

Streamflow prediction of the Yangtze River base on deep learning neural networks: Impact of the El Niño–Southern Oscillation

Si Ha

China University of Geosciences

Darong Liu (✉ lidr1169@cug.edu.cn)

China University of Geosciences

Lin Mu

China University of Geosciences

Research Article

Keywords: El Nino, Long short-term memory, Flood Forecast

Posted Date: February 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-227553/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Streamflow prediction of the Yangtze River base on deep learning neural networks: Impact of the El Niño–Southern Oscillation

Authors: Si Ha¹, Lin Mu^{1 2}, Darong Liu¹

¹ College of Marine Science and Technology, China University of Geosciences, Wuhan 430074, China

² College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China

Corresponding author: Darong liu(e-mail: lidr1169@cug.edu.cn), Lin Mu(e-mail: moulin1977@hotmail.com)

ABSTRACT

Accurate long-term streamflow and flood forecasting has always been an important research direction in hydrology research. Nowadays, with climate change, floods, and other anomalies occurring more and more frequently and bringing great losses to society. The prediction of streamflow, especially flood prediction, is important for disaster prevention. Current hydrological models based on physical mechanisms can give accurate predictions of streamflow, but the effective prediction period is only about one month in advance, which is too short for decision making. Previous studies have shown a link between the El Niño–Southern Oscillation (ENSO) and the streamflow of the Yangtze River. In this paper, we use ENSO and the monthly streamflow data of the Yangtze River from 1952 to 2016 to predict the monthly streamflow of the Yangtze River in two extreme flood years by using deep neural networks. In this paper, three deep neural network frameworks are used: Stacked LSTM, Conv LSTM Encoder-Decoder LSTM and Conv LSTM Encoder-Decoder GRU. Experiments have shown that the months of flood occurrence and peak flows predicted by these four models become more accurate after the introduction of ENSO. And the best results were obtained on the Convolutional LSTM + Encoder Decoder Gate Recurrent Unit model.

Key words: El Nino; Long short-term memory; Flood Forecast

Lin Mu designed experiments;

Si Ha and Darong Liu completed the coding;

Si Ha carried out experiments and analyzed experimental results;

Si Ha and Darong Liu wrote the manuscript

INTRODUCTION

The Yangtze River is one of the most important rivers in China, with a large, densely populated, and economically developed river basin. Flooding in the Yangtze River is of great concern to people, and China has invested heavily in flood prevention. However, thousands of people still died in several major floods in the past three decades, and the average direct loss is more than 100 billion RMB per year ^[1]. Yangtze River streamflow forecasting plays an important role in flood prevention and post-disaster relief, as well as in integrated water resources development and utilization, scientific management, and optimal scheduling. Because many factors affect the streamflow of the Yangtze River ^[2], researchers have used various methods to predict the streamflow of the Yangtze River over the years to obtain valuable prediction data.

Runoff is a natural signal, a complex non-linear time series that is simultaneously influenced by a variety of factors such as rainfall in the basin, the degree of erosion in the basin, atmospheric circulation, and urban and rural water use. Different methods of flow prediction have been proposed by researchers for predicting runoff. These methods can be divided into short-term prediction methods, dealing with prediction times of hours ^[3, 4] to days ^[5-7], and long-term prediction methods, dealing with scales of weeks ^[8], months ^[7, 9], and even years ^[10]. These methods can also be divided according to the type of model employed: hydrological models based on physical mechanisms and data-driven models based on data analysis. Hydrological models include the Soil and Water Assessment Tool (SWAT), Top Model ^[11], and the Xinanjiang model ^[12]. These models simulate the variability and transport of elements such as water quantity and quality in a region by collecting spatial and hydrological information about the river basin to obtain a prediction of river streamflow. This class of models has been widely validated and applied to achieve river streamflow prediction. Among the data-driven models, there are traditional black box time series models such as Auto-Regressive, Moving Average, Auto-Regressive Moving Average, Auto-Regressive Integrated Moving Average, and Auto-Regressive Integrated Moving Average with exogenous input models ^[13-16]. These models look for patterns by linearly decomposing the streamflow

data, and thus perform well when the data has periodic features; however, they perform poorly in the face of complex hydrological data. Data-driven models also include artificial intelligence (AI) models, which are good at dealing with nonlinear data and can find patterns in noise; therefore, AI-based models perform well when dealing with hydrological problems. Such models include artificial neural networks (ANNs) ^[17], the support vector machine^[18], fuzzy sets^[19], evolutionary computation (such as evolution strategies)^[20], and wavelet conjunction models^[21] etc. ANNs were inspired by the structure of biological neurons ^[17] and simulate these biological neurons, essentially constructing a mapping with a large number of parameters to fit the mapping between actual observed data and predicted data. This configuration of ANNs makes them excellent at handling nonlinear data with implicit patterns. The features of hydrological data are well matched to ANNs, and thus they perform well when dealing with hydrological data.

In recent years, both physics-based hydrological models and ANNs have been studied and applied in the processing of hydrological data. The physics-based hydrological model requires a large amount of existing hydrogeological data to construct a hydrogeological model of the study area, and even uses future hydrological data, such as future rainfall data. Hydrological models based on physical mechanisms have the following drawbacks: 1) they do not yield valid results in data-poor areas; 2) they require the use of high-precision rainfall predictions to support calculations; 3) although they are highly accurate in predicting normal streamflow, they are less sensitive to anomalies (e.g., floods and droughts) and less accurate in predicting extreme weather; 4) Short-term forecasts (hours, days) are highly accurate, while long-term forecasts are less accurate. ANNs solve the streamflow prediction problem differently, and their structure can well match streamflow data. Models such as long short-term memory (LSTM) models are excellent in dealing with the time series problem, and have been widely used in natural language processing ^[22, 23], image recognition ^[24], automated driving ^[25], and time series prediction ^[26]. ANN models have the following advantages over hydrological models in predicting streamflow: 1) they require less data, and most studies have achieved good results using only streamflow

data; 2) they predict on many time scales, such as daily, weekly, monthly, and yearly time scales; 3) they are better at capturing hidden features in historical data and more accurately predict outliers.

In recent years, the main focus of ANNs has been to improve the structure of models so that they can better exploit implicit connections in the data, and discover connections with longer time horizons, thus improving prediction performance. Many ANNs have emerged to offer more and better solutions to the time series processing problem. These methods include LSTM, which is good at dealing with continuous time series; convolutional neural networks (CNNs), which are good at dealing with spatially characterized data; in addition there are Convolutional Long Short Term Memory(Conv LSTM), Gate Recurrent Unit(GRU) and Encoder Decoder structure.

The accuracy of river streamflow prediction from the perspective of training data can be improved not only by exploring correlations in streamflow history data but also by including streamflow correlations other than streamflow data in the training, thereby improving the prediction results. Previous studies have found relationships between river streamflow and various data, such as precipitation, sea surface temperature, wetness, sea level pressure, evaporation, the El Niño–Southern Oscillation (ENSO), and the East Asian Summer Monsoon (EASM). Nalley et. al. revealed a relationship between streamflow ENSO, NAO and PDO ^[27], while Wei et. al. found a relationship between the EASM and ENSO and the Yangtze River's streamflow rate. Moreover, it was found that weak EASMs and ENSOs can lead to extreme floods, while strong EASMs and ENSOs can lead to extreme droughts ^[28].

Timo et al. studied the temporal and spatial effects of ENSO on precipitation fluctuations and flood occurrence in the Mekong River Basin, and their results showed that El Niño was negatively correlated with flooding while La Niña was positively correlated with flooding. Meanwhile, the average annual flood cycle in La Niña increased by one month compared to El Niño years, and the precipitation and streamflow anomalies during El Niño were found to be larger than those during La Niña ^[29]. In a study investigating the link between streamflow volume and the ENSO in the Yangtze River, Zhang Zizhan et al. used GRACE data to investigate the link between

terrestrial water storage and ENSO in the Yangtze River basin. The upstream streamflow and ENSO phases are inversely correlated while the downstream streamflow and ENSO phases are positively correlated^[30]. Furthermore, Jiang Tong et al. point out La Niña is strongly associated with drought events and El Niño related to floods in the middle and lower Yangtze River basin, while the opposite is true in the upper Yangtze River basin^[31].

The types of data used for solving streamflow prediction problems with artificial intelligence include streamflow, precipitation, sea surface temperature, wetness, sea level pressure, and evaporation. Such as S. Sharma compared the differences between the adaptive neuro-fuzzy inference system and the Loading Simulation Program in C++ model using these types of data, and found that the two methods produced similar results^[32]. Typically, the data used for streamflow prediction using ANNs are streamflow, evaporation, and precipitation; ENSO data has not been used^[3, 7, 33]. To investigate whether the introduction of ENSO values into the streamflow prediction problem will help improve the accuracy of streamflow prediction, the present paper adds ENSO values to the training data of several better-performing and widely used ANN models.

METHODOLOGY

A. LONG SHORT-TERM MEMORY

Long short-term memory (LSTM) was proposed by Sepp Hochreiter et al. in 1997^[34]. It is an algorithm based on the recurrent neural network (RNN). LSTM solves the vanishing gradient problem by introducing three thresholds and two memory states^[35].

