# Spatial and Temporal Aware Graph Convolutional Network for Flood Forecasting


1st Jun Feng
*College of Computer and Information*
*Hohai University*
Nanjing, China
fengjun@hhu.edu.cn

2nd Zhongyi Wang
*College of Computer and Information*
*Hohai University*
Nanjing, China
wangzhongyi@hhu.edu.cn

3rd Yirui Wu*
*College of Computer and Information*
*Hohai University*
Nanjing, China
wuyirui@hhu.edu.cn

4th Yuqi Xi
*College of Computer and Information*
*Hohai University*
Nanjing, China
xiyuqi@hhu.edu.cn


2021 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-3900-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/IJCNN52387.2021.9533694


***Abstract*—Intelligent flood forecasting systems provide an effective means to forecast flood disaster. Accurate flood flow value prediction is a huge challenge since it's influenced by both spatial and temporal relationship among flood factors. Popular deep learning structures like Long Short-Term Memory (LSTM) network lacks abilities of modeling the spatial correlations of hydrological data, thus cannot yield satisfactory prediction results. Moreover, not all the temporal information is always valuable for flood forecasting. In this paper, we proposed a novel spatial and temporal aware Graph Convolution Network (ST-GCN) for flood prediction, which is capable to extract spatial-temporal information from raw flood data. Moreover, a temporal attention mechanism is introduced to weight the importance of different time steps, thus involving global temporal information to improve flood prediction accuracy. Compared with the existing methods, results on two self-collected datasets show that ST-GCN greatly improves the prediction performance.**



***Index Terms*—flood forecasting, GCN, attention mechanism, data-driven model**


## I. Introduction

Natural disasters, especially flood, are gradually becoming one of the most important issues affecting our social and economic development. Due to global warming, flood events have occurred more frequently than before since the 1980s. Consequently, researchers in both hydrological and machine learning communities focus on improving our ability to forecast and prevent flood disasters.

Flood flow forecasting could offer future hydrological information on the basis of former hydrological and meteorological data. It's of great significance to predict flood events effectively and efficiently. However, the formation of flood is a complicated process, which is affected by topography, precipitation and other characteristics. Several approaches have been proposed to tackle the flood forecasting problem. These methods fall into two different categories, i.e., physical and data-driven model. The first category relies on hydrological prediction for specific rivers using physical mechanisms [1], which is quite sensitive to its internal parameters. Therefore, such physical models can be only applied in specific rivers. On the other hand, data-driven techniques have extensively been applied to the stream-flow forecasting [2], which achieves better performance in prediction accuracy by using machine learning approaches to capture temporal hydrological information.

Compared with traditional time series forecasting, flood prediction faces greater challenges, since it not only requires to consider temporal features, but also comprehensively takes a variety of hydrological related information into consideration. Among these information, spatial information makes great contribute to accurate prediction of floods. Take the upstream and downstream of the same river as an example, the runoff of the upstream will directly affect the runoff of the downstream. In the same river basins, the precipitation of different regions with close geographical location also has high similarity. In the process of construction of the intelligent flood prediction model, the spatial distribution of various hydrological monitoring stations should be considered.

Time-varying characteristics of flood requires researchers to consider flood as a dynamic process. In other words, the influence of related factors on flood should be different in various time periods of flood. For example, soil moisture, one of the flood factors, has a great influence on the runoff values at the initial stage of flood. After middle stage of flood, soil can be filled with water and factor of soil moisture keeps consistent. In other words, we should assign a higher weight to factor of soil moisture at the beginning, meanwhile offering a quite small weight to it in middle or final stage of flood. Therefore, it can be seen that the impact of the same hydrological features on the same flood may also change at

* indicates corresponding author

different times.

However, most data-driven models can only capture temporal dependency, and fail to describe spatial and temporal dependency . For example, long short-term memory (LSTM) network effectively retains the long-term hydrological characteristics so that successfully capture the temporal feature, but it cannot realize spatial relationship modeling. In this paper, we propose to firstly build a hydrological spatial node map, which is composed of geographical and statistical characteristics of hydrological stations. Based on the constructed node graph, we propose a spatial-aware graph convolution neural (S-GCN) network, which adopts GCN to extract distinguish spatial features. After the spatial modeling of GCN, we use LSTM to capture the hydrological temporal dependency. It's noted that GCN can extract spatial features form hydrological spatial node map, meanwhile LSTM can capture temporal features.

For the problem of dynamic feature extraction, attention mechanism is used to effectively capture the dynamic variation characteristics of the long-term hydrological data, which successfully assigns different weights to different temporal features, resulting in more accurate correlation analysis. Our main contribution can be concluded as follows:

- We propose a novel data-driven method to jointly learn spatial features with GCN and temporal features with LSTM for hydrological prediction. As far as we know, this is the first time to jointly use GCN and LSTM for flood prediction, which successfully model spatial and temporal relationship for flood forecasting task.
- We introduce a temporal-aware attention mechanism to describe the dynamics characteristics of floods, which effectively improves the ability of the proposed model to focus on informativeness ones of temporal information.

## II. Related Work

In this section, considering the relevance to the proposed method, we summarize the state of the art in the field of the data driven model for flood forecasting and discuss the related literature on graph convolutional network and attention-based models.

