Joint Spatial and Temporal Modeling for Hydrological Prediction

QUN ZHAO[1], **YUELONG ZHU**[1], **KAI SHU**[2], **DINGSHENG WAN**[1],
YUFENG YU[1], **XUDONG ZHOU**[3], **AND HUAN LIU**[2], **(Fellow, IEEE)**

[1]College of Computer and Information, Hohai University, Nanjing 211100, China
[2]College of Computer Science and Engineering, Arizona State University, Tempe, AZ 85281, USA
[3]Institute of Industrial Science, The University of Tokyo, Tokyo 113-8656, Japan

Corresponding author: Dingsheng Wan (dshwan@hhu.edu.cn)

ABSTRACT The accurate and timely estimation of river discharge plays an important role in hydrological modeling, especially for avoiding the consequences of flood events. The majority of existing work on hydrologic prediction focuses on modeling the inherent physical process for specific river basins, while the geographic-connections between rivers are largely ignored. Geographically connected rivers provide rich spatial information that can be used to predict discharge amounts. In this paper, we study a novel problem of exploiting both temporal patterns and spatial connections for hydrological prediction. We construct three relationship graphs for hydrological gauges in the study area: the hydraulic distance graph, the Euclidean distance graph and the correlation graph. We fuse these graphs into one hydrological network graph, and propose a novel framework **ST-Hydro** which exploits Graph Convolutional Networks (GCN) for learning the spatial feature representations, and Recurrent Neural Networks with carefully designed activation functions for capturing temporal features simultaneously for hydrological prediction. Experimental results on real world data set demonstrate that the proposed framework can predict the river discharge effectively and at an early stage.

INDEX TERMS Hydrologic prediction, spatial and temporal modeling, graph convolutional networks.

I. INTRODUCTION

With global warming, changeable weather and frequent extreme precipitation events in recent years, flood events have occurred more frequently and drastic than usual. For example, it is reported that the Midwestern United States has been experiencing major floods since mid-March 2019, and nearly 14 million people have been affected by the flooding.[1] To mitigate the detrimental effects caused by flood disasters, hydrologic prediction becomes particularly important. Hydrologic prediction is to forecast the hydrological information (e.g., runoff, water level) in a certain period of time (e.g., few hours) in the future according to the hydrological and meteorological data in the early stage. It can provide a basis for decision-making in flood control and disaster relief, and is important for the rational utilization of water resources.

Therefore, it is critical to predict hydrological events effectively and timely to help hydrologists for decision making.

The majority of existing methods fall into three types: conceptual model, physical model and data-driven model. The conceptual model is based on the physical concept and empirical formula of hydrological phenomena, and the physical model is based on the physical process. They can predict hydrological events for specific rivers. But these models may need a large number of hydrologic parameters, and can not be easily adapted to other water basins. Specifically, since the parameters of the physical process in different basins are quite disparate, and the parameters and even the structure of physical-based models may need to be significantly modified. Therefore, recently the data-driven hydrologic prediction model has gained increasing attention for predicting hydrological events. For example, Hwang *et al.* [24] proposed an improved support vector machine method for nonlinear hydrological time series prediction with better prediction accuracy and prediction performance, and Kratzert *et al.* [27] proposed a new Rainfall–Runoff

---

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Moinul Hossain.

[1]https://en.wikipedia.org/wiki/2019_Midwestern_U.S._floods

Model using Long Short-Term Memory (LSTM) networks. However, most of the existing data-driven models focus on the learning the temporal characteristics from historical data.

In fact, hydrological data has rich spatial information. For example, the upstream of a river can directly affect the runoff of the downstream, and the two nearby rivers are likely to have similar rainfall processes, which results in similar changes in runoff. The spatial relationships across rivers provide important signals and have great potential to advance the hydrological prediction performance [10], [37]. However, most of the current data-driven hydrological forecasting methods focus on modeling the temporal patterns, without considering the rich spatial characteristics. In essence,we investigate: (1) How to build a hydrological topology map and mine spatial information; (2) How to combine temporal features with spatial features to improve prediction accuracy. The solution to these challenges results in a novel framework **ST-Hydro**, which incorporates Graph Convolutional Networks (GCN) for learning the Spatial feature representations and Recurrent Neural Networks with carefully designed activation functions for capturing Temporal features simultaneously for Hdrologic prediction. The contributions are as follows:

- We study a novel problem of exploiting both spatial and temporal information for data-driven hydrological prediction;
- We propose a principled way to jointly learn spatial features with Multi-Graph Convolutional Networks and capture temporal features with an adaptive Gated Recurrent Unit for hydrological prediction;
- We conduct extensive experiments on real world datasets to demonstrate the effectiveness and timeliness of the proposed model.

The paper is organized as follows. Section 2 reviews the related research about hydrological prediction and the applications of GCN. Section 3 introduces the details of the model. In section 4, we compare the performance of our model with that of other prediction models. Section 5 concludes the paper and describes the future work.

II. RELATED WORK

In this section, we briefly discuss the related work on: (1) hydrologic prediction; and (2) graph neural networks for temporal modeling.

A. HYDROLOGIC PREDICTION

The existing hydrological models are mainly divided into conceptual model, physical model and data-driven model. The conceptual model is based on the physical concept of hydrological phenomena and some formulas, such as the Xin'anjiang model [6] and the HBV model [31]. The physical model is based on the laws of conservation of mass and energy in physics, as well as the characteristics of runoff generation and concentration, to construct a set of hydrodynamic equations to simulate the changes of time and space and calculate them, such as the SHE model [11] and the

SWAT model [1]. The data-driven model uses some methods to deduce the response function according to the input and output data. It produces the corresponding output results for some input data through empirical analysis, such as the ANN model [25]. Many hydrologic models are often designed for specific water basins and may not be easily adapted to other watersheds. The parameters used between different basins are quite different and not universal. When the same kind of model is applied in different watersheds, the model parameters and even the model structure may need to be greatly modified [7], [22]. In addition, hydrological time series-based prediction methods that analyze the laws in historical hydrological data and predict future trends have been developed for hydrologic forecasting recently. Compared with other hydrological models, data-driven models can make up for some shortcomings of physical model and conceptual model. For example, after training, a data-driven model can be applied to other basins without a lot of parameter adjustment. In the data age, data-driven models are becoming more and more popular [32], [38].

Yule [46] established an autoregressive (AR) model to lay the foundation for a time series model. Nowak *et al.* [33] first used the AR model to predict runoff. Yu *et al.* [44] combined chaos theory with SVM and applied it to chaotic time series analysis with large sample data records. Hwang *et al.* [24] proposed a least squares support vector machine (LS-SVM) method for nonlinear hydrological time series prediction with better prediction accuracy and prediction performance. Xing *et al.* [41] proposed a new heuristic optimization algorithm, BA algorithm, which is used to optimize SVM parameters and predict monthly average traffic in 2015. Compared with cross-validated support vector machines, the accuracy is improved. Li *et al.* [30]proposed an SVM flood forecasting model based on kernel principal component analysis(KPCA) and boosting algorithm. The nonlinear KPCA is applied to extract the useful information from historical flood data. Experiments show that the proposed SVM ensemble model based on KPCA and boosting learning can improve the flood forecasting accuracy effectively. Atiquzzaman and Kandasamy [4] [5] proposed a fast prediction method of hydrological time series by using the limit learning machine (ELM), aiming at the uncertainty of traditional gradient based slow learning algorithm in training and iteratively determining network parameters. Based on the basic principle of flooding formation, Chen *et al.* [13] proposed a data-driven hydrology forecasting model, Convolution Regression based on Machine Learning. This model could reflect the impact of hourly rainfall on the future flow changes and the flow changes are predicted by superimposing these impacts.

