hu, B.X., Feng, P., Jun, C., 2022. Evaluation and comparison of precipitation estimates and hydrologic utility of chirps, TRMM 3b42 v7 and PERSIANN-CDR products in various climate regimes. Atmos. Res. 265, 105881.

Zhang, Y., Ye, A., Nguyen, P., Analui, B., Sorooshian, S., Hsu, K., 2021. Error characteristics and scale dependence of current satellite precipitation estimates products in hydrological modeling. Remote Sens. 13 (16), 3061.

Zhang, Z., Zhang, Q., Singh, V.P., 2018. Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. Hydrological Sciences Journal 63 (7), 1091-1111.

Zhao, Q., Zhu, Y., Shu, K., Wan, D., Yu, Y., Zhou, X., Liu, H., 2020. Joint spatial and temporal modeling for hydrological prediction. IEEE Access 8, 78492-78503.

Zhao, R., 1992. The Xinanjiang model applied in China. J. Hydrol. 135 (1-4), 371-381.

Zhu, S., Wei, J., Zhang, H., Xu, Y., Qin, H., 2023. Spatiotemporal deep learning rainfall-runoff forecasting combined with remote sensing precipitation products in large scale basins. J. Hydrol. 616, 128727.