### *A. Data-driven Model for Flood Forecasting*

Since the rapid development of artificial intelligence technologies in the 21st century, various data driven algorithms have been widely used in flood forecasting. These methods can be broadly categorized in two different categories. The first category of work mainly utilizes the concept of probability and statistics, such as Bayesian model. Wu et al. [3] proposed to construct hierarchical Bayesian network to make a short-term stream-flow forecast. Comprehensive consideration of hydrology and statistics is of great importance to the prediction of flow value. The second category of work includes Artificial intelligence-based approaches using machine learning and deep learning. Commonly used machine learning algorithms include Support Vector Machines (SVM) [4], Decision Tree (DT) [5] , and Back Propagation Neural Networks (BPNN) [6], which solve the problem by abstracting actual flood forecasting tasks as non-linear regression tasks. With the massive accumulation of hydrological data and meteorological data, Long and Short-Term Memory (LSTM) has better applicability and accuracy than the traditional neural network model by introducing a gating unit to alleviate the gradient disappearance. Liu et al. [7] proposed a context and temporal aware attention LSTM network model, using the attention module in each step of LSTM, so as to achieved a high prediction accuracy based on the collected characteristics of various hydrological data.

### *B. Graph Convolutional Network*

The convolutional neural network (CNN) is a novel method for deep learning of graph data developed in recent years. The reason why CNN can be successfully applied in the field of computer vision is that it can extract spatial features well. The main function of GCN is also to extract spatial features, but different from traditional CNN, GCN mainly deals with graph data.

Bruna et al. [8] proposed the first graph convolutional neural network in 2013. The core of spectral domain graph convolution is to use a symmetric normalized Laplacian matrix. In the field of natural language processing, Marcheggiani et al. [9] proposed to use GCN model to act on syntactic dependency graphs and super-impose them with long and short-term memory networks. In the field of traffic prediction, Sun et al. [10] comprehensively considered various factors, constructing a Multi-View Graph Convolutional Network (MVGCN) for pedestrian flow prediction. At the same time, this method designed a fusion module that took into account both temporal and spatial characteristics. Temporal and spatial characteristics needs to be considered comprehensively in hydrology (different monitoring stations on the same river have spatial connections). In this paper, we use GCN to mine spatial features in the node graph of the hydrological monitoring stations.

### *C. Attention-based Models*

Human vision systems contain complex information processing mechanisms that can always quickly locate important target areas for detailed analysis. Relying on this powerful visual system, human process various data from the outside world into images, and they can quickly focus on the important target area, so that put more attention to the target area. This signal processing mechanism can enable human to acquire valuable information more efficiently so that greatly improves the efficiency and accuracy of human visual processing.

The attention mechanism is similar to human selective visual information processing systems in the field of deep learning. The core of this mechanism is to select information that is more critical to the current task goal from a large amount of information. With this advantage, it is widely used in image description, natural language processing and other fields. Choi et al. [11] proposed a fine-grained attention mechanism, so that improves the translation quality of BLEU scores and reveals how the fine-grained attention mechanism

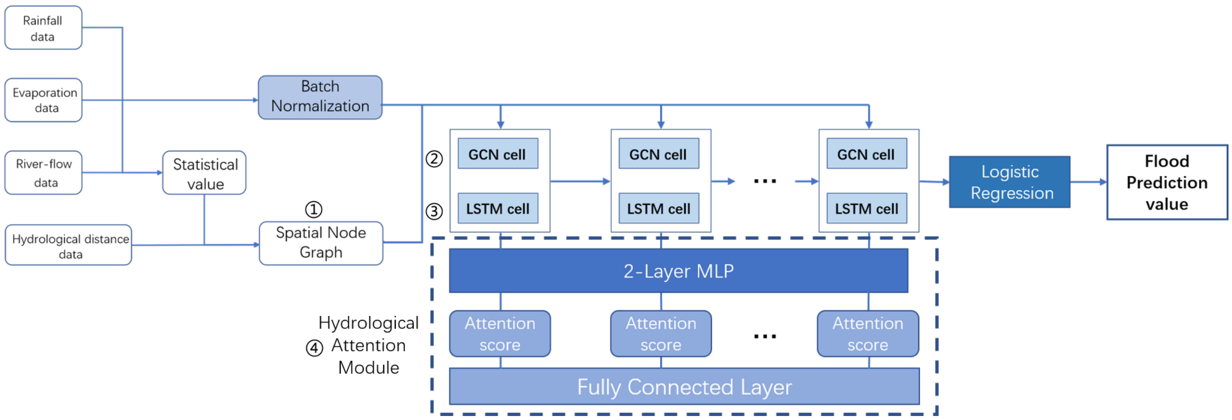



Fig. 1. After data preprocessing by ① spatial graph generation module and batch normalization, data will be input into the intelligent model we built for processing, which consists of ② GCN module, ③ LSTM module and ④ Hydrological Attenion module.

utilizes the correlation of context vectors. The accuracy of stock time series prediction is very important in the field of finance. Cheng et al. [12] proposed a hybrid model consisting of empirical mode decomposition and attention-based long-term and short-term memory network,which can effectively decompose the financial time series into multiple levels of inherent mode functions. The linear regression analysis of stock market index verifies the prediction performance of the proposed model.