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$

$$h_t = o_t \circ \tanh(C_t) \quad (4)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (5)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (6)$$

LSTM consists of three gates: input gate i_t ; forget gate f_t ; and output gate o_t . The two mnemonic states are the cell state C_t and candidate state \tilde{C}_t . The formulas used by LSTM are Equations (1)–(6). W_i , W_f , and W_o comprise the matrix of parameters to be trained. b_i , b_f , and b_o are the biases to be trained. x_t is the entered

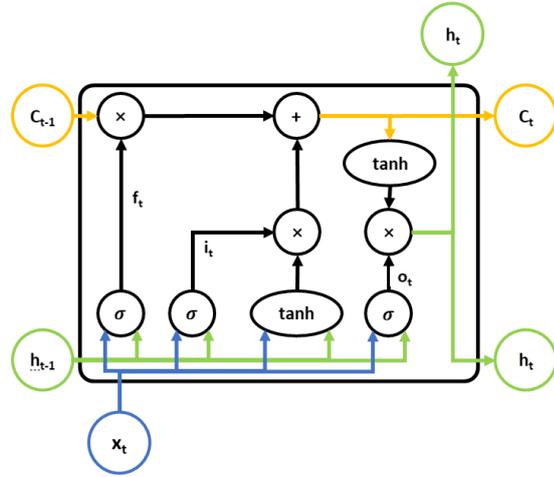


Figure 1 The networks of LSTM memory unit

data. h_{t-1} is the result of the last moment of memory. h_t represents the short term memory and the C_t cell state represents the long term memory.

The formula for the input gate is (1), the formula for the forget gate is (2), and the formula for the output gate is (3).

The tanh activation function limits the output to between -1 and 1, and can be replaced by other activation

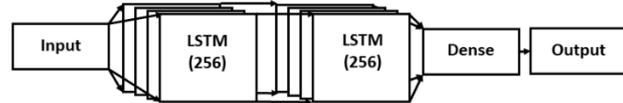


Figure 2 Stacked LSTM Structure

functions. The three gates multiply the input data and the memory of the previous moment and output. Equation (4) is the formula for memory, which is the result of multiplying the output data from the current output gate with the cell state that has undergone the tanh function; the memory represents the short term memory resulting from the action of the output and the long term memory. The cell state represents the long term memory, and is calculated as in (5) by multiplying the cell state at the previous moment through the forget gate by the candidate state. The candidate state represents the information to be deposited in the cell state, and is calculated as in (6); it is the result of the action of the current input data and the output data from the previous moment. Fig.1 shows the structure of an LSTM memory unit.

In this paper, the LSTM model is used in stacked LSTM and convolutional LSTM encoder–decoder LSTM. Fig.2 illustrates the Stacked LSTM used for the experiments

in this paper. The body of the model is a two-layer LSTM containing 256 memory cells. And output through a layer of Dense. Three layers of LSTMs are used as decoder in convolutional LSTM encoder–decoder LSTM to decode the encoded vectors and output them through the Dense layer.

B. GATE RECURRENT UNIT

Gate Recurrent Unit (GRU) was proposed by Cho et al. in 2014 [36] to solve the vanishing gradient problem in RNN networks. The GRU can be regarded as deformation of LSTM. It has fewer parameters than LSTM, and can produce the same excellent results as LSTM in some cases. The features of the GRU make it possible to shorten the computation time without

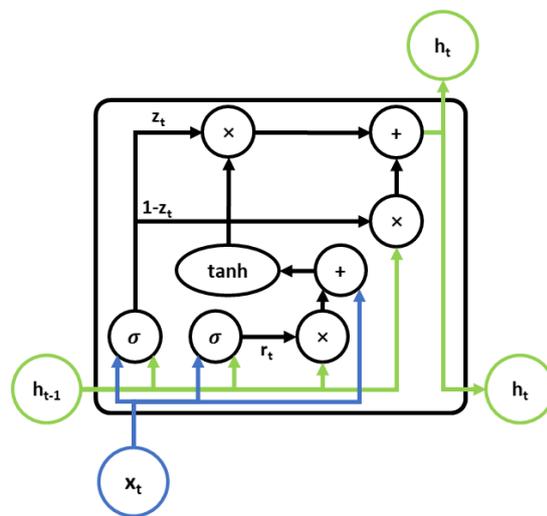


Figure 3 The networks of GRU memory unit

affecting the prediction performance and even produce better results, thus making it a frequently used model in machine learning [37].

The GRU is similar in principle to LSTM, with an update gate (7), a reset gate (8), a memory (9), and a candidate hidden layer (10). σ is the Sigmoid function, which limits the output to the range 0 to 1, and the tanh function limits the output to the range -1 to 1. W_z , W_r and W is

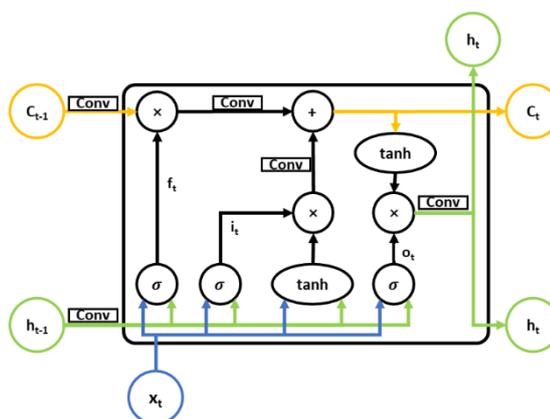


Figure 4 The networks of Conv LSTM memory unit

the parameter matrix. Both the update gate and the reset gate are calculating the memory of the current input and the previous moment. But the updates gate determines the update of the memory and controls how much of the previous moment's memory and the current input data can be retained in the current memory. The reset gate also

determines the update of the memory by controlling the candidate state, but controls how much of the information from the previous memory is forgotten. The candidate hidden layer represents the memory formed at the current moment. Fig.3 shows the structure of a GRU memory unit.

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (7)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (8)$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \quad (9)$$

$$\tilde{h}_t = \tanh(W[r_t \circ h_{t-1}, x_t]) \quad (10)$$

In the experiments set up in this paper, the convolutional LSTM encoder–decoder GRU (Conv LSTM Encoder-Decoder GRU) uses a three-layer GRU as the decoder structure to decode the encoded vectors and output them through the dense layer.

C. CONVOLUTIONAL LSTM NETWORK

The convolutional LSTM network (Conv LSTM) was proposed by Shi Xingjian et al. in 2015 [38]. In the past, LSTM was used as the encoder layer when building encoder–decoder models; however, LSTM has no special design for spatial-temporal sequences and uses full connections between layers to transform data. Meanwhile, Conv LSTM uses convolution instead of full connections to transform data. Conv LSTM has roughly the same formula as LSTM, using formulas (11)– (16), but the * stands for convolution instead of a full-connection operation; otherwise, the meaning and function of each formula is as in the LSTM and described above. According to Shi Xingjian's article, a larger kernel can perceive features with larger spatial variation in the data while a smaller kernel can perceive features with a small spatial variation. Fig.4 show the structure of a Conv LSTM memory unit.

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (11)$$

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (12)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = o_t \circ \tanh(C_t) \quad (14)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (15)$$

$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (16)$$

Conv LSTM was originally developed to process a series of radar wave images and extract the motion of clouds according to the time series of radar wave images, thus giving accurate short-term predictions. In this paper, the streamflow data and ENSO data are 1-dimensional data that change with time. When using Conv LSTM, the time series are first grouped according to different periods, and then the grouped 1-dimensional data are treated as special 2-dimensional data, and the streamflow data and ENSO data are composed of a sequence with two channels fed into the Conv LSTM network. After the above procedure, the convolutional kernel extracts the feature information from the time series as spatial features, thus increasing the accuracy of the prediction.

D. Conv LSTM ENCODER–DECODER RNN

The encoder-decoder model was proposed by Ilya Sutskever et al. to solve the problem of needing a sufficient amount of annotation data for training traditional deep neural networks (DNNs)^[39]. The encoder-decoder structure is shown in Fig. 5. The encoder encodes the input field into a vector and

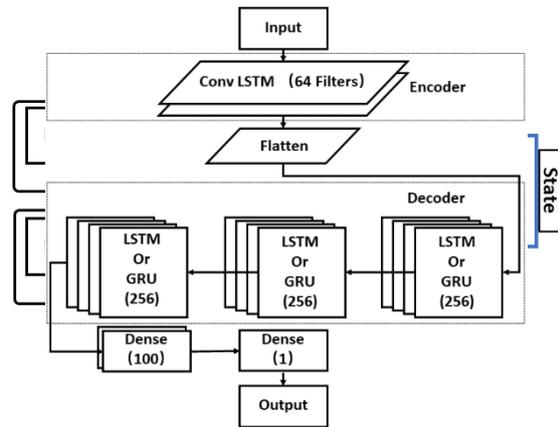


Figure 6 Conv LSTM Encoder Decoder

the decoder decodes the encoded vector into the output field. Ilya Sutskever et al. found encoder-decoder structure constructed by the LSTM model handles the translation results similar to the best translation results at that time. Therefore, the encoder-decoder structure is often used to handle the sequence-to-sequence problem. The encoder-decoder model has one feature when dealing with the sequence-to-sequence problem: It is sensitive to the order of the input sequences, which means that encoder-decoder may perform well in dealing with the time series problem. Since streamflow prediction

using time series data consisting of streamflow and ENSO values to predict future streamflow data can also be used as a sequence-to-sequence problem, the encoder-decoder structure is chosen for our experiments.

The Conv LSTM encoder-decoder RNN used in this paper uses encoder-decoder as the model framework (Fig.6), and the encoder uses a Conv LSTM with 64 convolutional kernels, and the size of the convolutional kernels is adopted as $(n,3)$, where n is the number of training data feature values. $n=1$ when the data is only streamflow data, and $n=2$ when the training data contains both flow data and ENSO data. The step size of the extracted time series features increases when the convolution kernel becomes large, while the performance of the extracted time features is close to that of an ordinary LSTM when the size of the convolutional kernel is too small. Therefore, the size of the convolution kernel is chosen to be 3. The encoder output is transformed into a one-dimensional feature vector through the Flatten layer and then fed into the decoder, which uses a three-layer LSTM or GRU with 256 memory units, and finally the decoder data is transformed into a prediction output through the Dense layer. The structure combines multiple models while inheriting the advantages of each model. The Conv LSTM as an encoder has excellent temporal feature extraction capability and can sense the change of data over time, while the LSTM and GRU as a decoder have similar time series processing capability, and the GRU simplifies the number of parameters compared to the LSTM and saves computational resources.