B. GRAPH CONVOLUTIONAL NETWORKS FOR TEMPORAL PREDICTION

Convolutional Neural Networks (CNN) [28] is a type of feedforward neural network that includes convolution calculations and has a deep structure. It is one of the representative algorithms of deep learning [21]. However, there are many

graph structure or network structure data in the real world, which are difficult to deal with by traditional Convolutional Neural Networks(CNN). In 2017, Thomas N. Kipf proposed a semi supervised classification with graph structural network [26], which can process graph structure. Recently, Graph Convolutional Networks(GCN) is widely used in various domains such as social networks and information networks for feature learning [16]. GCN is mainly used in unstructured data such as social networks and information networks to extract features for learning. However, few people use GCN in temporal dependence data. For example, Yan *et al.* [42] proposed the spatiotemporal convolution network model ST-GCN, it solves the problem of human motion recognition which is based on the key points of the human skeleton. This model is built on the skeleton map sequence and each node corresponds to a joint of the human body. There are two sides, one is the spatial edge, the other is the temporal edge. Average the eigenvectors of all nodes in the neighborhood, and then calculate the inner product with the parameter vectors of convolution kernel. Chai *et al.* [12] used the inbound and outbound flows before t time to predict the inbound and outbound flows at t time. This paper constructs a variety of spatial relationships, and obtains the prediction effect by constructing multiple maps. But the structure of encoder decoder is limited, and the dimension of input and output is fixed. Zhao *et al.* [48] proposed a temporary graph revolutionary network, which combines GCN and Gated Recurrent Unit(GRU) to predict traffic information. Kim *et al.* [36] considered the spatial and temporal influence, and the influence of global variables, such as weather and weekday/weekend to reflect non-station-level changes. And then used graph convolutional network to predict bike demands. They obtain the time and space dependence from the T-GCN, and predict the vehicle speed in future one hour.

Spatial features are also used for processing in many aspects in hydrology (such as hydrological similarity analysis [49]). In this paper, we construct a novel hydrological topological structure, and use improved GCN to mine the spatial characteristics, improved GRU to mine the time characteristics, and further perform hydrologic prediction.

III. METHODOLOGY

In this section, we present the details of the proposed framework **ST-Hydro** for hydrological prediction, which mainly consists of three components (see Figure 1): (1) a spatial feature learning component; (2) a temporal feature learning component; and (3) a hydrological prediction component.

A. MODELING SPATIAL FEATURES

In hydrology, since hydrological process has a great relationship with geographical location and spatial features, it is important to capture spatial dependence between rivers. In this section, we establish the hydrological topology graph and use GCN to mine spatial features.
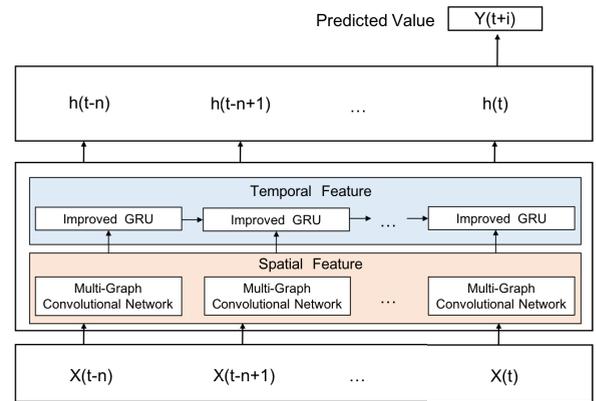

FIGURE 1. The proposed framework ST-Hydro for hydrological prediction consists of three components including a spatial feature learning module, a temporal feature learning module, and a hydrological prediction module.

1) HYDROLOGICAL TOPOLOGY GRAPH GENERATION

Graph generation is an important first step for training graph convolutional network model. If the graph can not reflect the relationship between nodes well, the model will not mine information effectively. The hydrological topology graph G is represented as a weighted graph, whose nodes are stations and edges are the geographical relationships between stations. We use $G = (V, E)$ to describe the topological structure of the hydrological network, where V is a set of station nodes, $V = \{v_1, v_2, \cdots, v_N\}$, N is the number of the stations, and E is a set of edges. The weights of edges represent the strength of relationship between stations.

From the hydrological point of view, if a gauge is downstream to another gauge in the same river, the similarity of the flow at these two gauges will be high because of their hydraulic connections [34], [40]. However, if two gauges are in different catchments without hydraulic connections (For example, A and C in Figure2), the similarity of the flow at two gauges is determined by the rainfall-runoff process in each own catchment. If the two gauges are close in space with short Euclidean distance, their discharge tends to have high similarities because of the similar climatic and land cover conditions. However, the short Euclidean distance is not the degerminator because the topographic conditions can significantly vary even in small local scale, the discharge series may still significantly differ in this case if they are not with the upstream-downstream connection.

Here we construct three different graphs to show the relationships between hydrological stations.

a: HYDRAULIC DISTANCE GRAPH

Here we use latitude and longitude to calculate the upstream and downstream relationship, and the distance of the river channel. The hydraulic connection between hydrological is estimated with regards to Digital Elevation Model (DEM). As water in each grid flows to its neighbouring grid with the steepest slope. Then, we can find the flow path of water
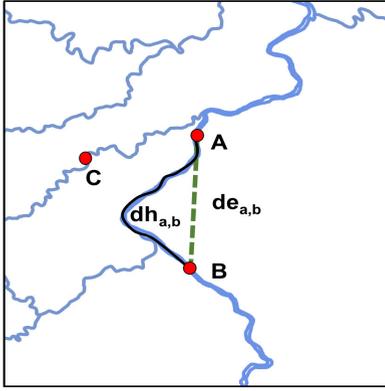
FIGURE 2. The case of different distances. Best viewed in colors.

from any point to the ocean or out of the study area. If the water path of an upstream gauge (e.g., G_i) goes through a downstream gauge (e.g., G_j), the hydraulic distance is the length of the flow path between the two gauges ($dh_{i,j}$). If they have no hydraulic connections, the $dh_{i,j}$ is ∞. The hydraulic distance of a certain gauge to itself is considered as 0. As an example, the $dh_{A,B}$ between A and B is shown in Figure 2. The matrix for the hydraulic distance is D_H:

$$D_H = \begin{bmatrix} dh_{1,1} & dh_{1,2} & \ldots & dh_{1,k} & \ldots & dh_{1,n} \\ dh_{2,1} & dh_{2,2} & \ldots & dh_{2,k} & \ldots & dh_{2,n} \\ & & \ldots & \ldots & & \\ dh_{m,1} & dh_{m,2} & \ldots & dh_{m,k} & \ldots & dh_{m,n} \\ & & \ldots & \ldots & & \\ dh_{n,1} & dh_{n,2} & \ldots & dh_{n,k} & \ldots & dh_{n,n} \end{bmatrix} \quad (1)$$

The adjacency matrix of hydraulic distance graph is A_H:

$$A_H = \begin{bmatrix} \dfrac{1}{dh_{1,1}} & \dfrac{1}{dh_{1,2}} & \ldots & \dfrac{1}{dh_{1,k}} & \ldots & \dfrac{1}{dh_{1,n}} \\ \dfrac{1}{dh_{2,1}} & \dfrac{1}{dh_{2,2}} & \ldots & \dfrac{1}{dh_{2,k}} & \ldots & \dfrac{1}{dh_{2,n}} \\ & & \ldots & \ldots & & \\ \dfrac{1}{dh_{m,1}} & \dfrac{1}{dh_{m,2}} & \ldots & \dfrac{1}{dh_{m,k}} & \ldots & \dfrac{1}{dh_{m,n}} \\ & & \ldots & \ldots & & \\ \dfrac{1}{dh_{n,1}} & \dfrac{1}{dh_{n,2}} & \ldots & \dfrac{1}{dh_{n,k}} & \ldots & \dfrac{1}{dh_{n,n}} \end{bmatrix} \quad (2)$$

b: EUCLIDEAN DISTANCE GRAPH
The matrix for Euclidean distance (D_e) is similar but the hydraulic distance is replaced.

$$de_{i,j} = 2R * atan2(\sqrt{a}, \sqrt{1-a}) \quad (3)$$

where $atan2$ is a function,[2] R is the earth radius, where

$$a = sin^2(\Delta\theta/2) + cos\alpha_i \cdot cos\alpha_j \cdot sin^2(\Delta\alpha/2) \quad (4)$$

where $\Delta\theta$ is the difference of latitude in two geographic points, α_i and α_j are longitudes of the two geographic points.