The attention mechanism can effectively capture the dynamic change characteristics of the data, making the correlation analysis more accurate. In this paper, we introduce the hydrological attention mechanism based on Spatial and Temporal aware Graph Convolutional Network (ST-GCN) , with the aim of improving the selection of ST-GCN network for historical information, thus screening out information with higher value.

## III. The Proposed Method

This part will introduce the details of the proposed framework for flood prediction with four categories: 1)Overall structure; 2) Hydrological spatial graph generation module; 3) Graph convolutional network module; 4) S-GCN module; 5) Hydrological attention module.

### A. Overall Structure

Take Tunxi river and Changhua river as examples, we describe the steps to predict and analyze the hourly flood using the proposed method. For the purpose of flood forecasting, our model (ST-GCN) consists of four components. The structure of our model is shown in Fig. 1. A necessary step for flood forecasting is data preprocessing, consisting mainly includes spatial map generation and data normalization processing. We use various hydrological values as the input data of the model, including rainfall data, evaporation data, river-flow data and hydrological distance data.

$$I_t = D(v) \tag{1}$$

$$S_t = M(d, s) \tag{2}$$

where $v$ denotes the first three types of data, $I_t$ denotes the result after the normalization of $D$ at time $t$.We extract statistical features $s$ from data $v$ . Then we use $s$ and distance data $d$ to generate a spatial node graph $S_t$ after processing $M()$ . Then they are input into our model for processing.

The S-GCN module, consisting of GCN cell and LSTM cell, can effectively solve the problem of learning temporal and spatial features at the same time. First, the spatial features of the data are processed by GCN, and then the obtained values are input into the LSTM cell to learn temporal features. The above workflow is described as follow:

$$F(I_t, S_t) = LSTM\left(I_t, GCN(S_t)\right) \tag{3}$$

where $F()$ denotes the procedure of S-GCN , $GCN()$ is the convolution operation at GCN cell, $LSTM()$ is the sequential operation at LSTM cell. However, this cannot make S-GCN able to determine whether the retained information is useful. Our model uses the hydrological attention module to selectively focus information, so the model can process useless and retain useful information, thereby obtaining more accurate output prediction results. $h_t$ is the hidden state of S-GCN cell. Here, we use the attention function to assign different scores to $h_t$ at each moment.

$$A(F(I_t, S_t)) = FCL\left(C(h_t)\right) \tag{4}$$

$C()$ is the operation that can get a state vector that represents development trend information of flood process. $FCL()$ is fully connected layer, and $A()$ is the operation of hydrological attention mechanism.

$$Model(v, s) = L\left(R\right) \tag{5}$$

Finally, we utilize the logistic regression classifier connected to get flood prediction value. $R$ is the result after hydrological attention mechanism, $L()$ is the logistic regression classifier.

### B. Hydrological Spatial Graph Generation Module

First of all, we need to construct the topology of the hydrological station to explore the spatial correlation of flood flow. The spatial correlation we describe here mainly refers to the spatial connection between hydrological stations. Due to river flow states vary over time, it is better to let the graph nodes possess the varying river flow states and keep the graph structure fixed. Thus, to ensure the consistency of the definition in a graph, we use nodes to represent the hydrological stations, which can be rainfall stations or river gauging station. Then, the edges in a graph represent the strength of relationship between stations. The hydrological network and the relationship between hydrological stations can be described as an undirected graph $G = (V, E)$, where $V = \{v_1, v_2, \dots, v_N\}$ is a set of nodes, $N$ is the number of hydrological stations, and $E$ is the set of edges.

The question of how to generate topological diagrams requires comprehensive consideration of relevant knowledge in the field of hydrology and statistics. Here we construct two different graphs to show the relationships between hydrological stations. Firstly, from the hydrological point of view, if a station is in the same river as another station, the similarity of the water flow between the two stations will be high due to their hydraulic connections [13]. Secondly, from the statistical point of view, we need to calculate the degree of correlation between various stations based on the multi-year data of hydrological stations.

*1) Hydrological Distance Graph:* The hydrological connection between hydrological stations is estimated with regards to Digital Elevation Model (DEM). Here, the hydrological distance graph is mainly based on the flow distance between two stations. In the case where the water path of the upstream station (e.g., $S_i$) passes the downstream station (e.g., $S_j$), the hydraulic distance is the length of the flow path between the two stations ($d_{i,j}$).In other words, $d_{i,j}$ is the length of the river between two stations $i$ and $j$. If they have no hydrological connections, the $d_{i,j}$ is $\infty$. The adjacency matrix of Pearson correlation graph is $A_D$:

$$A_D = \begin{bmatrix} 1 & \cdots & \frac{1}{d_{1,j}} & \cdots & \frac{1}{d_{1,N}} \\ \frac{1}{d_{2,1}} & \cdots & \frac{1}{d_{2,j}} & \cdots & \frac{1}{d_{2,N}} \\ \vdots & & \vdots & & \vdots \\ \frac{1}{d_{i,1}} & \cdots & \frac{1}{d_{i,j}} & \cdots & \frac{1}{d_{i,N}} \\ \vdots & & \vdots & & \vdots \\ \frac{1}{d_{n,1}} & \cdots & \frac{1}{d_{n,j}} & \cdots & 1 \end{bmatrix} \tag{6}$$