EXPERIMENT

A. STUDY AREA AND DATA

The Yangtze River is the most important water system in China and the fifth-largest in the world in terms of streamflow volume. The source region is in the alpine zone with 300–400 mm precipitation; the upper reaches are mostly in the sub-humid zone with 400–800 mm precipitation; and the middle and lower reaches are in the humid zone with 800–1600 mm precipitation. The middle and lower reaches of the Yangtze

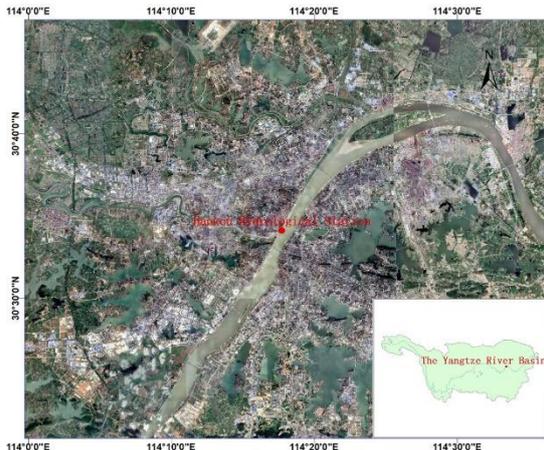


Figure 7 The location of the Yangtze River basin and Hankou Hydrological Station in Wuhan



Figure 8 The location of the Yangtze River basin and Datong Hydrological Station in Chi Zhou

River are the areas with the most severe flooding, especially the area between the confluence of the Yangtze and Han rivers and Datong [40]; the floods that occurred in 1998 and 2016 caused great economic losses in the Yangtze River basin.

In this experiment, we use streamflow data from the Hankou and Datong hydrological stations. The streamflow data are the monthly streamflow data of the Yangtze River from January 1952 to December 2016 recorded at the Hankou and Datong hydrological stations. The Hankou hydrological station is located in the middle reaches of the Yangtze River at the confluence of the Han and Yangtze rivers (Fig. 7), and controls a watershed area of 1,488,000 km². The Datong hydrological station is located in the lower reaches of the Yangtze River, at the upper end of the Chaohe section of the Yangtze River (Fig. 8), and is the main control station for the streamflow of the main stem of the Yangtze River, with a control basin area of 1.705 million km². The prediction time intervals are January–December 1998 and January–December 2016. In 1998, the second basin-wide flooding occurred in the Yangtze River basin, and was characterized by high volume, prolonged flooding, and severe coastal flooding [41]. The average monthly water level at Hankou and Datong stations in April was a record high, and the average monthly water level at Hankou and Datong stations in June was about

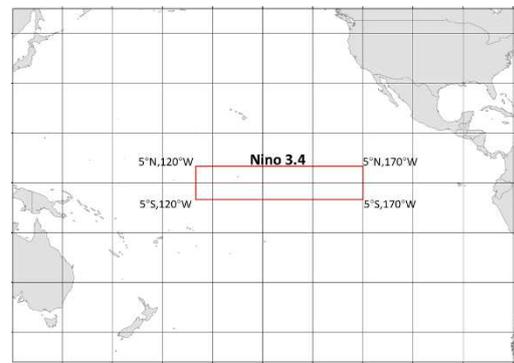


Figure 9 The Nino 3.4 area

2 m higher than that of the same period in history^[42]. The model performance is verified by predicting these two abnormal years.

As a large-scale ocean-atmosphere phenomenon in the tropical Pacific, the El Niño–Southern Oscillation (ENSO) is the most important source of interannual climate variability. El Niño represents oceanic warming in the tropical central-eastern Pacific and La Niña is the opposite. Southern oscillation is characterized by a seesaw of sea level pressure between the tropical western and eastern Pacific. The occurrence of ENSO is accompanied by a series of high-intensity climate anomalies. ENSO events influence the ecosystem, agriculture, and extreme weather of a region.

Generally, ENSO can be described by the Niño index. That is a 3-month running mean of sea surface temperature anomalies in the Niño 3.4 region (5°N–5°S, 120°W–170°W). Fig.9 shows the Niño 3.4 area. The data collected here are used as ENSO values and for the training of the model.

B. NORMALIZATION

Before the data are fed into the neural network, the streamflow data and ENSO data from Hankou and Datong stations are normalized using Equation (1) for fast convergence and stability of the model during training. Z_i is the normalized data, ranging from 0-1, X_i is each data, $\max(X)$ and $\min(X)$ are the maximum and minimum values of the data respectively.

$$Z_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (17)$$

C. NEURAL NETWORK CONSTRUCTION

In this paper, three neural networks are used: the LSTM model, the GRU model, and the Conv LSTM model. These three neural networks are used to build three model frameworks: stacked LSTM, Conv LSTM encoder–decoder LSTM model and Conv LSTM encoder–decoder GRU model. These three model frameworks, ranging from simple to complex, are used to compare the effects of different numbers of eigenvalues of training data on the accuracy of flood prediction. The training data are used for both the 1-feature training data and the 2-features training data: 1-feature data contains only

monthly streamflow data, while 2-feature data contains monthly streamflow data and ENSO values. We divide the training sets into four cycles: 6m-min-pd(6 months minimum prediction periods), 12m-min-pd, 18m-min-pd, and 24m-min-pd. We then compare the effects of the different number of features on the accuracy of flood prediction by assessing the training results produced under these four cycles. The “N” m-min-pd indicates a set of “N” months of data for training.

D. PERFORMANCE EVALUTION

After the model has completed its predictions outputting data in the range 0-1, the normalized data Z_i needs to be reduced to the original size data X_i using equation (18) when performing the evaluation.

$$X_i = Z_i * (\max(X) - \min(X)) + \min(X) \quad (18)$$

To measure the difference between the true and predicted values, we used the following four statistics.

The root mean square error (EMSE) is defined as Equation (19).

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (19)$$

The RMSE is the inverse square of the mean square error. The inverse square method reduces the MSE by an order of magnitude so that the scale of the result is the same as that of the original data, making it possible to compare the results more intuitively. When evaluating data that are expected to follow a Gaussian distribution, the RMSE is more suitable than the MAE to reflect the model performance^[43].

The coefficient of determination (R^2) is defined as Equation (20).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (20)$$

The coefficient of determination reflects what percentage of the fluctuations in the predicted value y_i can be explained by the fluctuations in the observed values^[44]. The decision coefficient takes values in the range $-\infty$ to 1. R^2 close to 1 indicates that the fluctuations in the predicted values are well explained by the fluctuations in the

observed values. On the contrary, a smaller R^2 value means that the fluctuations in the predicted values are less linearly related to the observed values and the predicted values are not well explained by the observed values.

Willmott's Index of agreement (WI) is as shown in Equation (21).

$$WI = 1 - \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \right] \quad (21)$$

WI is often used in the measurement of hydrological data. It is dimensionless, and is bounded by $-\infty$ and 1.0. It is also quite flexible and is suitable for a wide range of model performance problems. In general, it is more related to model accuracy than are other indices. It is proposed by Nash and Sutcliffe in 1970, Watterson in 1996, Legates and McCabe's in 1999, Mielke and Berry in 2001 and refined by Willmott in 2011^[45].

Legates–McCabe's Index (LMI) is written as Equation (22).

It is not oversensitive to extreme values and can reflect additive and proportional between model predictions and observations. The index is better suited as a complement to assessment instruments than other correlation-based assessment instruments. It is also dimensionless, bounded by 0 and 1.0. And higher the LMI value, the better the fitting effect of the model^[46].

$$LMI = 1 - \left[\frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{\sum_{i=1}^n |y_i - \bar{y}|} \right] \quad (22)$$

Among all the equations, where n represents the number of data pairs, y_i is the observed values, \hat{y}_i represents the forecasted value and \bar{y} represents the mean of observed values.

RESULTS

As mentioned above, the monthly streamflow forecasts of the Yangtze River have important reference value for flood prevention, and the trained model needs to provide accurate forecasts not only in normal months but also relatively accurate forecasts of

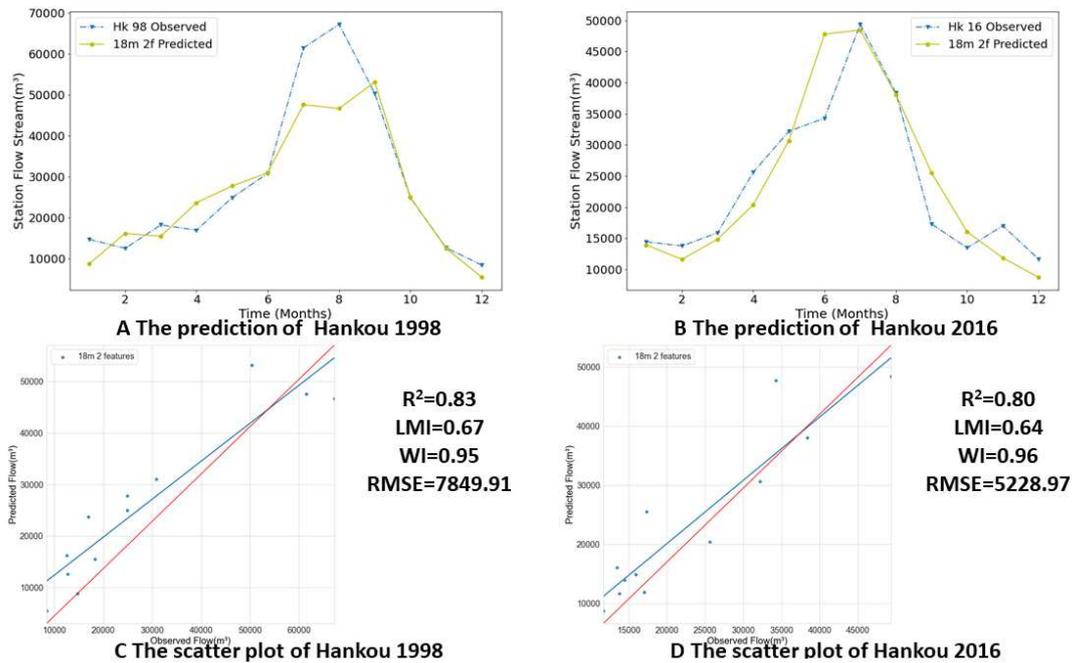


Figure 10 The result of Hankou Station

peak flows. In our experiment, the monthly streamflow forecasts of the Hankou and Datong stations, two important control stations in the middle and lower reaches of the Yangtze River, are made for the years 1998 and 2016 using Yangtze River monthly streamflow data and ENSO values. The dataset was split using 18-month groups and fed into the Conv LSTM encoder-decoder GRU model for prediction. Fig.10 shows the