[2]https://en.wikipedia.org/wiki/Atan2

$\Delta\alpha$ is the difference in longitude. Note that the angles need to be in radians rather than in numerical latitudes or longitudes. For example, the $de_{A,B}$ between A and B is shown in the Figure 2. The matrix for the Euclidean distance is D_E:

$$D_E = \begin{bmatrix} de_{1,1} & de_{1,2} & \ldots & de_{1,k} & \ldots & de_{1,n} \\ de_{2,1} & de_{2,2} & \ldots & de_{2,k} & \ldots & de_{2,n} \\ & & \ldots & \ldots & & \\ de_{m,1} & de_{m,2} & \ldots & de_{m,k} & \ldots & de_{m,n} \\ & & \ldots & \ldots & & \\ de_{n,1} & de_{n,2} & \ldots & de_{n,k} & \ldots & de_{n,n} \end{bmatrix} \quad (5)$$

The adjacency matrix of Euclidean distance graph is A_E:

$$A_E = \begin{bmatrix} \dfrac{1}{de_{1,1}} & \dfrac{1}{de_{1,2}} & \ldots & \dfrac{1}{de_{1,k}} & \ldots & \dfrac{1}{de_{1,n}} \\ \dfrac{1}{de_{2,1}} & \dfrac{1}{de_{2,2}} & \ldots & \dfrac{1}{de_{2,k}} & \ldots & \dfrac{1}{de_{2,n}} \\ & & \ldots & \ldots & & \\ \dfrac{1}{de_{m,1}} & \dfrac{1}{de_{m,2}} & \ldots & \dfrac{1}{de_{m,k}} & \ldots & \dfrac{1}{de_{m,n}} \\ & & \ldots & \ldots & & \\ \dfrac{1}{de_{n,1}} & \dfrac{1}{de_{n,2}} & \ldots & \dfrac{1}{de_{n,k}} & \ldots & \dfrac{1}{de_{n,n}} \end{bmatrix}$$
$$(6)$$

c: CORRELATION GRAPH
The correlation between stream flow is used as a means to understand spatial patterns of stream flow dynamics [3], [8]–[10], [45]. In addition to distance, we also calculate correlations based on the daily steam flow of the stations for the last five years. We use Pearson 's correlation to calculate the correlation. $c_{m,k}$ is the correlation result between station m and station k. The adjacency matrix of correlation graph is A_C:

$$A_C = \begin{bmatrix} 1 & c_{1,2} & \ldots & c_{1,k} & \ldots & c_{1,N} \\ c_{2,1} & 1 & \ldots & c_{2,k} & \ldots & c_{2,N} \\ & & \ldots & \ldots & & \\ c_{m,1} & c_{m,2} & \ldots & c_{m,k} & \ldots & c_{m,N} \\ & & \ldots & \ldots & & \\ c_{n,1} & c_{n,2} & \ldots & c_{n,k} & \ldots & 1 \end{bmatrix} \quad (7)$$

2) GRAPH FUSION
Then we need to merge different graphs into one graph. We combine different graphs by the weighted summing their adjacency matrices at the element level. We use the channel distance between the upstream and downstream rivers, the Euclidean distance between the common stations and correlation values between stations to build the adjacency matrix.

$$a_{i,j} = \begin{cases} 1, & i = j \\ \alpha * \dfrac{1}{dh_{i,j}} + \beta * \dfrac{1}{de_{i,j}} + \gamma * c_{i,j}, & i \neq j \end{cases} \quad (8)$$

where $a_{i,j}$ is the elementary of the ith row and jth column in the adjacency matrix A. Because the hydraulic distance is more important than the Euclidean distance [34], we take

0.4 for α and 0.2 for β and 0.4 for γ. Then we can obtain the adjacency edges and nodes of the graph, and construct the hydrological topology graph.

3) GRAPH CONVOLUTION

In this section, we need to do graph convolution. The essence of discrete convolution is weighted sum. Convolution in CNN essentially uses a filter with shared parameters to extract spatial features by calculating the weighted sum of central pixels and neighboring pixels. However, CNN cannot process the non-structure data. Different from CNN, GCN can extract the spatial features of topological graphs. Here we use the GCN to capture the spatial features from the hydrological information(shown in Figure 3). In the graph structure data, we should consider the characteristic information and structure information of the nodes at the same time, that is, the adjacency matrix A and the characteristic matrix X. For example, when predicting runoff, X is daily runoff data and rainfall data. The model can be defined as:

$$X^{(l+1)} = f\left(X^{(l)}, A\right) \tag{9}$$

where l is the number of layers, $X^{(l)}$ is the feature of the node of the first layer, A is adjacency matrix.
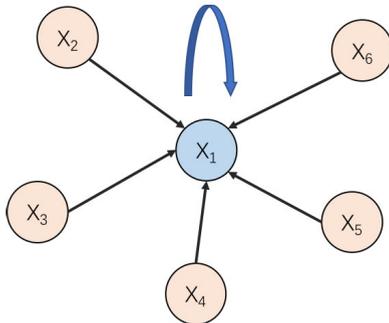

FIGURE 3. The idea of GCN. X_1 is the target station, others are the stations in the hydrological graph.

Then, the feature transformation of nodes is carried out first, and the adjacency matrix is normalized by the degree matrix. After adding the self cycle, the relationship between each node and the adjacency node is considered, and the specific model formula is as follows:

$$x_i^{(l+1)} = \sigma\left(\sum_{j \in N_i} \hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}x_j^{(l)}W^{(l)} + b^{(l)}\right) \tag{10}$$

where $x_i^{(l+1)}$ is the feature of node i in $l+1$ layer, $x_j^{(l)}$ is the feature of all neighbor nodes of node (including itself)in layer l, σ is a nonlinear transformation, A is the adjacency matrix, $\hat{A}$ represents self circulation, $\hat{D}$ is the degree matrix corresponding to $\hat{A}$, N_i is all neighbors of node i (including itself), $W^{(l)}$ is the weight of layer l, $b^{(l)}$ is the intercept of layer l.

B. LEARNING TEMPORAL FEATURES

We have demonstrated the spatial feature extraction of hydrological topological structure, now we demonstrate how to learn temporal features from time series. Gated Recurrent Unit (GRU) [14] can improve the gate design of the Long Short-Term Memory [19](LSTM) to overcome the gradient disappearance problem. In GRU, the activation function is very important. Existing commonly used activation functions include *sigmoid*, *tanh*, *softsign*, and *ReLu*. *Sigmoid* activation function [43]:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{11}$$

The *sigmoid* function is soft saturated, which can be guided everywhere in the definition domain. When x takes a large value, the corresponding y value will not be very different. In other words, when the input value tends to be positive infinity or negative infinity, the gradient will approach zero. So the gradient will disappear. In addition, the mean shift problem is one of the shortcomings of the *sigmoid* function.

Tanh activation function [18]:

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{12}$$

The activation function *tanh* has a value range of -1 to 1, so its output mean is close to 0, overcoming the mean shift problem generated by the *sigmoid* function. It makes the stochastic gradient fall closer to the natural gradient, reduces the number of iterations required to solve the network parameters, shortens the training time of the deep network. However, it still has soft saturation, which is easy to cause the gradient to disappear.