*2) Pearson Correlation Graph:* In this paper, we not only consider the distance, but also calculate the correlations between the stations based on the flow rate per unit time for the last five years. Here we use the Pearson correlation coefficient to calculate the correlation. The adjacency matrix of Pearson correlation graph is $A_P$:

$$A_P = \begin{bmatrix} 1 & \cdots & p_{1,j} & \cdots & p_{1,N} \\ p_{2,1} & \cdots & p_{2,j} & \cdots & p_{2,N} \\ \vdots & & \vdots & & \vdots \\ p_{i,1} & \cdots & p_{i,j} & \cdots & p_{i,N} \\ \vdots & & \vdots & & \vdots \\ p_{n,1} & \cdots & p_{n,j} & \cdots & 1 \end{bmatrix} \tag{7}$$

where $p_{i,j}$ is the Pearson correlation result between station $i$ and station $j$.

*3) Hydrological Spatial Graph :* Finally, we need to merge hydrological distance graph and Pearson correlation graph into one graph. We combine these two graphs by the weighted summing their adjacency matrices at the element level.

$$a_{m,k} = \begin{cases} 1, & m = k \\ \alpha * \frac{1}{d_{m,k}} + \beta * p_{m,k}, & m \neq k \end{cases} \tag{8}$$

where $\alpha_{m,k}$ is the elementary of the mth row and $k$th column in the adjacency matrix $A$. Considering hydrological knowledge and statistical knowledge comprehensively, here we take the values of $\alpha$ and $\beta$ as 0.5 and 0.5 respectively. Now, we can get the adjacent edges and nodes of the graph through the matrix $A$, thus generating the hydrological topology graph.

### C. Graph Convolutional Network Module

Acquiring spatial correlation has an important influence on hydrological status value. In this section, we use graph convolutional network to capture the spatial features from the hydrological information. The existing graph convolutional neural networks are divided into two types: spectral convolution [8] and spatial convolution [14]. The former was applied in this study, because it is more universal and does not require complicated calculations.

The spectral method uses the convolution theorem on the graph to define graph convolution from the spectral domain. The main idea of the convolution theorem is that the Fourier transform of signal convolution is equivalent to the product of the signal Fourier transform.Using the convolution theorem, we can multiply the signal in the spectral space, and then use the inverse Fourier transform to convert the signal to the original space to achieve graph convolution. For the reason that the original spectral method has a disadvantage of high temporal and spatial complexity, Kipf et al. [15] parameterized the convolution kernel in the spectral method, which greatly reduces the temporal and spatial complexity:

$$x_j^{m+1} = h\left(\sum_{i=1}^{p} \theta \widehat{D}^{-\frac{1}{2}} \hat{A} \widehat{D}^{-\frac{1}{2}} x_i^m\right) \tag{9}$$

where $j = 1, \dots, q; x_i^m \in R^n$ represents the $i$-th input feature of the node in the $m$-th layer on the graph; $A$ is the adjacency matrix, $\hat{A}$ is a matrix with self-connection structure; $\widehat{D}$ is the degree matrix corresponding to $\hat{A}$; $p$ is the dimension of the input feature; $\theta$ is the parameter to be learned; $h$ is the activation function.

In the field of hydrology, the distribution of various stations is uneven, and spatial dependence needs to be considered. In this study, we use the GCN model to learn spatial features from hydrological data, mainly by obtaining the topological relationship between hydrological stations and surrounding stations. A 2-layer GCN model can be expressed as:

$$f\left(X_t, A\right)=\sigma\left(\widehat{D}^{-\frac{1}{2}} \hat{A} \widehat{D}^{-\frac{1}{2}} \operatorname{Relu}\left(\widehat{D}^{-\frac{1}{2}} \hat{A} \widehat{D}^{-\frac{1}{2}} X_t W_0\right) W_1\right) \tag{10}$$

where $X_t$ is the feature matrix, $\sigma()$, Relut() is nonlinear transformation, $W_0$ represents the weight matrix of first layer, $W_1$ represents the weight matrix of the second layer.

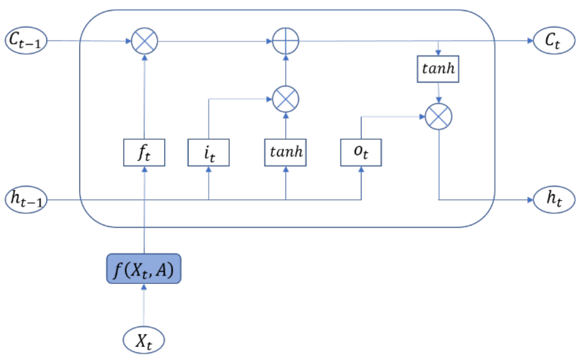



Fig. 2. The structure of S-GCN.