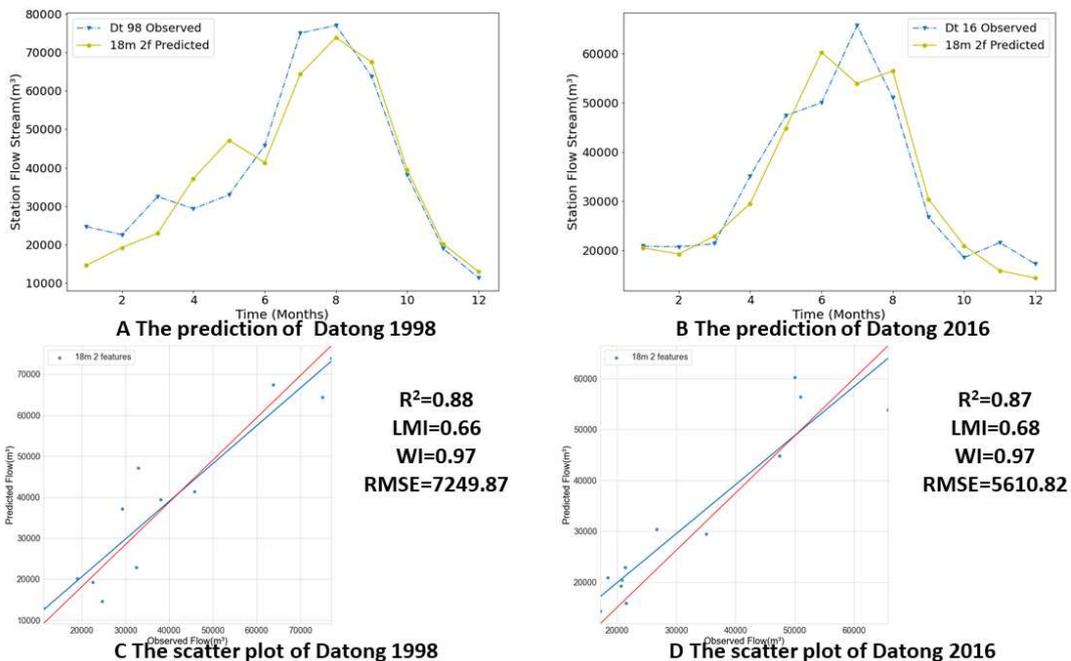


Figure 11 The result of Datong station

prediction results for 1998 and 2016 for the Hankou station, and Fig. 11 shows the prediction results for 1998 and 2016 for the Datong station.

The flooding trends of the Yangtze River in the past years show that flooding usually begins to converge in the middle reaches of the Yangtze River, with flood peaks in the middle and lower reaches of the river posing a great threat. As an important control station of the middle reaches of the Yangtze River, the Hankou station reflects the overall level of flooding in the middle reaches of the Yangtze River; this station's streamflow is also an important indicator for flood control in Wuhan as well as downstream areas. The 1998 flood streamflow was huge, reaching a peak flow of nearly 70,000 m³ in August, and the streamflow in July nearly as high. The Hankou 1998 forecast has large deviations for July and August: 13,837.25 m³ between July's forecasted and observed values, and 20,548.57 m³ between August's forecasted and observed values. In September the predicted flow was close to the observed value and differed by 2758.31 m³. the predicted time of occurrence of the flood peak differed by one month from the observed value and the predicted value of the flood peak differed from the observed value by 14077.69 m³. In other months, the predicted value fluctuates with the observed value but remains very close. From the evaluation index, it can be found that the values predicted by Hankou station in 1998 is close to the observed values as a whole, but there is still a big gap in the flood prediction. The R² value reaches 0.83; the LMI reaches 0.67; the WI value is low, only 0.95; and the RMSE value is 7849.91 m³. Furthermore, the slope of the regression line of the scatter plot is big, and therefore the overall prediction results have some degree of accuracy, but there is a big difference between the prediction and observation results of the flood peak. Fig.10 B and D show the prediction results of the Hankou station in 2016, when the Yangtze River's streamflow was smaller than that during the 1998 flood. In 2016, the peak flow occurred in July (49355 m³), and the duration of the flood was shorter than that of the 1998 flood. The flood's peak passed by August, and dropped to 49355 m³ in September. The model predictions are approximately the same as the observed values; only in June is there is a large deviation between the predicted value and the observed value (a difference of 13493.98m³). Additionally, the predicted values for July and

August are very close to the observed values. The peak flood prediction in July has only a 909.99m^3 difference from the observed value. The overall predicted RMSE was 5228.97m^3 , the R2 value reached 0.80, the LMI was 0.64, and the WI was 0.96. Thus, better forecasting results are obtained for the 2016 flood, not only in non-flood months but also in flood months, and the forecasts can be considered accurate.

The 1998 floods not only caused damage in the middle reaches of the Yangtze River, but also resulted in persistently high-water levels in the lower basin compared to previous years, with monthly flows of nearly $80,000\text{ m}^3$ observed at the Datong station, which continued from July to August and remained above $60,000\text{ m}^3$ until September. Fig.10 A, B shows the results of the 1998 flow forecast using the model with the same parameters as the Hankow station forecast model above. The predicted values above and below the observed values from January to June, and the predicted values for July, August, and September are close to the observed values. A large difference of $10,655.86\text{ m}^3$ is seen between July's predicted and observed values, and the peak flow prediction for August was accurate and differed from the observed value by 3186.73m^3 . Meanwhile, the predicted values from September to December are almost the same as the observed values. The regression line of the predicted and observed values is very close to the red line representing the case where predicted values perfectly match observed values.

Fig 11 B and D show the 2016 prediction results of the Datong station. It can be seen that the overall predicted values are close to the observed values, but the predicted values have a different trend than the observed values. For example, June's predicted values are larger than its observed values, while July's predicted values are smaller than its observed values. Fig.11 C, D shows that the model gives more accurate predictions at Datong Station in 1998 and 2016. Both predictions have an R2 of approximately 0.88, an LMI greater than 0.65, a WI of 0.97 and RMSE values between $5,000$ and $7,000\text{ m}^3$, which mean predictions is accurate.

COMPARISONS AND ANALYSIS

The proposed model uses different segmentation methods to divide the training

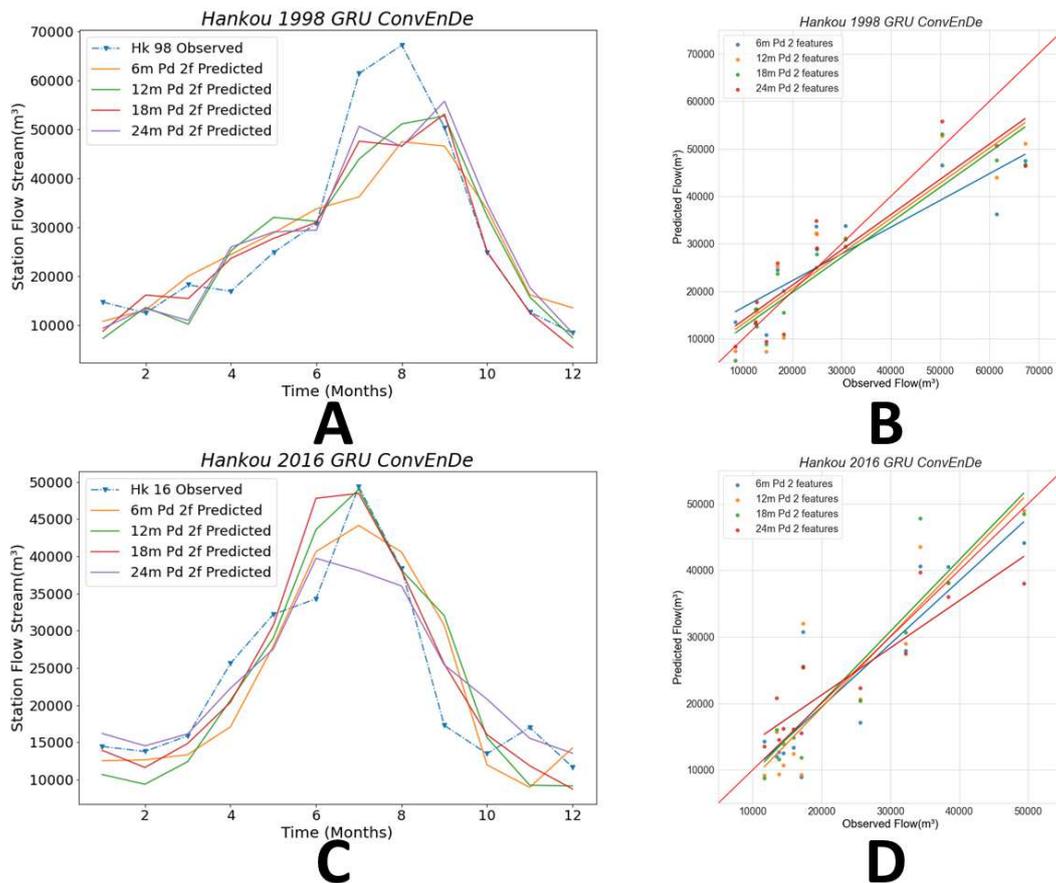


Figure 12 Comparison among the Conv Encoder Decoder GRU with different period in Hankou station

set's samples. The segmented time series data of different lengths contain input and output sequences, and are fed to the neural network for training. Finally, the trained network is verified by using a validation set. Disordering the training data is a necessary

operation for the neural network since the disordered data can increase the stability and robustness of the neural network and prevent the model from converging to the local optimal solution too quickly and overfitting. Due to the different lengths of the time series, the implicit links contained in the time series are also