Softsign activation function [17]:

$$\text{softsign}(x) = \frac{x}{|x| + 1} \tag{13}$$

The *softsign* function is also antisymmetric, and can be differentiated. Its value range is between -1 and 1, and the output mean is close to 0, which overcomes the mean shift problem generated by *sigmoid*. Its curve is flatter than the *tanh* function. This makes the derivative decrease slower and can better solve the gradient disappearance problem.

ReLu activation function [2]:

$$\text{Re}\,Lu(x) = \max(0, x) \tag{14}$$

When the input value falls into the negative half axis, the gradient of the neuron is always 0, and no activation is activated for any data, which means that the neuron is dead. This will cause the calculation result to not converge. When the input value falls into the positive half axis, its derivative value is always 1, which can keep the gradient from falling and avoid the gradient disappearing problem. The value range of the *ReLu* function is non-negative, so the output mean is greater than 0, it has the disadvantage of the mean shift.

Although the *ReLu* function can effectively avoid the gradient disappearance, it also has some shortcomings. Since the result of the *ReLu* function is always 0 in the case of $x < 0$,
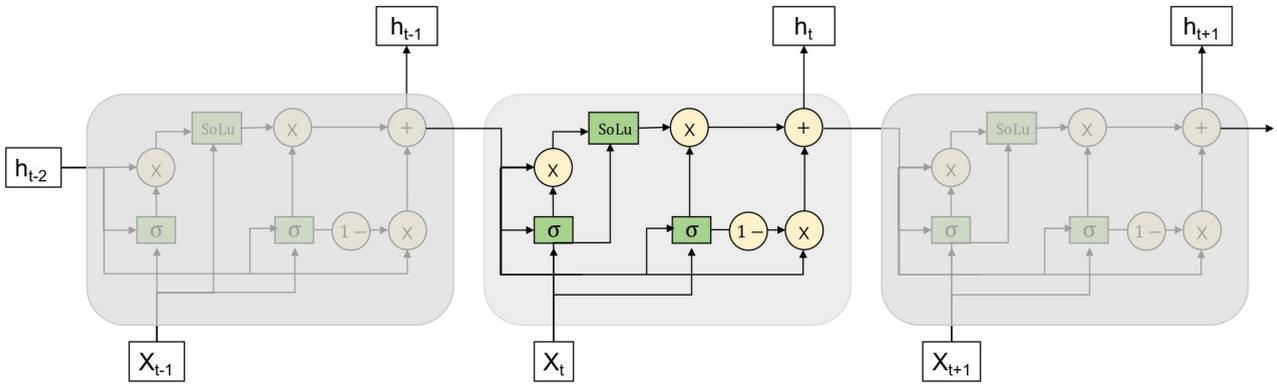
FIGURE 4. The structure of improved GRU model.

the output mean is greater than 0, resulting in a mean shift phenomenon. The output value of the *softsign* function on the negative half axis is less than 0, and the derivative decreases slowly. Combining the *softsign* function with the right half of the *ReLu* function not only mitigates the mean shift problem, but also prevents the left half of the neurons from dying. Therefore, in the case of $x < 0$, the *softsign* function is used instead, and we obtain a new activation function *SoLu*, which is defined as follows (β is 0.25):

$$\text{SoLu}(x) = \begin{cases} x, & x \geq 0 \\ \beta * \text{softsign}(x), & \text{otherwise} \end{cases} \tag{15}$$

The *SoLu* function combines the advantages of two commonly used activation functions, alleviating the gradient disappearance and output mean shift problems, and the robustness is also better.

The new activation function *SoLu* proposed in this paper is applied to the GRU instead of *tanh*. We propose the SoLu-GRU model to capture the temporal dependence. The model uses the *SoLu* to improve the output activation function of the GRU model in the hidden layer(shown in Figure 4).

In Figure 4, x_t is the input of the current time neuron, h_t is the output value of the current time neuron, h_{t-1} is the output of neurons in the previous moment, σ is the *sigmoid* function.

The construction steps are as follows:

Step 1: Build an update gate.

$$z_t = s\left(W_1 \cdot [h_{t-1}, x_t]\right) \tag{16}$$

W_1 is the weight of the update gate,h_{t-1} is the output of neurons in the previous moment. When the value of Z_t is larger, the less information neurons can leave in the previous time, and the more information neurons can leave in the current time.

Step 2: Build a reset gate.

$$r_t = s\left(W_2 \cdot [h_{t-1}, x_t]\right) \tag{17}$$

W_2 is the weight matrix of reset gate. When the value of r_t is 0, it means that the useless information transmitted by the neuron at the previous moment is discarded, and only the

input of the neuron at the current moment is reserved as the input.

Step 3: Build pending output section.We use *Solu* instead of *Tanh*, that is:

$$\widetilde{h}_t = \text{SoL}\,u\left(W_3 \cdot [r_t * h_{t-1}, x_t]\right) \tag{18}$$

W_3 is the weight vector of update door.

Step 4: Build the output section.

$$h_t = z_t * \tilde{h}_t + (1 - z_t) * h_{t-1} \tag{19}$$

C. HYDROLOGIC PREDICTION

The **ST-Hydro** model proposed in this paper adds spatial features to time series prediction. GCN is used to capture spatial dependence, while GRU, which is improved by *SoLu* activation function, captures temporal features.

1) SPATIAL FEATURE EXTRACTION

We take a 2-layer GCN as an example:

$$f(X_t, A) = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}\sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}X_t W_0\right) W_1\right) \tag{20}$$

where X_t is the feature matrix, A is the adjacency matrix, σ is a nonlinear transformation, we can use *ReLu* or other functions, $\hat{A}$ represents self circulation, $\hat{D}$ is the degree matrix corresponding to $\hat{A}$, W_0 is the weight of first layer, W_1 is the weight of the second layer.

2) LEARNING TEMPORAL FEATURES

Based on the spatial feature, we use the improved GRU to extract the temporal feature:

$$z_t = s\left(W_z \cdot [h_{t-1}, f(X_t, A)]\right) \tag{21}$$

$$r_t = s\left(W_r \cdot [h_{t-1}, f(X_t, A)]\right) \tag{22}$$

$$\tilde{h}_t = \text{SoLu}\left(W_{\tilde{h}} \cdot [r_t * h_{t-1}, f(X_t, A)]\right) \tag{23}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{24}$$

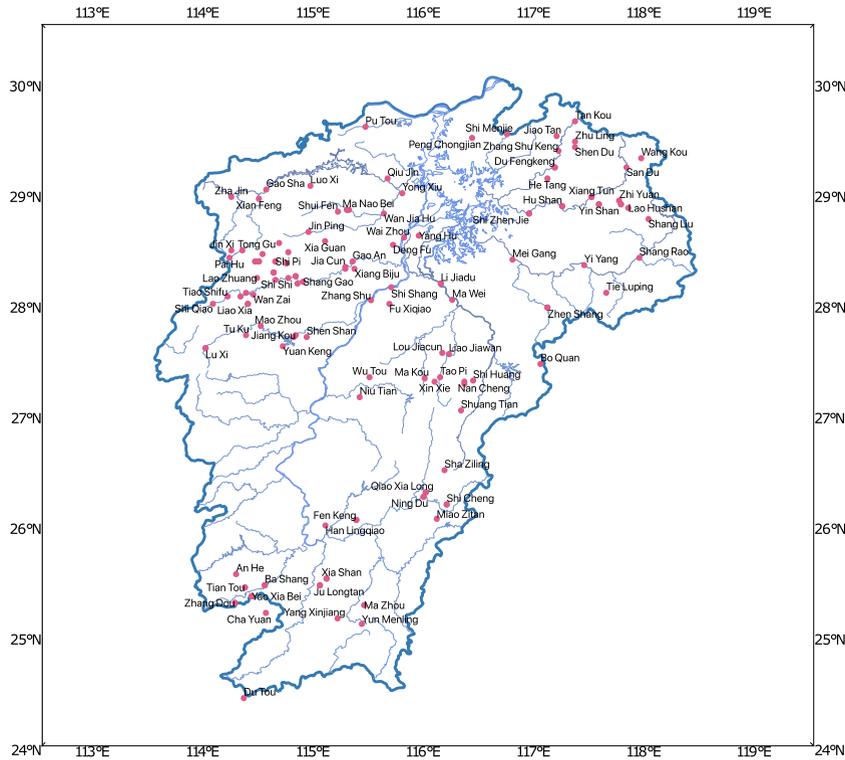where W_z, and W_r are the weights of the update gate and reset gate.