### D. S-GCN Module

To fully exploit the spatial and temporal correlations of the hydrological flow data, spatial aware GCN for flood prediction (S-GCN) was proposed, based on Graph convolution network and LSTM network. The memory cell unit structure in the hidden layer of S-GCN is shown in Fig.4.

$$i_t=\sigma\left(W_i\left(f\left(A, X_t\right), h_{t-1}\right)\right)+b_i \tag{11}$$

$$o_t=\sigma\left(W_o\left(f\left(A, X_t\right), h_{t-1}\right)\right)+b_o \tag{12}$$

$$f_t=\sigma\left(W_f\left(X_t, h_{t-1}\right)\right)+b_f \tag{13}$$

$$g_t=\tanh \left(W_g\left(f\left(A, X_t\right), h_{t-1}\right)\right)+b_g \tag{14}$$

$$C_t=f_t \odot C_{t-1}+i_t \odot g_t \tag{15}$$

$$h_t=\tanh \left(C_t\right) \odot o_t \tag{16}$$

where $f\left(A, X_t\right)$ represents the operation of GCN. For details, please refer to equation (10). And $i_t$ is the input gate at time $t$ that decides how much new input information is added to the next step, $o_t$ is the output gate at time $t$ which is responsible for whether the current cell value is output, $f_t$ is the forget gate at time $t$ that decides whether the current cell state is abandon. All three of them are calculated by activation function $\sigma()$. $W_i$, $W_o$ and $W_f$ is the recurrent weight matrix from the input, output and forget gates to the hidden layer, respectively. Similarly, $b_i$, $b_o$ and $b_f$ represent the bias. $X_t$ denotes the input of the current cell and $h_t$ is the output state vector at time $t$. $h_{t-1}$is the state of the hidden layer at the previous moment. $g_t$ is the result obtained by activating the tanh activation function, after performing element-wise point multiplication on $X_t$ and $h_{t-1}$, respectively. $C_{t-1}$ and $C_t$ are the state variables of the memory cell passing through the memory cell unit at the previous moment and $t$ moment.

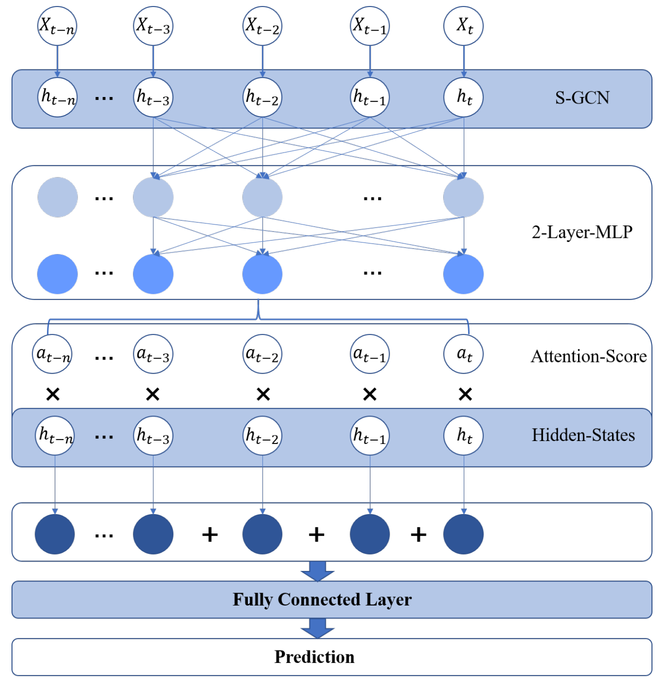



Fig. 3. The structure of our ST-GCN model.

### E. Hydrological Attention Module

In this part, we introduce the temporal attention mechanism based on S-GCN model. This attention mechanism calculates state vector that express the overall development trend of flood flow, after uses the multilayer perception (MLP) to solve the importance of hydrological information at every moment, so as to take the obtained value as the modeling object of the time series model. The model with the attention mechanism is shown in Fig. 3.

$$e_k=\sigma\left(W_k\left(\sigma\left(W_m H_t+b_m\right)\right)+b_k\right) \tag{17}$$

Here, let $h_k\ (k=1,2, \ldots, n)$ represents the hidden states at different moments, and $H_t=\left\{h_k \mid k=1,2, \ldots, n\right\}$. Let $e_k$ represents the score of the relationship between $h_m$ and $h_k\ (m, k=1,2, \ldots, n)$, and the higher value denotes the higher correlation. $W_m$ and $W_k$ denote the weights matrix of the first and second layers respectively. Similarly, $b_m$ and $b_k$ denote the bias vectors matrix of the first and second layers respectively.

$$a_k=\frac{\exp \left(e_k\right)}{\sum_{i=1}^n \exp \left(e_m\right)} \tag{18}$$

$$S_t = \sum_{i=1}^{n} a_k h_k \tag{19}$$

where $a_k$ is the attention coefficient corresponding to $e_k$, it can be calculated using a $softmax$ function. Then, the attention coefficient $a_k$ is assigned to different hidden layer states $h_k$, and the state vector $S_t$, denoting the overall development trend of flood flow, is obtained by summation. Finally, the state vector $S_t$ calculates the predicted value through the fully connected layer. This hydrological attention model can self-judge the specific important time series information through this attention mechanism, and increase the weight of the influence of this part of the information on the result, so as to achieve the purpose of improving the prediction accuracy.