Table 1 Comparison of criteria in Conv Encoder Decoder GRU with different period in Hankou station

Year		6m min pd	12m min pd	18m min pd	24m min pd
1998	RMSE	10224.25	8536.23	7849.91	8568.73
	WI	0.89	0.94	0.95	0.94
	LMI	0.54	0.58	0.67	0.58
	R2	0.72	0.80	0.83	0.80
	2016	RMSE	5974.99	6157.50	5228.97
	WI	0.94	0.94	0.96	0.94
	LMI	0.53	0.54	0.64	0.61
	R2	0.74	0.72	0.80	0.80

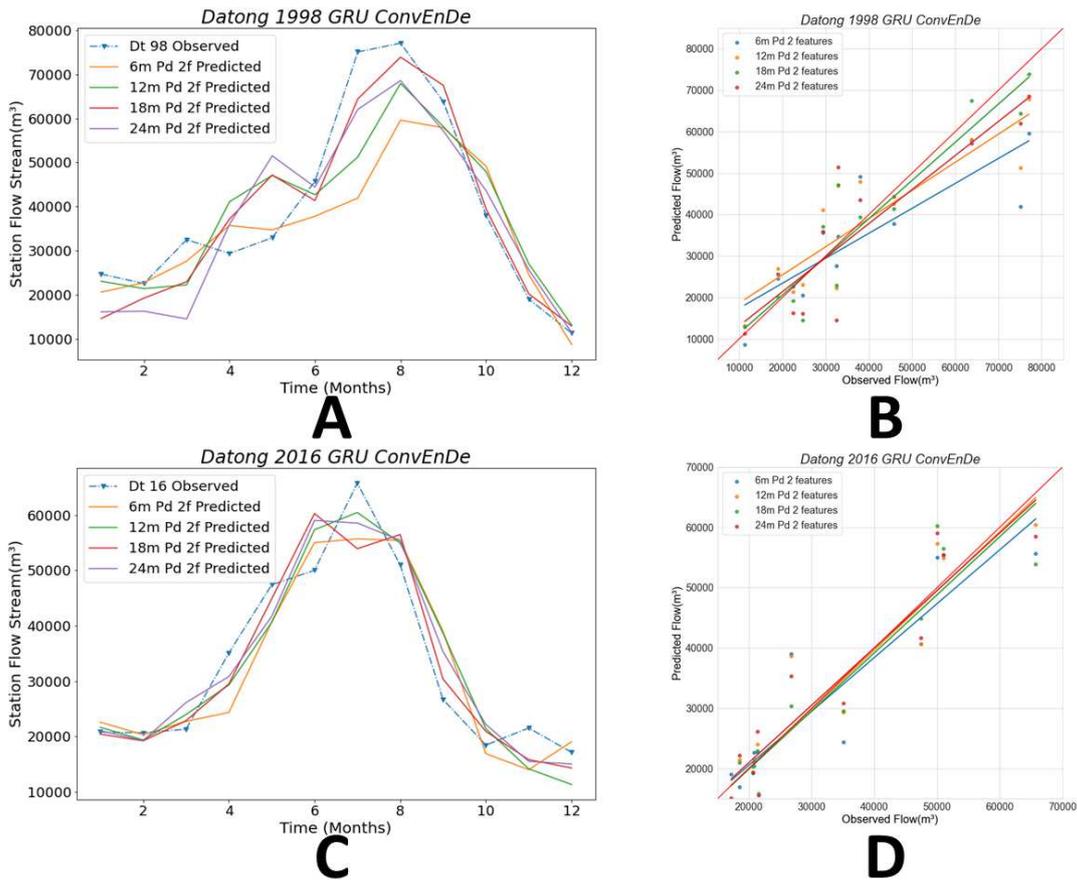


Figure 13 Comparison among the Conv Encoder Decoder GRU with different period in Datong station

different. The streamflow data and ENSO data have corresponding implicit rules in different period scales, and these implicit rules directly affect the training effect and prediction accuracy of the model. It is clear that streamflow vary on an annual cycle. Therefore, in our experiment, we select the period multiplied by the annual cycle to

Table 2 Comparison of criteria in Conv Encoder Decoder GRU with different period in Datong station

Year		6m min pd	12m min pd	18m min pd	24m min pd
1998	RMSE	12074.85	10430.62	7249.87	9954.34
	WI	0.88	0.92	0.97	0.94
	LMI	0.52	0.52	0.66	0.52
	R2	0.67	0.75	0.88	0.77
2016	RMSE	6618.38	5957.79	5610.82	5462.29
	WI	0.95	0.97	0.97	0.97
	LMI	0.62	0.63	0.68	0.66
	R2	0.82	0.86	0.87	0.88

observe the prediction accuracy of different cycle time series, which have four lengths: 6 months, 12 months, 18 months, and 24 months. By doing this, we can make accurate predictions with results close to those obtained by non-time series models. When the length of the selected time series is

too long, the number of time series segments that can be segmented from the data decreases, and the monthly streamflow data from 1952 to 2016 is too small for machine learning. Therefore, the aforementioned lengths of time series data are selected. The Conv LSTM encoder-decoder GRU model, which is the most complex model, is selected for comparison. Below, we present the prediction results of the Conv LSTM encoder-decoder GRU model on the streamflow + ENSO dataset for the Hankou station and Datong station for 1998–2016 with different time series lengths.

Fig.12 shows the prediction results for the four time series lengths for the Hankou station in 1998 and 2016. Figs 12 A and B show the predicted streamflow of the Hankou station in 1998 using the streamflow data. It can be seen that the predicted values obtained with the four lengths are close to each other. The overall trend of the predicted values obtained using 6m-min-pd is flat. The overall trend of predicted values using 12m-min-pd, 18m-min-pd, and 24m-min-pd fluctuates widely. 24m-min-pd is closest to the observed value in the peak flow prediction, followed by 12m-min-pd and 18m-min-pd. The largest differences between predicted and observed flood values were obtained for 6m pd, while 18m pd gave near-unanimity with observations from September to December. In Fig.12B, the regression lines for 12m-min-pd, 18m-min-pd, and 24m-min-pd are nearly identical to the red line; meanwhile, the regression line for 6m-min-pd is very far from the red line. By observing Tab.1, it can be found that the four evaluation indexes of 18m-min-pd are better than the results obtained from other datasets. The RMSE reaches 7849.91 m^3 , the WI value reaches 0.95, the LMI value is significantly different from those of other cases, and the LMI is 16%–24% higher than in other cases, and the R2 is 4%–15% higher than those of other cases. The best prediction was achieved in the 1998 results for 18m-min-pd at Hankow Station. Fig 12 C and D show the predicted streamflow of the Hankou station in 2016 using the streamflow data. It can be seen that the most accurate prediction is 12m-min-pd, followed by 18m-min-pd, 6m-min-pd, and finally 24m-min-pd. Similarly, the scatterplots and corresponding regression curves in D show that the results for 6m pd, 12m pd, and 18m pd are very close to the red line; meanwhile, 24m-min-pd has the poorest results and deviates greatly from the red line. Furthermore, 24m-min-pd has the

smallest RMSE value (5196.74m³), with 18m-min-pd close behind; the other two cases have large RMSEs. Additionally, 18m-min-pd has the highest WI and LMI values (0.96 and 0.64, respectively). Therefore, 18m-min-pd showed the best results and highest accuracy for predictions.

Fig.13 shows the prediction for the four segmentation methods of streamflow + ENSO data at the Datong station. Tab.2 also shows that the 18m-min-pd indicator is excellent, with a WI of 0.97 and an LMI of 0.66. The maximum WI is 0.94 and the maximum LMI is 0.52. Fig.13C and D show the results of the four data sets for predicting 2016 flows at the Datong station, which are very close, with poor overall prediction accuracy for 6m-min-pd. 12m-min-pd, 18m-min-pd, and 24m-min-pd predictions are similar. The 18m-min-pd and 24m-min-pd predicted peak flows one month earlier than observation. 12m-min-pd was accurate for the month of peak flow, and followed observed values; meanwhile, 18m-min-pd has the largest difference between forecast and true values in July. In Fig. 13 D, the regression line for 6m-min-pd is slightly off the red line, while the regression lines for the remaining results are very close to the red line. The evaluation indicators presented in Tab.2 show that all four results have reached very high values, with the best results for 18m-min-pd and 24m-min-pd.

Combining the prediction results for two years, we can find that 18m-min-pd outperforms the other datasets in most cases, and gives predictions suitable for later reference. We can conclude that the model with 18m-min-pd performs well on the streamflow + ENSO dataset.