FIGURE 5. The representative 139 stations in Jiangxi Province.

3) LOSS FUNCTION

We use Y_t and Y_t' to donate the observation value and predicted value of runoff. The loss function is shown as follows:

$$Loss = \frac{1}{N} \sum_{i=1}^{N} \left(Y_t - Y_t' \right)^2 \qquad (25)$$

Then we get the predicted value through the fully connected layer:

$$Y_t' = O \left(W_c h_t + b \right) \qquad (26)$$

$O(\cdot)$ is a linear function, and W_c is the weight.

IV. EXPERIMENTS AND RESULTS

In this section, we present the experiments to evaluate the effectiveness and timeliness of the proposed **ST-Hydro** framework. Specifically, we aim to answer the following evaluation questions:

- **EQ1**: Can **ST-Hydro** improve the prediction performance of hydrological prediction by modeling the spatial and temporal features simutaneously?
- **EQ2**: How robust is **ST-Hydro** for hydrological prediction in one certain flood?
- **EQ3**: Whether the multi-graph convolutional network is better than single graph convolutional network?
- **EQ4**: Whether the proposed activation function is more effective for temporal modeling?

A. DATA DESCRIPTION

In this section, we introduce the data sets used in the experiment, the hydrological data sets of China Jiangxi Province. The experimental data includes runoff data, longitude and latitude data, and rainfall data of China Jiangxi Province. To eliminate the missing data in stations, station relocation, stoppages of stations, we finally selected 139 stations as the research stations in Jiangxi Province(see in Figure 5). The discharge and rainfall data of summer floods from 1998 to 2010 are selected as the experimental data. Because part of the rainfall is incomplete, we use the linear interpolation method commonly used in hydrology to supplement the missing rainfall. We take each hydrological station as a node, the input adjacency matrix is the station relationship, and use the characteristic matrix to represent the historical runoff flow and rainfall information. We use 80 percent of the dataset as the training dataset and 20 percent as the testing dataset.

B. EVALUATION METRICS AND BASELINES

We use the widely-adopted evaluation metrics that are based on the flood forecasting indicators given in the "Hydrological Information Forecasting Specification (GB/T 22482-2008)", including flood peak flow, flood peak occurrence time and flood process, etc.

1) EVALUATION METRICS

a: THE DETERMINISTIC COEFFICIENT

The deterministic coefficient (DC) represents the degree of coincidence between the predicted value and the measured

value. The closer the value is to 1.0, the better the prediction accuracy is.The formula is:

$$DC = 1 - \frac{\sum_{i=1}^{n} \left[y_p(i) - y(i) \right]^2}{\sum_{i=1}^{n} [y(i) - \bar{y}]^2} \qquad (27)$$

where $y_p(i)$ is the predicted value, $y(i)$ is the measured value, $\bar{y}$ is the average value of the measured value, n is the number of samples.

b: THE ROOT MEAN SQUARE ERROR

The root mean square error (RMSE) represents the deviation between the predicted value and the actual value. The calculation formula is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[y_p(i) - y(i) \right]^2} \qquad (28)$$

c: THE FLOOD PEAK ERROR

The flood peak error(error) represents the difference between the maximum value and the maximum value of the actual value in the flow forecast of a flood peak.

$$error = y_p^{\max} - y^{\max} \qquad (29)$$

where $y_p^{\max}$ is the maximum value of the predicted value, $y^{\max}$ is the maximum value of the measured value.

d: THE TIME DIFFERENCE OF THE FLOOD PEAK

The time difference of the peak (time) represents the difference between the moment when the maximum value of the predicted flow value in the flood occurs and the moment when the maximum value of the true value appears.

$$time = t_1 - t_2 \qquad (30)$$

where t_1 is the time when the predicted maximum value appears, and t_2 is the time when the real maximum value appears.

2) BASELINES

Here we introduce some baselines.

a: SUPPORT VECTOR MACHINE

Support vector machine(SVM) [39] Support Vector Machine (SVM) is a generalized linear classifier that performs binary classification of data in a supervised learning manner. In recent years, SVM has been used in many fields, including hydrology [15].

b: LONG SHORT-TERM MEMORY

Long Short-Term Memory (LSTM) [23] is a kind of time-recurrent neural network, which is specially designed to solve the long-term dependency problem of general Recurrent Neural Network. Many researches [27], [35], [47] find the potential of the LSTM for hydrological modelling applications.

c: GATED RECURRENT UNIT

Gated recurrent unit (GRU) was proposed by Chung *et al.* [14] to make each recurrent unit to adaptively capture dependencies of different time scales. Because of its simplicity and low computational overhead, it was quickly applied to many fields and also achieved good results in the hydrological field [20], [29].

C. EVALUATING OVERALL EFFECTIVENESS

To answer **EQ1**, we conduct the prediction performance comparison of different models. The Poyang Lake Basin in Jiangxi Province is the region with the most frequent floods and floods in the middle and lower reaches of the Yangtze River. The analysis of the hydrological data in this basin can provide strong support for the prevention and control of flood disasters. We choose the waizhou station in the Poyang Lake Basin as the display of hydrologic prediction results.

Training parameters of SVM model are configured as follows: penalty coefficient C interval is [0.23,0.3], kernel function is radial basis kernel function, parameter g interval is [0.05,0.08]. The number of hidden nodes in LSTM is 155, the interval of learning factor LR is [0.005,0.009], the batch size is 256, and the number of iterations is 40. The number of hidden nodes in GRU is 50, the learning rate LR interval is [0.005,0.007], the batch size is 16, and the number of iterations is 50. The number of hidden nodes in **ST-Hydro** is 32, the learning rate LR is [0.005,0.007], the batch size is 64, and the number of iterations epochs is 200.

The results are shown in Figure 6 and Table 1. Figure 6 is the runoff prediction results of WaiZhou station, and Table 1 shows the evaluation metric values of different models in the prediction.

It can be seen from the Figure 6 that the prediction result of SVM model is more jittery, and the prediction curve obtained is not as stable as LSTM, GRU, and **ST-Hydro**. The prediction results of LSTM and GRU are similar,**ST-Hydro**is better than GRU, and is more stable than LSTM.

From Table 1, we can see that the 1 hour runoff forecast is relatively the most accurate, and the later the forecast time is, the lower the accuracy is. **ST-Hydro** is the best model in all different forecast periods. The performance of LSTM and GRU is slightly better than that of GRU, but the difference between them is not significant. SVM performed the worst. Through the evaluation metric of the prediction results in the table, it can be seen that the DC coefficient of **ST-Hydro** is the highest, the difference between LSTM and GRU is not significant, and SVM is slightly worse. Therefore, the **ST-Hydro** model is the best in the overall prediction results.