## IV. EXPERIMENT

We have six models, KNN, DTree, LSTM, IndRNN, S-GCN and ST-GCN to compare and analyze in our experiments. This part will introduce the details of experiments, which includes dataset and measurement, implementation details and results.

### A. Dataset and Measurement

In order to better prove the validity of the experiment,,we chose two watershed data as the data set of this experiment, namely Tunxi watershed and Changhua watershed.

We have collected floods happened from 1981 to 2003 except 2001 in Tunxi catchment as our original dataset. In this study, we take the data of Tunxi from 1981 to 1999 as the training set, and the rest as the test set. Moreover, at another dataset, we take the data of Changhua from 1998 to 2003 as the training set, and the 2004 and 2005 as the test set. Since the hydrological data in the training set may have different magnitide, we need preprocess the hydrological data of the time series by using the method of Min-Max Normalization to standardize the data to the range of [0–1]:

$$x^* = \frac{x - \min}{\max - \min} \tag{20}$$

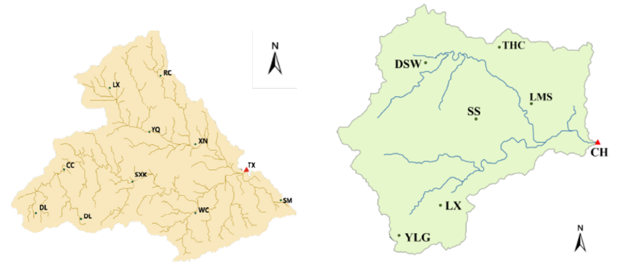
where $x$ is the value of primitive variables, max and min represent the maximum and minimum values in input values, respectively.

Here two evaluation indicators are used to evaluate the prediction performance of the model, Relative Mean Square Error (RMSE) and mean absolute error (MAE), which are used to measure prediction error.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i^{test} - y_i^{pre} \right| \tag{21}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\bar{y}_i^{pre} - y_i^{test}\right)^2}{n}} \tag{22}$$

In equation (21) and (22), $y_i^{test}$ is the actual observed value of the $i$-th sample river water flow, $y_i^{pre}$ is the $i$-th sample river water flow prediction value, $\bar{y}_i^{pre}$ is the average value of the $i$ -th sample river flow forecast, and $n$ is the number of test samples. The smaller the value of RMSE and MAE are, the more accurate the prediction is.




(a) Tunxi river dataset (b) Changhua river dataset

Fig. 4. The map of various types of hydrological stations in Tunxi and Changhua. Note that the Tunxi River station and Changhua River station are represented by red triangles, and other rainfall stations are represented by green dots.

### B. Implementation Details

This system uses Python language as the actual coding language. All experiments are carried out on a Linux server equipped with 2.10GHz 8-core Xeon CPU, 60GB RAM and Nvidia GeForce GTX 1080 Ti. For LSTM , IndRNN model, S-GCN model and ST-GCN model, the number of cycles is set to 32, the learning rate was set to 0.001. Besides, the maximum depth of Dtree we choose is 7 . Our model and baseline are implemented in the environment of TensorFlow [16]. By inputting 32 samples at a time, the error back propagation and parameter update are completed, and the epoches are set to 500 to converge to reach the final prediction target.

### C. Ablation Experiments

This part will list the performances of KNN, DT, LSTM network, IndRNN network model and S-GCN model, ST-GCN model. Note that we implement some baseline models for comparison. Firstly, we use the K-Nearest Neighbor (KNN) algorithm, which is a classic regression algorithm. The second one is Decision Tree (DT), which is widely used in the field of hydrological prediction . The third one, Long Short-Term Memory (LSTM), is a time cyclic neural network, which is specially designed to solve the long-term dependence problem of general RNN. There are many scholars who use LSTM to solve hydrological prediction problems. The next one is Independently Recurrent Neural Network (IndRNN) [17], which can retain long-term memory and handle long sequences. Its biggest advantage is that it can stack multiple layers to build a deeper network than traditional RNNs. Moreover, to emphasize the importance of the temporal information judgment ability of the flood forecasting intelligent model, we also import S-GCN model in our experiment.

The Table 1 and Table 2 shows in detail the comparison between our proposed ST-GCN network and other networks. In this study, we adjust the parameters to make various models achieve better results. From the average column , the top four models in the table are baseline models, and the bottom of the table is our model. Compared with the other models, our model has a lower average value of both RMSE and MAE, it indicates the performance of our network structure has obvious advantages over all the other baseline models. Besides, we clearly

TABLE I
PERFORMANCE OF DIFFERENT MODELS AT TUNXI DATASET

| Model | T+1 (Hour) | | T+3 | | T+6 | | T+9 | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| **KNN** | 82.07 | 17.13 | 124.08 | 35.09 | 177.54 | 55.23 | 222.28 | 67.48 | 151.69 | 43.73 |
| **Dtree** | 49.60 | 13.29 | 127.71 | 37.97 | 193.42 | 58.60 | 236.03 | 70.73 | 151.69 | 45.15 |
| **LSTM** | 39.12 | 23.81 | 96.78 | 26.45 | 128.22 | 38.89 | 179.89 | 51.84 | 101.01 | 35.25 |
| **IndRNN** | **21.60** | **12.89** | 92.05 | 36.07 | 154.48 | 41.59 | 205.98 | 55.60 | 118.53 | 36.54 |
| **S-GCN** | 45.9 | 16.13 | 93.29 | 22.32 | 121.40 | 38.19 | 149.75 | 41.1 | 102.59 | 29.44 |
| **ST-GCN** | 47.84 | 19.95 | **89.20** | **21.44** | **117.24** | **32.61** | **136.73** | **39.20** | **84.76** | **24.67** |