In this paper, we experiment with stacked LSTM, Conv LSTM encoder-decoder LSTM, and Conv LSTM encoder-decoder GRU, and select the model with the most

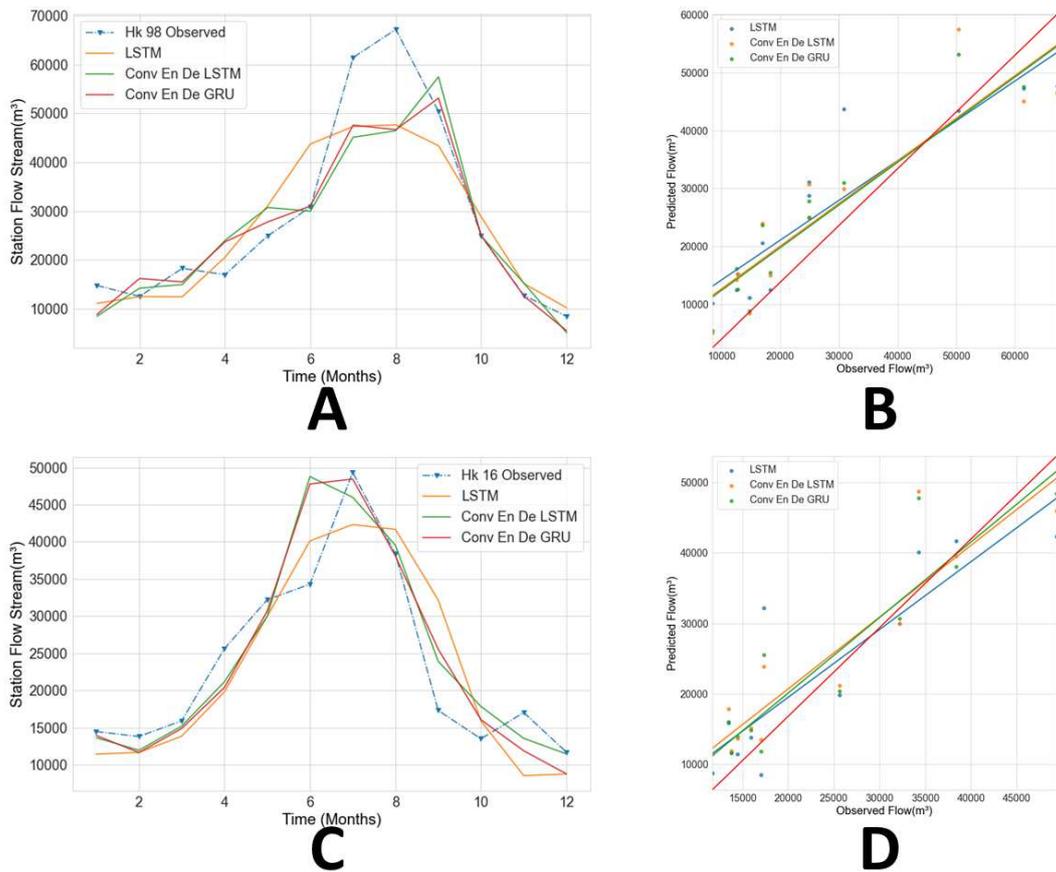


Figure 14 Comparison with different models in Hankou station

accurate predictions. The results are presented in Figs. 14 and 15 and Tabs.3,4. Fig.14 A shows the line graphs of the predictions of the three models for the Hankou station, and it can be seen that the predictions of LSTM are quite different from those of the other models. Conv LSTM encoder-decoder LSTM and Conv LSTM encoder-decoder

Table 3 Comparison of criteria with different models in Hankou station

Hankou				
Year	Model	Conv En De GRU	Conv En De LSTM	Stacked LSTM
1998	RMSE	7849.91	8675.98	8757.40
	WI	0.95	0.94	0.93
	LMI	0.67	0.61	0.58
	R2	0.83	0.80	0.79
2016	RMSE	5228.97	5216.75	6184.42
	WI	0.96	0.95	0.93
	LMI	0.64	0.65	0.51
	R2	0.80	0.80	0.72

GRU have similar predictions, but with Conv LSTM encoder-decoder LSTM having better predictions in September, and Conv LSTM encoder-decoder GRU having better predictions in July. Meanwhile, the regression lines of the three models are

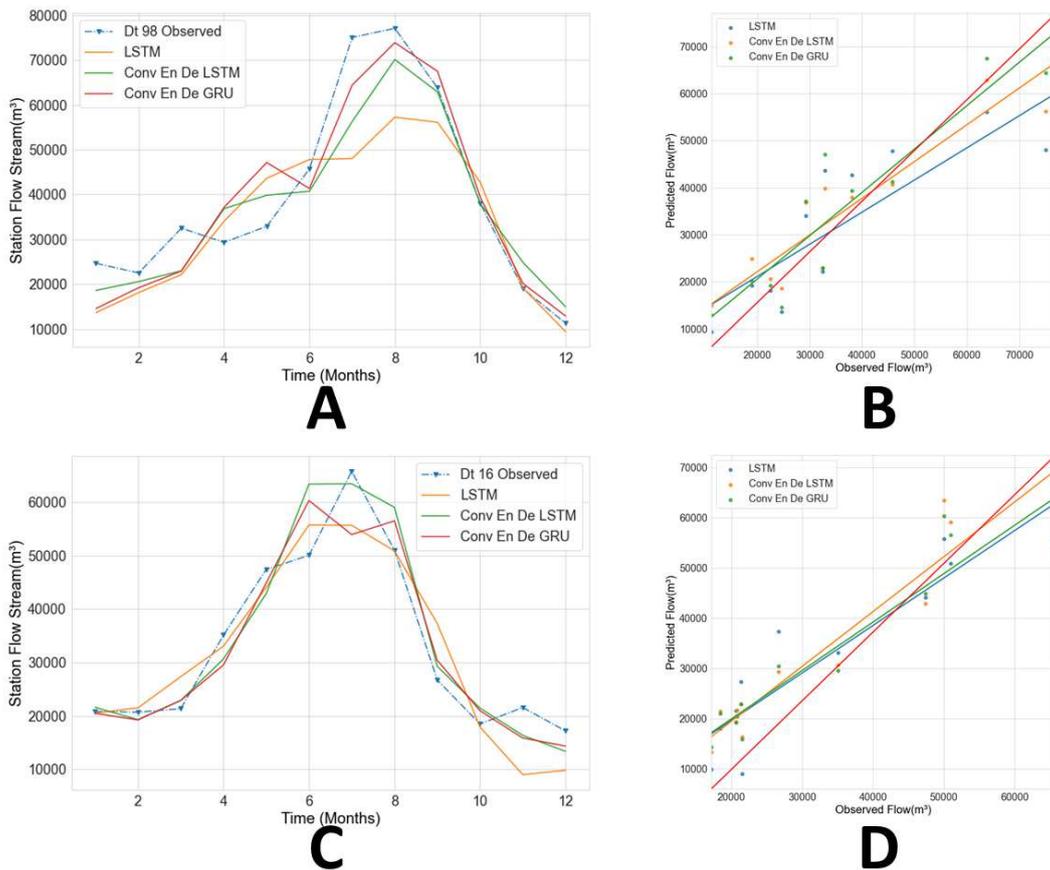


Figure 15 Comparison with different models in Datong station

similar, and only the LSTM value is slightly far away from the red line, which means that the overall performance of the three models is similar. The R2 values of the three models are around 0.8, which means that the predictions of the three models are close to the observed values; additionally, the Conv LSTM encoder-decoder GRU model has

Table 4 Comparison of criteria with different models in Datong station

		Datong		
Year	Model	Conv En De GRU	Conv En De LSTM	Stacked LSTM
1998	RMSE	7249.87	7697.48	11542.46
	WI	0.97	0.96	0.90
	LMI	0.66	0.65	0.50
	R2	0.88	0.86	0.69
2016	RMSE	5610.82	5400.94	6523.28
	WI	0.97	0.97	0.96
	LMI	0.68	0.70	0.65
	R2	0.87	0.88	0.83

the best results regarding the other three evaluation indexes. Fig.14 C shows the predictions of the three models on the monthly streamflow of the Hankou station in 2016. The predictions of the three models are similar from January to May. Stacked

LSTM's prediction during the flood season is different from those of the other models and is far from the observed value; meanwhile, the LSTM prediction differs from the observed value by 7038.91 m³, the Conv LSTM encoder-decoder LSTM prediction is similar to the observed value, and the Conv LSTM encoder-decoder LSTM prediction is different from the observed value by 0.67 m³. The flood peak occurs in June, and the observed flood peak occurs in July. Conv LSTM encoder-decoder GRU's flood peak prediction is the same as observed flood peak. Fig 14 D shows that the LSTM model has a different regression line than the other models. Tab.3 shows that Stacked LSTM has the poorest results and the worst fit to observations; the other models perform better and the predictions are nearly identical. Combining the two-year prediction data from the Hankow station, it can be concluded that the prediction of Conv LSTM encoder-decoder GRU is slightly better than that of Conv LSTM encoder-decoder LSTM, and the prediction of Stacked LSTM is worst.