This may be because GCN captures the spatial characteristics of runoff flow, adds some rainfall of several related stations in the space to the prediction, and the function proposed in this paper avoids the problems of gradient disappearance and mean shift.
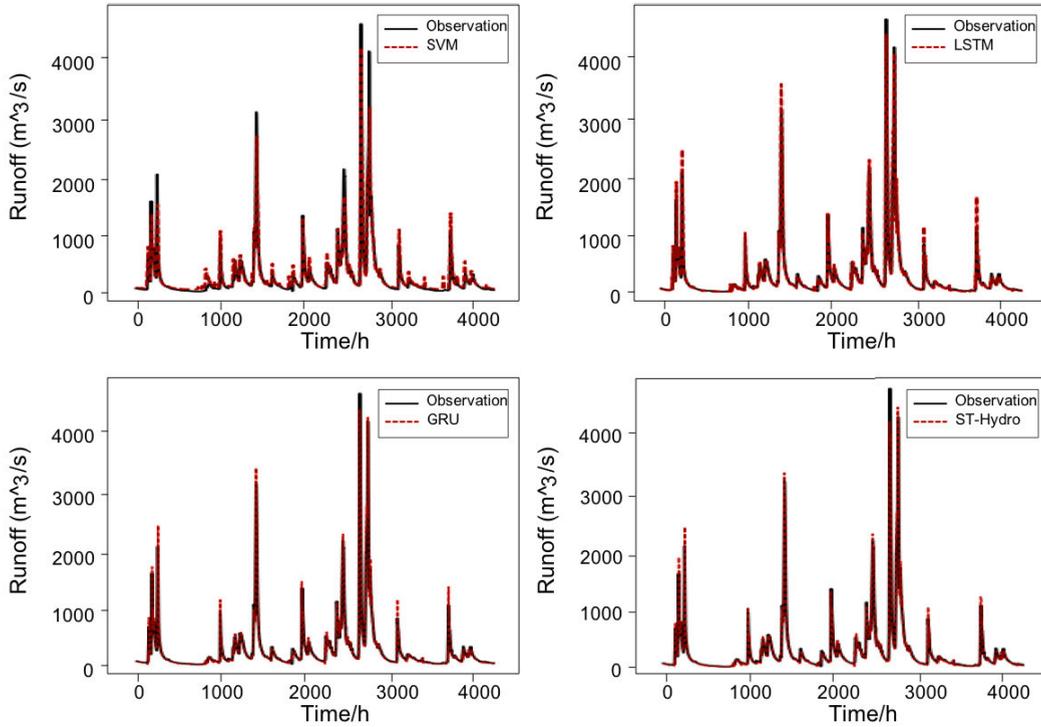
FIGURE 6. Take WaiZhou station as an example, the prediction results of four models are shown. *Observation* is the observation value of runoff. Best viewed in colors.

TABLE 1. Prediction comparisons of WaiZhou station.

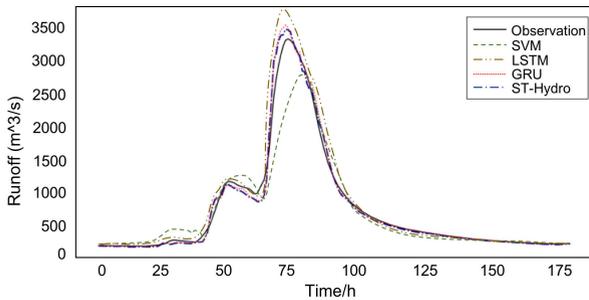| Period	1 hour		2 hour		3 hour		4 hour		5 hour	
Metric	DC	RMSE	DC	RMSE	DC	RMSE	DC	RMSE	DC	RMSE
SVM	0.95	128.34	0.93	135.49	0.92	162.04	0.91	175.48	0.89	191.4
LSTM	0.97	91.66	0.97	101.23	0.96	105.76	0.96	119.9	0.95	128.82
GRU	0.98	92.73	0.96	98.54	0.96	104.71	0.97	105.85	0.95	126.96
ST-Hydro	0.99	72.34	0.98	81.67	0.98	87.79	0.97	104.71	0.96	117.03

FIGURE 7. Take one flood (the next 3 hour) as an example, the comparison of different models are shown.

D. EVALUATING EFFECTIVENESS OF ONE FLOOD

To answer **EQ2**, we choose one flood in Waizhou station to show the result details of different models. The results are compared in Figure 7 and Table 2. Figure 7 is one flood prediction of different models, Table 2 is the results of the same flood prediction for different future periods.

In the process of hydrological prediction, the time of flood peak appears is very important, that is, the peak time. If the predicted peak time is ahead of the real peak time, it can

prevent the occurrence of flood in advance and prepare for the flood control work. The lag of predicted peak time will bring some difficulties and challenges to the flood control work.

From the Figure 7 and Table 2, we can see that the effect of SVM is the worst and the peak time lag in the data set, even 5 hours' prediction lag in Waizhou station. LSTM, GRU and **ST-Hydro** can reach the peak in time or one hour in advance, which provides guidance for flood control operation. The peak error of **ST-Hydro** is the smallest and the accuracy is the highest, so **ST-Hydro** can predict more accurately in all aspects.

We also can see from the Table 2 that SVM always has the problem of delay in peak time prediction. Both LSTM and GRU can reach the peak value exactly or one hour in advance. The most surprising thing is that **ST-Hydro** is able to reach the peak point two hour ahead of time in the 5hours prediction, which is very meaningful, and the peak error of **ST-Hydro** is smaller. So the effect of **ST-Hydro** is better.

E. COMPARING DIFFERENT GRAPHS PERFORMANCE

In the experiments, to answer **EQ3**, we predict the same data by using different graphs in capturing spatial feature.

TABLE 2. Prediction comparison of one WaiZhou Flood for different future periods.

Model		SVM	LSTM	GRU	ST-Hydro
1 hour	DC	0.94	0.98	0.99	0.99
	RMSE	156.45	89.33	86.52	57.33
	error	-203	212	198	123
	time	0	1	1	1
2 hour	DC	0.93	0.98	0.98	0.98
	RMSE	177.38	97.25	95.23	76.96
	error	-358	227	232	147
	time	-1	1	1	1
3 hour	DC	0.93	0.97	0.97	0.99
	RMSE	218.58	119.58	114.43	85.88
	error	-422	416	260	142
	time	-5	1	1	1
4 hour	DC	0.92	0.98	0.98	0.98
	RMSE	236.63	122.85	119.58	114.43
	error	-462	307	438	260
	time	-6	1	1	1
5 hour	DC	0.90	0.97	0.97	0.98
	RMSE	266.95	138.75	132.97	118.12
	error	-581	556	378	156
	time	-5	0	1	2

Here we use the data set of Jiangxi Province to compare the hydrological prediction models established by different maps. We use hydraulic distance, Euclidean distance, correlation and fusion graphs for experiments. The discharge and rainfall data of summer flood from 2000 to 2010 are selected as the experimental data, and we take the average *RMSE* and *DC* of 15 stations for comparison. The result is shown in the Figure 8.
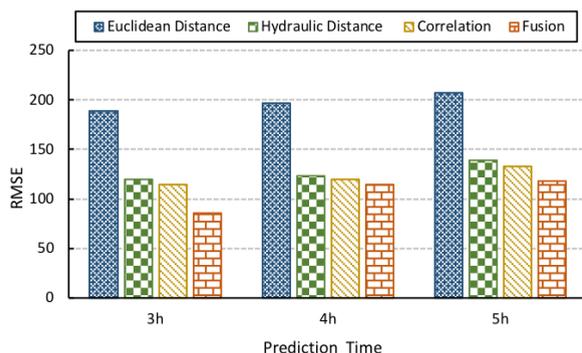

FIGURE 8. The comparison of multi-graph model and different single gragh model.

From Figure 8, we can see that the model using Euclidean distance graph has the worst performance, it may be because the Euclidean distance mostly shows the difference of rainfall processes, but we have added some rainfall information as the input in the whole prediction model. The models using hydraulic distance and correlation have the similar effect, and the model using correlation is slightly better. The model using fusion graph is the best.

F. ASSESSING ACTIVATION FUNCTIONS
In the experiments, to answer **EQ4**, *Tanh*, *softsign*, *ReLu* and *SoLu* are respectively used as activation functions to predict.

Because the training results do not converge when *sigmoid* function is used as activation function of the network, the experimental results are compared by *Tanh*, *softsign*, *ReLu* and *SoLu*.