TABLE II
PERFORMANCE OF DIFFERENT MODELS AT CHANGHUA DATASET

| Model | T+1 (Hour) | | T+3 | | T+6 | | T+9 | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| **KNN** | 57.78 | 17.13 | 76.06 | 27.36 | 107.83 | 44.25 | 153.03 | 69.56 | 98.67 | 39.58 |
| **Dtree** | 42.74 | 13.56 | 87.06 | 29.39 | 130.58 | 49.45 | 151.506 | 65.46 | 102.97 | 39.47 |
| **LSTM** | 60.53 | 19.26 | 90.14 | 28.72 | 96.91 | 33.89 | 116.10 | 27.15 | 90.92 | 27.26 |
| **IndRNN** | **49.79** | **13.29** | 99.71 | 26.84 | 102.60 | 35.42 | 127.77 | 41.19 | 94.96 | 29.18 |
| **S-GCN** | 58.06 | 18.01 | **68.95** | **27.57** | 75.13 | 33.84 | 105.2 | 34.77 | 76.83 | 28.55 |
| **ST-GCN** | 60.23 | 18.31 | 70.07 | 28.83 | **73.93** | **28.88** | **97.71** | **32.89** | **75.84** | **27.22** |

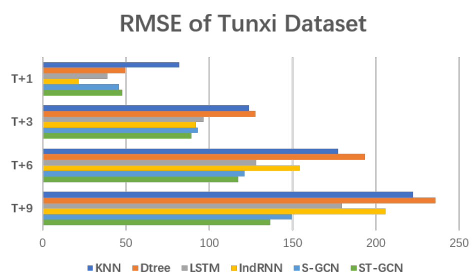



Fig. 5. RMSE comparison of models performance at Tunxi.

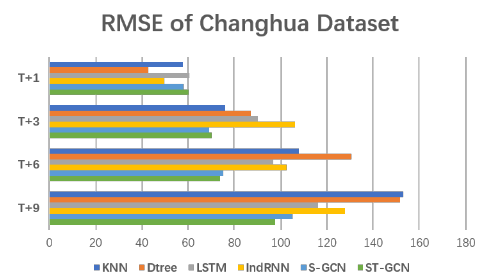



Fig. 7. RMSE comparison of models performance at Changhua.

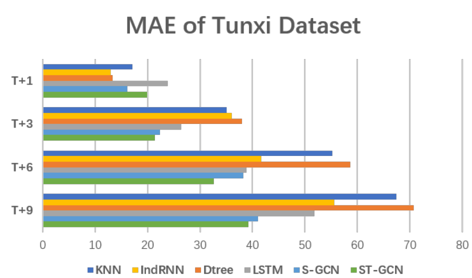



Fig. 6. MAE comparison of models performance at Tunxi.

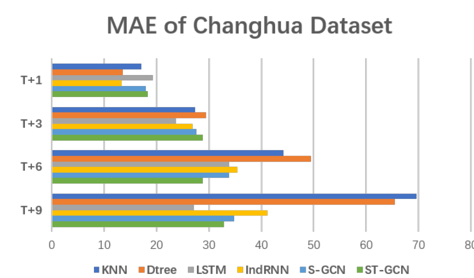
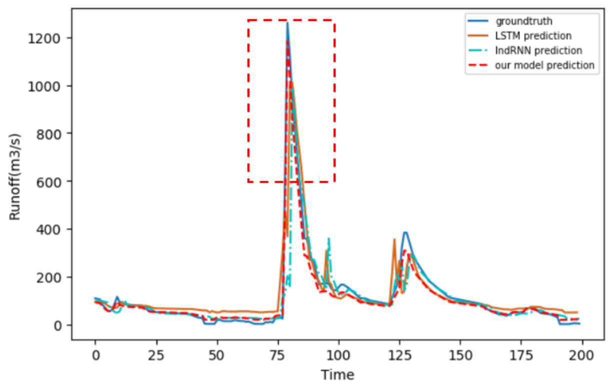


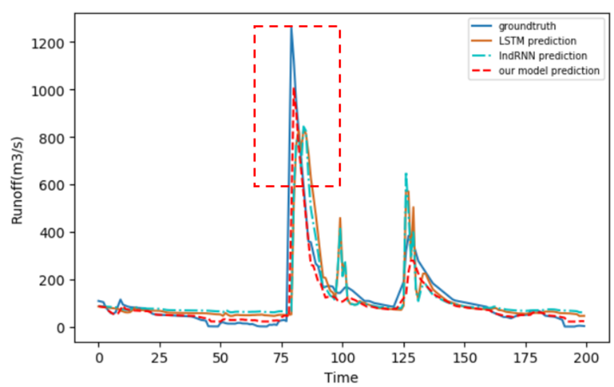

Fig. 8. MAE comparison of models performance at Changhua.

observe that DT model performs better than both LSTM and our model for flood prediction at T+1, meanwhile performs much worse than LSTM and our model for flood prediction at T+3, T+6 and T+9, which means the deep learning model based on LSTM conducted more accurate prediction in the flood disaster. Because those models not only learn the timing characteristics of the runoff sequence, but also retain long-term dependency information.The effect of the attention model is not obvious in the flood forecasting experiment with a short forecast period. After T+3, our method's RMSE and MAE are basically the smallest of all comparison experiments. Our model can be applied to the flood prediction, on the one hand, it can further explore the potential spatial relationship of the known geographic information, and on the other hand, it can enhance the ability to capture dynamic flood information over time.