Fig.15 shows a comparison of the predictions of the three models for 1998 and 2016 for the Datong station. The non-flood season shows similar results for the three models. the flood season shows that the Conv LSTM encoder-decoder GRU prediction is closest to the observed value, followed by the Conv LSTM encoder-decoder LSTM. The worst is prediction is produced by Stacked LSTM, and the slope of the regression line for Conv LSTM encoder-decoder GRU on scatter plot Fig. 15 B is closest to the red line, followed by Conv LSTM encoder-decoder LSTM. Stacked LSTM has the largest difference. In Fig. 15 C, it is found that the results of the three models are still similar during the non-flood season, with the Conv LSTM encoder-decoder LSTM results being closer to the peak flow during the flood season; Stacked LSTM's predicted streamflow has a larger gap between the peak flow and the flood. Conv encoder-decoder GRU's flood predictions fall between the results of the other models, with the predicted flood peak occurring in June, one month away from the observed flood peak, with a small difference in flood flows, but the largest difference between the predicted July flows and the observed values. The regression line for the Conv LSTM encoder-decoder LSTM are closest to the red line in Fig. 15 D. The four evaluation indicators

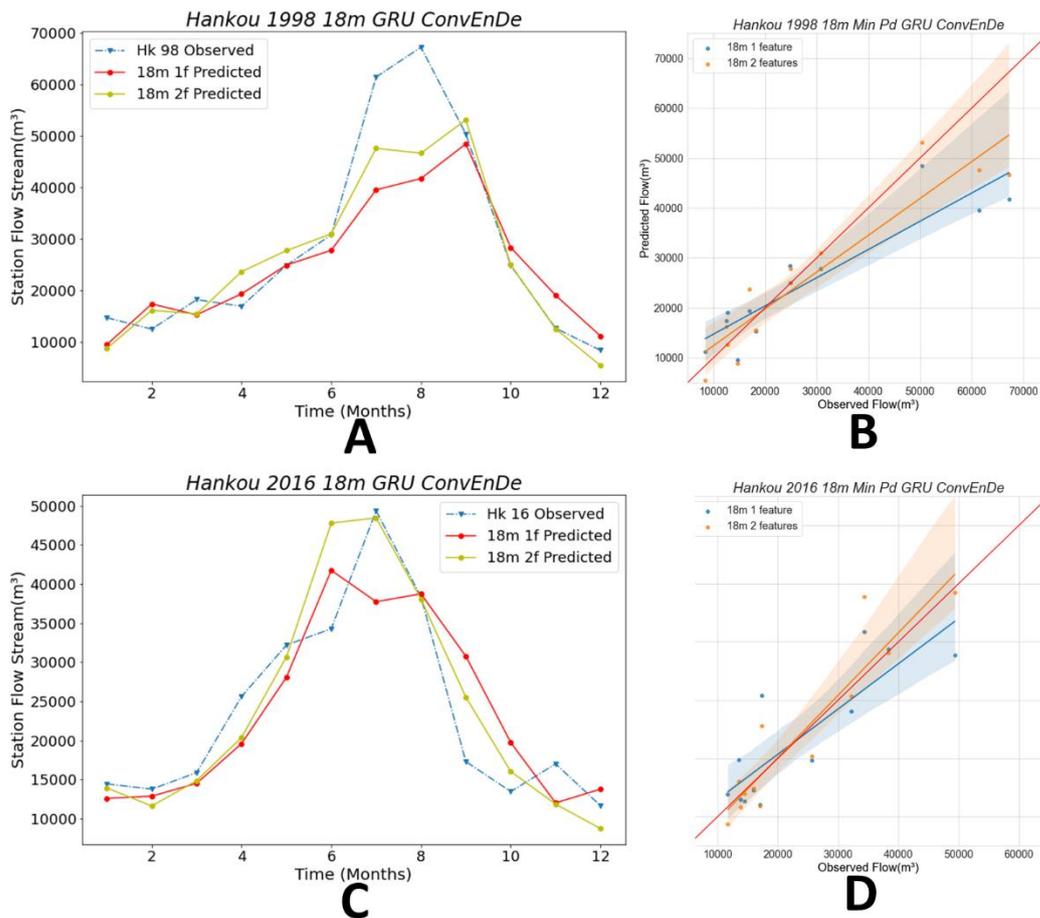


Figure 16 Comparison with different features in Hankou station

listed in Tab.4 show that the 1998 forecast results from Conv LSTM encoder-decoder GRU are the best, and those from Conv LSTM encoder-decoder LSTM are the best for the 2016 forecast.

Table 5 Comparison of criteria with different features in Hankou station

Hankou			
Year		1 feature	2 features
1998	RMSE	10272.88	7849.91
	WI	0.89	0.95
	LMI	0.58	0.67
	R2	0.71	0.83
2016	RMSE	6453.44	5228.97
	WI	0.91	0.96
	LMI	0.51	0.64
	R2	0.70	0.80

By comparing the model predictions, Conv LSTM encoder-decoder GRU has the best 2-year prediction among the three models. Conv LSTM encoder-decoder GRU has a similar prediction to Conv LSTM encoder-decoder LSTM with a 1-year prediction, Conv LSTM encoder-decoder LSTM has the best prediction with a 1-year prediction, and Stacked LSTM has the worst prediction in all cases. Conv LSTM encoder-decoder GRU performs slightly better than Conv LSTM encoder-

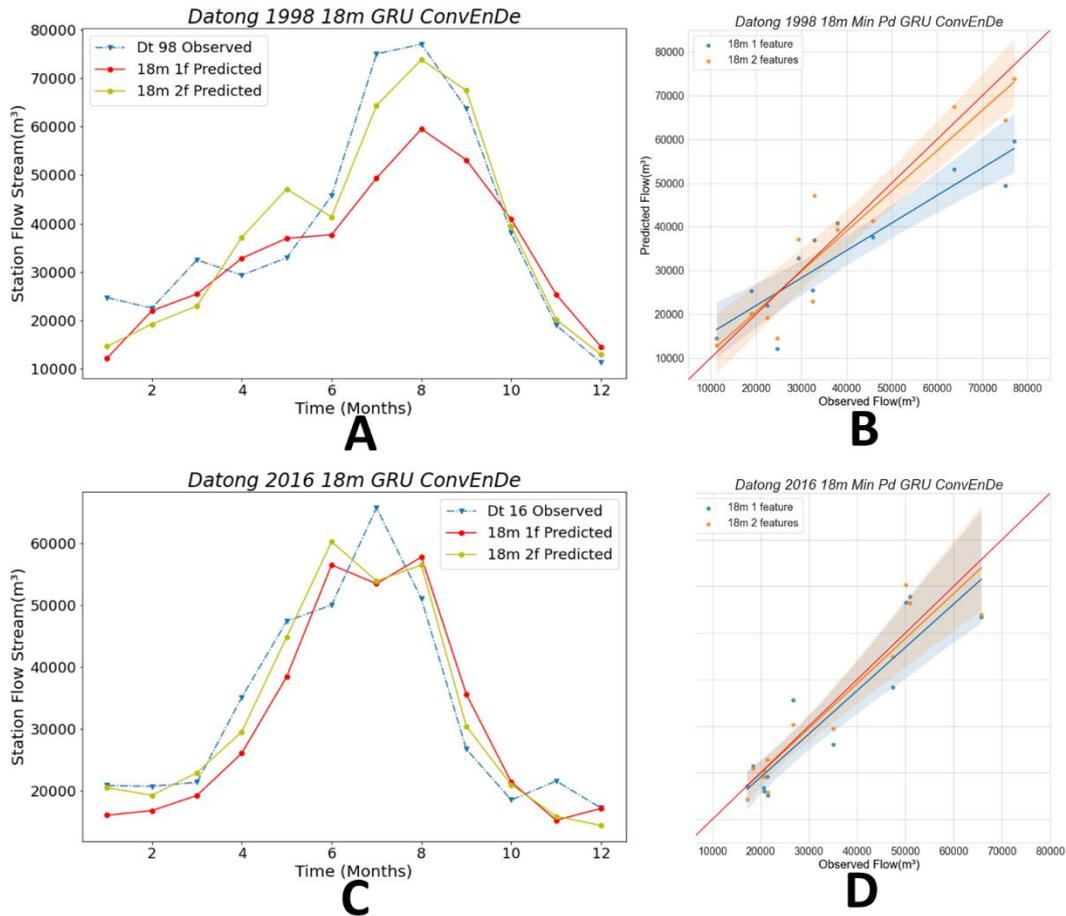


Figure 17 Comparison with different models in Datong station

decoder LSTM.

In this paper, we introduce ENSO values that are implicitly related to the streamflow data, in addition to the previous machine learning approach of using only

Table 6 Comparison of criteria with different features in Datong station

Year	Model	1 feature	2 features
1998	RMSE	10951.82	7249.87
	WI	0.90	0.97
	LMI	0.51	0.66
	R2	0.72	0.88
2016	RMSE	6905.13	5610.82
	WI	0.95	0.97
	LMI	0.57	0.68
	R2	0.81	0.87

streamflow data for training and prediction. Through this, we can enhance the training effect by increasing the data dimensions and get more accurate monthly streamflow predictions, and hopefully more accurate flood predictions. The best-performing Conv LSTM encoder-decoder GRU model is used in this next experiment, and the best-performing 18m-min-pd data partitioning method is used to compare the difference in prediction results between the ENSO + streamflow dataset and the

streamflow dataset. The experimental results are presented in Figs. 16 and 17 and Tab. 5,6, where A and C plot the line graphs of monthly flows, and B and D present scatter plots, regression lines, and confidence intervals.

Fig.16 A plots the 1998 prediction results from Hankou station. It can be seen that the results from the 2features data are closer to the observed values during the flood months, while the October–November prediction is almost equal to the observed values. From Fig. 16 B, we also find that the regression line of the 2features results is closer to the red line, but the confidence intervals are similar in both cases. By comparing the four evaluation indicators of Tab.5, it can be found that the RMSE of 2features is reduced by 2422.97 m³, WI is increased by 6% to 0.95, LMI is increased by 16% to 0.67, and R2 is increased by 17% to 0.83. In Fig. 16 C, the prediction results of Hankow station in 2016 are presented. The model using 2features data accurately predicts the timing and streamflow of flood peaks, while the 1features model flood peak prediction differs by 1 month from the observed value and gives a value that differs significantly from the observed value in the month in which the flood peak occurs. In addition, in non-flood seasons, the 2features predictions are closer to the observations. In Fig.16 D it can be seen that the confidence intervals are about the same, but the regression line for 2features is very close to the red line; meanwhile, the regression line for 1features differs more from the red line. Moreover, all the four evaluation indices are greatly improved (Tab.5): the RMSE is reduced by 1224.47 m³, WI is increased by 5% to 0.96, LMI is increased by 25% to 0.64, and R2 is increased by 14% to 0.80. Based on these results, we can obtain the monthly streamflow rate of the model in Hankou in 1998 after adding ENSO data. Predictions have significantly improved, with better results not only in non-flood months but also in flood months.