It can be seen in Figure 9 that under the condition of 50 iterations, the training convergence time of *SoLu* activation function is the shortest and and the training error is the lowest. The order of error values with different activation functions is *SoLu* < *ReLu* < *softsign* < *Tanh*. Therefore, using *SoLu* as the activation function can not only effectively shorten the convergence time, but also reduce the training error.
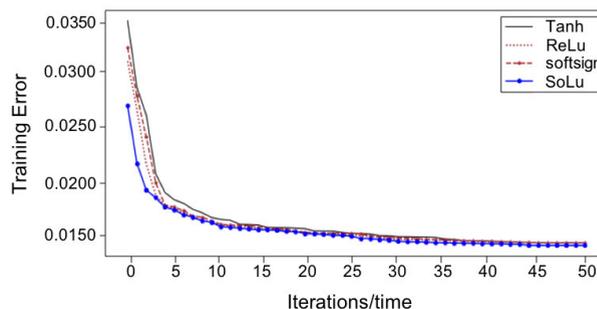

FIGURE 9. Model training error comparison of different activation functions.

V. CONCLUSION AND FUTURE WORK
In this paper, we propose a novel river discharge prediction model **ST-Hydro** which incorporates network-driven method Graph Convolutional Networks (GCN) with the data-driven method improved Gated Recurrent Unit (GRU). GCN is used to capture the spatial features and GRU is used to capture the temporal features of the discharge samples. We also use multi-graph to construct the GCN and use proposed activation function SoLu to improve Gated Recurrent Unit. Experiments show that the **ST-Hydro** model outperforms several state-of-the-art baselines.

There are several interesting directions need to be explored. For example, when building hydrological topological structure map, seasonal factors can be considered, different maps can be built in different time periods. In addition, we will consider using GCN to capture spatial features for the analysis of ungaged basins.

REFERENCES
[1] K. C. Abbaspour, J. Yang, I. Maximov, R. Siber, K. Bogner, J. Mieleitner, J. Zobrist, and R. Srinivasan, "Modelling hydrology and water quality in the pre-alpine/alpine thur watershed using SWAT," *J. Hydrol.*, vol. 333, nos. 2–4, pp. 413–430, Feb. 2007.
[2] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375
[3] S. A. Archfield and R. M. Vogel, "Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments," *Water Resour. Res.*, vol. 46, no. 10, Oct. 2010, Art. no. W10513.
[4] M. Atiquzzaman and J. Kandasamy, "Prediction of hydrological time-series using extreme learning machine," *J. Hydroinformatics*, vol. 18, no. 2, pp. 345–353, Mar. 2016.
[5] M. Atiquzzaman and J. Kandasamy, "Robustness of extreme learning machine in the prediction of hydrological flow series," *Comput. Geosci.*, vol. 120, pp. 105–114, Nov. 2018.

[6] H.-J. Bao, L.-N. Zhao, Y. He, Z.-J. Li, F. Wetterhall, H. L. Cloke, F. Pappenberger, and D. Manful, "Coupling ensemble weather predictions based on TIGGE database with grid-xinanjiang model for flood forecast," *Adv. Geosci.*, vol. 29, pp. 61–67, Feb. 2011.

[7] A. Bárdossy, "Calibration of hydrological model parameters for ungauged catchments," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 3, no. 3, pp. 1105–1124, 2007.

[8] A. Betterle, D. Radny, M. Schirmer, and G. Botter, "What do they have in common? Drivers of streamflow spatial correlation and prediction of flow regimes in ungauged locations," *Water Resour. Res.*, vol. 53, no. 12, pp. 10354–10373, Dec. 2017.

[9] A. Betterle, M. Schirmer, and G. Botter, "Characterizing the spatial correlation of daily streamflows," *Water Resour. Res.*, vol. 53, no. 2, pp. 1646–1663, Feb. 2017.

[10] A. Betterle, M. Schirmer, and G. Botter, "Flow dynamics at the continental scale: Streamflow correlation and hydrological similarity," *Hydrol. Processes*, vol. 33, no. 4, pp. 627–646, Feb. 2019.

[11] K. Beven, R. Warren, and J. Zaoui, "She: Towards a methodology for physically-based distributed forecasting in hydrology," *IAHS Publ*, vol. 129, pp. 133–137, Apr. 1980.

[12] D. Chai, L. Wang, and Q. Yang, "Bike flow prediction with multi-graph convolutional networks," in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2018, pp. 397–400.

[13] C. Chen, Q. Hui, Q. Pei, Y. Zhou, B. Wang, N. Lv, and J. Li, "CRML: A convolution regression model with machine learning for hydrology forecasting," *IEEE Access*, vol. 7, pp. 133839–133849, 2019.

[14] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: http://arxiv.org/abs/1412.3555

[15] S. Raghavendra. N and P. C. Deka, "Support vector machine applications in the field of hydrology: A review," *Appl. Soft Comput.*, vol. 19, pp. 372–386, Jun. 2014.

[16] K. Ding, J. Li, R. Bhanushali, and H. Liu, "Deep anomaly detection on attributed networks," in *Proc. Int. Conf. Data Mining (SIAM)*, 2019, pp. 594–602.

[17] D. L. Elliott, "A better activation function for artificial neural networks," Univ. Maryland, College Park, MD, USA, Tech. Rep. TR.93-8, 1993.

[18] E. Fan, "Extended tanh-function method and its applications to nonlinear equations," *Phys. Lett. A*, vol. 277, nos. 4–5, pp. 212–218, Dec. 2000.

[19] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, Oct. 2000.

[20] I. Groenen, "Representing seasonal patterns in gated recurrent neural networks for multivariate time series forecasting," M.S. thesis, Univ. Amsterdam, Amsterdam, The Netherlands, 2018.

[21] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[22] G. Heuvelmans, B. Muys, and J. Feyen, "Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets," *J. Hydrol.*, vol. 319, nos. 1–4, pp. 245–265, Mar. 2006.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] S. H. Hwang, D. H. Ham, and J. H. Kim, "Forecasting performance of LS-SVM for nonlinear hydrological time series," *KSCE J. Civil Eng.*, vol. 16, no. 5, pp. 870–882, Jul. 2012.

[25] R. S. Govindaraju, "Artificial neural networks in hydrology. II: Hydrologic applications," *J. Hydrologic Eng.*, vol. 5, no. 2, pp. 124–137, Apr. 2000.

[26] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: http://arxiv.org/abs/1609.02907

[27] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (lstm) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022 2018.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[29] X. H. Le, H. V. Ho, and G. Lee, "Application of gated recurrent unit (GRU) network for forecasting river water levels affected by tides," in *Proc. Int. Conf. Asian Pacific Coasts*. Singapore: Springer, 2019, pp. 673–680.

[30] S. Li, K. Ma, Z. Jin, and Y. Zhu, "A new flood forecasting model based on SVM and boosting learning algorithms," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2016, pp. 1343–1348.

[31] G. Lindström, B. Johansson, M. Persson, M. Gardelin, and S. Bergström, "Development and test of the distributed HBV-96 hydrological model," *J. Hydrol.*, vol. 201, nos. 1–4, pp. 272–288, Dec. 1997.

[32] N. J. Mount, H. R. Maier, E. Toth, A. Elshorbagy, D. Solomatine, F.-J. Chang, and R. J. Abrahart, "Data-driven modelling approaches for socio-hydrology: Opportunities and challenges within the Panta Rhei Science Plan," *Hydrological Sci. J.*, vol. 61, no. 7, pp. 1192–1208, 2016.

[33] K. C. Nowak, B. Rajagopalan, and E. Zagona, "Wavelet auto-regressive method (WARM) for multi-site streamflow simulation of data with non-stationary spectra," *J. Hydrol.*, vol. 410, nos. 1–2, pp. 1–12, Nov. 2011.