As shown in Fig.5 to Fig.8, our model has higher accuracy and better robustness among the models after the first step. For IndRNN, we could find it get higher accuracy labeled by gray rectangles, when predicting at T+1. This model is more suitable for single-step prediction, since the IndRNN model of this time period deviate from the actual value due to the calculation of the neuron weight independence. In addition, from the perspective of time step analysis, the performance




(a) Result at T+3.




(b) Result at T+6.

Fig. 9. Take one flood as an example,the comparison of the ground truth flow rates and predicted flow rates computed by ST-GCN, LSTM and IndRNN of Changhua dataset.Note that the red rectangles indicates the time of flood peak.

trends of the six models change over time, with the best performance in the less time steps and the worst in the more time steps.

### D. Performance Comparison

To further illustrate the performance of the deep learning models more intuitively, Fig. 9 plots the predicted river flow of a flood in the Changhua River Basin. In the process of hydrological forecasting, the time of flood peak is very important for that the delay of flood peak forecasting will bring difficulties to flood work. We mark the peak time of the flood with a red border in Fig. 9. Here we compare the results of T+3 and T+6. In particular, for LSTM and IndRNN methods, we find that it is easy to make wrong prediction during flood peak. Instead,ST-GCN model has better performance in predicting the time of flood peak. Therefore, our model is suitable to predict floods of rivers.

## V. CONCLUSION

This paper has presented a deep learning approach based on the use of Graph Convolution Network (GCN) and Long Short-Term Memory Network (LSTM) for flood prediction to predict the stream flow which is called S-GCN. The role of GCN is to capture the spatial correlation of each feature in the hydrological data and LSTM is use to record the temporal information. Besides, our model imports the attention model to consider dynamic variation of stream flow by assign weights to adjust attention score in the hidden layer of the S-GCN at every moment. Under the task of flood prediction, our method achieves better predictive performance under different lengths of prediction horizon compared to existing baselines. For future work, we can use the attention mechanism to replace the RNN model, that is to say, improve the novel Transformer model to flood forecasting.

## REFERENCES

[1] N. A. Pierini, E. R. Vivoni, A. Robles-Morua, R. L. Scott, and M. A. Nearing, "Using observations and a distributed hydrologic model to explore runoff thresholds linked with mesquite encroachment in the sonoran desert," *Water Resources Research*, vol. 50, no. 10, pp. 8191–8215, 2014.

[2] Y. Bai, Z. Chen, J. Xie, and C. Li, "Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models," *Journal of hydrology*, vol. 532, pp. 193–206, 2016.

[3] Y. Wu, W. Xu, J. Fengt, S. Palaiahnakote, and T. Lu, "Local and global bayesian network based model for flood prediction," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 225–230.

[4] Y. Shi, K. Taalab, and T. Cheng, "Flood prediction using support vector machines (svm)." GIS Research UK (GISRUK), 2016.

[5] W. Chen, Y. Li, W. Xue, H. Shahabi, S. Li, H. Hong, X. Wang, H. Bian, S. Zhang, B. Pradhan *et al.*, "Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods," *Science of The Total Environment*, vol. 701, p. 134979, 2020.

[6] J. Wang, P. Shi, P. Jiang, J. Hu, S. Qu, X. Chen, Y. Chen, Y. Dai, and Z. Xiao, "Application of bp neural network algorithm in traditional hydrological model for flood forecasting," *Water*, vol. 9, no. 1, p. 48, 2017.

[7] Z. Liu, Y. Wu, Y. Ding, J. Feng, and T. Lu, "Context and temporal aware attention model for flood prediction," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 545–555.

[8] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.

[9] D. Marcheggiani, J. Bastings, and I. Titov, "Exploiting semantics in neural machine translation with graph convolutional networks," *arXiv preprint arXiv:1804.08313*, 2018.

[10] J. Sun, J. Zhang, Q. Li, X. Yi, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *arXiv preprint arXiv:1903.07789*, 2019.

[11] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, 2018.

[12] L. Chen, Y. Chi, Y. Guan, and J. Fan, "A hybrid attention-based emd-lstm model for financial time series prediction," in *2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*. IEEE, 2019, pp. 113–118.

[13] 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?" *Journal of hydrology*, vol. 400, no. 1-2, pp. 1–9, 2011.

[14] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in neural information processing systems*, 2017, pp. 1024–1034.

[15] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[16] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

[17] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (indrnn): Building a longer and deeper rnn," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5457–5466.