[34] S. Patil and M. Stieglitz, "Hydrologic similarity among catchments under variable flow conditions," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 3, pp. 989–997, 2011.

[35] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting," *Acta Geophysica*, vol. 67, no. 5, pp. 1471–1481, Oct. 2019.

[36] T. S. Kim, W. K. Lee, and S. Y. Sohn, "Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects," *PLoS ONE*, vol. 14, no. 9, 2019, Art. no. e0220782.

[37] J. O. Skøien and G. Blöschl, "Spatiotemporal topological kriging of runoff time series," *Water Resour. Res.*, vol. 43, no. 9, Sep. 2007, Art. no. W09419.

[38] D. P. Solomatine, "Data-driven modeling and computational intelligence methods in hydrology," *Encyclopedia Hydrological Sci.*, vols. 1–2, no. 19, pp. 293–306, Apr. 2006.

[39] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[40] H. Tao, M. Gemmer, Y. Bai, B. Su, and W. Mao, "Trends of streamflow in the tarim river basin during the past 50 years: Human impact or climate change?" *J. Hydrol.*, vol. 400, nos. 1–2, pp. 1–9, 2011.

[41] B. Xing, R. Gan, G. Liu, Z. Liu, J. Zhang, and Y. Ren, "Monthly mean streamflow prediction based on bat algorithm-support vector machine," *J. Hydrologic Eng.*, vol. 21, no. 2, Feb. 2016, Art. no. 04015057.

[42] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI*, 2018, pp. 7444–7452.

[43] X. Yin, "A flexible sigmoid function of determinate growth," *Ann. Botany*, vol. 91, no. 3, pp. 361–371, Feb. 2003.

[44] X. Yu, S.-Y. Liong, and V. Babovic, "EC-SVM approach for real-time hydrologic forecasting," *J. Hydroinformatics*, vol. 6, no. 3, pp. 209–223, Jul. 2004.

[45] L. L. Yuan, "Using correlation of daily flows to identify index gauges for ungauged streams," *Water Resour. Res.*, vol. 49, no. 1, pp. 604–613, Jan. 2013.

[46] G. U. Yule, "VII. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers," *Phil. Trans. Roy. Soc. London. Ser. A*, vol. 226, nos. 636–646, pp. 267–298, 1927.

[47] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas," *J. Hydrol.*, vol. 561, pp. 918–929, Jun. 2018.

[48] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. L. T-Gcn, "A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[49] Q. Zhao, Y. Zhu, D. Wan, Y. Yu, and Y. Lu, "Similarity analysis of Small- and medium-sized watersheds based on clustering ensemble model," *Water*, vol. 12, no. 1, p. 69, 2020.

QUN ZHAO was born in Taixing, Jiangsu, China, in 1994. She received the B.S. degree from the College of Software Engineering, Jinling Institute of Science and Technology, Nanjing, China, in 2016. She is currently pursuing the Ph.D. degree in computer science and technology with the College of Computer and Information, Hohai University, Nanjing.

Her current research interests include data management and data mining, water informatization, and computer network application. She was awarded the National Encouragement Scholarship of China, in 2014, the National Scholarship of China, in 2015, the Team Winning Award of the China Service Outsourcing Innovation Competition, in 2015, the Second Prize of the Academic Scholarship of Hehai University, in 2017, and the First Prize of the Academic Scholarship of Hohai University, in 2018.

YUELONG ZHU was born in Jianhu, Jiangsu, in December 1959. He received the B.S. degree in automation engineering from Hohai University, Nanjing, China, in 1982.

He taught in Hohai University. He is currently a Professor and a Ph.D. Supervisor with Hohai University, the Deputy Chairman of the Water Resources Informatization Special Committee of the China Hydraulic Engineering Society, and the Vice Chairman of the Jiangsu Water Resources Protection Association. He was promoted to a Lecturer, in 1987, an Associate Professor, in 1995, and a Professor, in 2001. In 1993, he has served as the Deputy Director of the Department of Computer Engineering, Hohai University. In 1995, he has served as the Dean (presided over the work) with the School of Computer and Information Engineering, Hohai University, and served as the Dean, in 1998. In April 2003, he was selected as a member of the party committee of the School at the eleventh party congress of the school and served as the Vice President of Hohai University. He was awarded the National Second Prize for Scientific and Technological Progress, the First Prize for Excellent Achievements of the Ministry of Education, and the Dayu Water Science and Technology First Prize.
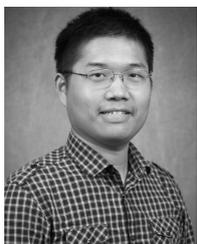
KAI SHU received the B.S./M.S. degrees from Chongqing University, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Data Mining and Machine Learning (DMML) Laboratory, Arizona State University. He is currently a Research Assistant with the DMML Laboratory, Arizona State University. His research interests include fake news detection, social computing, data mining, and machine learning. He was awarded the ASU CIDSE Doctorial Fellowship, in 2015, the First Place of SBP Disinformation Challenge, in 2018, the University Graduate Fellowship, and various scholarships. He co-presented two tutorials in KDD 2019 and WSDM 2019, and has published innovative works in highly ranked journals and top conference proceedings, such as ACM KDD, WSDM, WWW, CIKM, IEEE ICDM, IJCAI, AAAI, and MICCAI. He also worked as a Research Intern with Yahoo! Research and Microsoft Research, in 2018 and 2019, respectively. More can be found at http://www.public.asu.edu/~skai2/.

DINGSHENG WAN was born in Liyang, Jiangsu, China in 1963. He received the B.S. degree in automation engineering from Hohai University, Nanjing, China, in 1986.

He is currently a Professor with the College of Computer and Information, Hohai University. His current research interests include database technology, data management and data mining, and computer network application. He was awarded the Hehai University Science and Technology Progress Award, in 2001, the Jiangsu Engineering Excellent Consulting Award, in 2006, the National Engineering Excellent Consulting Award, in 2006, the China Surveying and Mapping Science and Technology Progress Award, in 2013, and the Jiangsu Water Resources Science and Technology Excellent Achievement Award, in 2013.

YUFENG YU was born in Huanggang, Hubei, China, in 1979. He received the B.S. degree in computer science from Wuhan University and the Ph.D. degree in computer science from Hohai University, Nanjing, China. He is currently an Associate Professor with the College of Computer and Information, Hohai University. His current research interests include data quality management, anomaly detection, social media mining, and water informatization.

XUDONG ZHOU was born in Rudong, Jiangsu, China, in 1991. He received the B.S. degree in water engineering and the Ph.D. degree in hydrology and water resources from Hohai University, in 2013 and 2019, respectively.

He is currently a Project Researcher with the Institute of Industrial Science, The University of Tokyo, Japan. His research interests are in hydrological modeling, flood risk assessment and data science in hydro-climatic aspects. He was awarded the National Scholarship of China, in 2010, 2013, and 2015. He obtained the First Prize of the Dayu Water Science and Technology Progress Award, in 2013. He was also awarding the Outstanding Student from the Liuguangwen Foundation, the Baogang Foundation, and the Yankai Foundation.

HUAN LIU (Fellow, IEEE) is currently a Professor of computer science and engineering with Arizona State University. Before he joined ASU, he worked at Telecom Australia Research Labs and was on the Faculty with the National University of Singapore. His research interests are in data mining, machine learning, social computing, and artificial intelligence, investigating interdisciplinary problems that arise in many real-world, data-intensive applications with high-dimensional data of disparate forms, such as social media. He is also a coauthor of *Social Media Mining: An Introduction* (Cambridge University Press). He is also a Founding Organizer of the International Conference Series on Social Computing, Behavioral-Cultural Modeling, and Prediction, Field Chief Editor of Frontiers in Big Data and its Specialty Chief Editor in Data Mining and Management. He is also a Fellow of ACM, AAAI, and AAAS. More can be found about him at http://www.public.asu.edu/~huanliu.

• • •