Fig.17 A plots the 1998 prediction results of the Datong station. The 2features data make the prediction much more accurate than 1 feature data, and the predictions for July, August, and September are very close to the observations; meanwhile, the 1features data make the predictions significantly different from the observations. In October–December, the prediction of 2features is almost equal to the observed value. Fig.17B shows that the regression line for the 2features is very close to the red line. On

Tab.6, it is clear that the 2features evaluation index has improved significantly, with the RMSE shrinking by 3701.95 m³, WI increasing by 8% to 0.97, LMI increasing by 29% to 0.66, and R2 increasing by 22% to 0.88. In Fig. 17C, it is clear that the two eigenvalues yield similar prediction results. The 2features result flood peak occurs in June, the 1features result flood peak occurs in August, and the observed flood peak occurs in July; meanwhile, the two predictions for July are almost identical and differ significantly from the observations. The comparison of the predicted flood peaks shows that the 2features results are closer to the observed maximum flows. In non-flood

Table 7 The criteria of Hankou station

		Hankou Station								
		6m min pd		12m min pd		18m min pd		24m min pd		
Model	Year	1feature	2features	1feature	2features	1feature	2features	1feature	2features	
Conv Encoder	1998	RMSE	11885.94	10224.25	11069.12	8536.23	10272.88	7849.91	10528.94	8568.73
		WI	0.84	0.89	0.87	0.94	0.89	0.95	0.89	0.94
		LMI	0.54	0.54	0.55	0.58	0.58	0.67	0.54	0.58
		R2	0.62	0.72	0.67	0.80	0.71	0.83	0.70	0.80
GRU Decoder	2016	RMSE	6840.26	5974.99	6449.84	6157.50	6453.44	5228.97	6227.46	5196.74
		WI	0.90	0.94	0.91	0.94	0.91	0.96	0.93	0.94
		LMI	0.52	0.53	0.53	0.54	0.51	0.64	0.53	0.61
		R2	0.66	0.74	0.70	0.72	0.70	0.80	0.72	0.80
Conv Encoder	1998	RMSE	11873.32	10465.17	10912.88	10003.99	10115.54	8675.98	9645.38	8623.09
		WI	0.84	0.88	0.87	0.90	0.89	0.94	0.91	0.93
		LMI	0.54	0.56	0.56	0.59	0.57	0.61	0.59	0.62
		R2	0.62	0.70	0.68	0.73	0.72	0.80	0.75	0.80
LSTM Decoder	2016	RMSE	6722.49	5347.60	6448.34	5859.16	6131.52	5216.75	6211.28	5484.72
		WI	0.90	0.95	0.92	0.94	0.92	0.95	0.93	0.94
		LMI	0.52	0.59	0.53	0.57	0.54	0.65	0.50	0.55
		R2	0.67	0.79	0.70	0.75	0.72	0.80	0.72	0.78
Stacked LSTM	1998	RMSE	11343.89	9967.40	11541.24	9417.74	11820.25	8757.40	11128.30	9293.75
		WI	0.86	0.90	0.85	0.91	0.85	0.93	0.87	0.92
		LMI	0.52	0.54	0.52	0.57	0.54	0.58	0.53	0.53
		R2	0.65	0.73	0.64	0.76	0.62	0.79	0.66	0.77
Stacked LSTM	2016	RMSE	8354.20	5428.27	7404.95	5720.43	7745.91	6184.42	6997.11	6311.20
		WI	0.83	0.94	0.87	0.94	0.85	0.93	0.89	0.93
		LMI	0.38	0.62	0.51	0.60	0.38	0.51	0.44	0.51
		R2	0.49	0.78	0.60	0.76	0.56	0.72	0.64	0.71

months, most of the 2features results are more accurate. The regression lines for the results on Fig. 17 D for both data are very close, but the regression line for the predicted results for 2features is slightly more accurate. A slight improvement in the evaluation metrics for 2features over 1feature can also be seen through Tab.6. Ultimately, these results illustrate that the accuracy of the prediction is improved by adding ENSO data, with a significant improvement in the 1998 prediction and a small improvement in the 2016 prediction.

In the above comparison, the addition of ENSO data to the 18m-min-pd division

Table 8 The criteria of Datong station

Datong Station										
Model	Year	6m min pd		12m min pd		18m min pd		24m min pd		
		1feature	2features	1feature	2features	1feature	2features	1feature	2features	
Conv Encoder	1998	RMSE	14132.73	12074.85	12393.00	10430.62	10951.82	7249.87	10869.14	9954.34
		WI	0.82	0.88	0.86	0.92	0.90	0.97	0.91	0.94
		LMI	0.43	0.52	0.46	0.52	0.51	0.66	0.50	0.52
		R2	0.54	0.67	0.65	0.75	0.72	0.88	0.73	0.77
GRU Decoder	2016	RMSE	8744.39	6618.38	6902.22	5957.79	6905.13	5610.82	6730.49	5462.29
		WI	0.91	0.95	0.95	0.97	0.95	0.97	0.96	0.97
		LMI	0.54	0.62	0.63	0.63	0.57	0.68	0.59	0.66
		R2	0.69	0.82	0.81	0.86	0.81	0.87	0.82	0.88
Conv Encoder	1998	RMSE	13821.62	11995.99	11995.13	10219.74	10681.81	7697.48	10685.21	9603.82
		WI	0.83	0.88	0.87	0.92	0.91	0.96	0.91	0.94
		LMI	0.45	0.50	0.46	0.52	0.50	0.65	0.52	0.55
		R2	0.56	0.67	0.67	0.76	0.74	0.86	0.74	0.79
LSTM Decoder	2016	RMSE	8738.14	6837.50	7116.51	5752.76	7052.87	5400.94	6741.70	5134.78
		WI	0.91	0.95	0.94	0.97	0.95	0.97	0.96	0.97
		LMI	0.52	0.60	0.60	0.66	0.58	0.70	0.57	0.69
		R2	0.69	0.81	0.80	0.87	0.80	0.88	0.82	0.89
Stacked LSTM	1998	RMSE	15392.32	10913.63	13839.18	11219.90	13645.69	11542.46	13986.39	11354.88
		WI	0.80	0.90	0.83	0.90	0.85	0.90	0.84	0.91
		LMI	0.37	0.55	0.39	0.50	0.39	0.50	0.39	0.52
		R2	0.46	0.73	0.56	0.71	0.57	0.69	0.55	0.70
LSTM Decoder	2016	RMSE	10168.99	5996.88	9671.06	7012.46	10642.35	6523.28	8847.33	8042.89
		WI	0.87	0.96	0.88	0.95	0.85	0.96	0.90	0.92
		LMI	0.44	0.65	0.47	0.58	0.37	0.65	0.48	0.50
		R2	0.58	0.85	0.62	0.80	0.54	0.83	0.68	0.74

in the Conv LSTM encoder-decoder GRU significantly improves the prediction accuracy.

Also, in this paper, in order to investigate the effect of Enso values on flow prediction, the predicted monthly flows for both Hankow and Datong stations in 1998 and 2016 are compared under three models and four data divisions, and these results are shown on Tabs.7-10. Tabs.7,8 show the comparison of the four-evaluation metrics for the prediction results, and Tabs.9,10 show the comparison of the maximum flows for the prediction results. tabs.7,8 show that the prediction results are all improved to varying degrees by the addition of Enso values, and when the model originally had poor predictions, adding data feature values will result in a greater improvement in the accuracy of the predictions, such as Stacked LSTM for Hankou station in 2016 using 6m-min-pd data and the model's predictions for Datong station in 1998 and 2016. When the model's original predictions are more accurate, the effect of increasing the value of the data features on the prediction accuracy is diminished, a feature that can be observed in the prediction results of Conv encoder-decoder LSTM and Conv encoder-decoder GRU. In the case of the maximum flow prediction, the maximum flows predicted by each model were closer to the observed maximum flows with the addition of Enso data, and the months in which the maximum flows were predicted to occur were more accurate.

Table 9 The peak flow prediction of Hankou

Model	Station	Hankou							
	Year	1998 Max Flow(M ³)/Month				2016 Max Flow(M ³)/Month			
	Min Pd	6	12	18	24	6	12	18	24
	Observed	67219.00/8				49355.00/7			
Stacked LSTM	1 feature	47949.05/9	42226.30/8	38507.58/8	44690.96/8	34898.43/8	36758.89/8	31422.67/8	33545.99/7
	2 features	49315.43/9	47666.97/8	47669.21/8	48944.84/7	40608.11/7	40982.64/7	42316.09/7	42020.28/7
Conv En De LSTM	1 feature	42814.24/9	43659.18/9	47798.03/9	45552.51/8	36811.79/8	40672.40/6	41455.50/6	41421.19/8
	2 features	46918.94/8	45675.12/8	57500.74/9	47941.18/7	45235.79/7	44817.63/7	48776.30/6	43024.21/6
Conv En De GRU	1 feature	42117.32/9	42880.25/9	48480.34/9	47258.80/9	36257.22/8	38662.68/6	41743.10/6	40968.08/8
	2 features	47496.47/8	52758.74/9	53141.31/9	55788.13/9	44141.22/7	48966.08/7	48445.01/7	39724.06/6

Table 10 The peak flow prediction of Datong

Model	Station	Datong							
	Year	1998 Max Flow(M ³)/Month				2016 Max Flow(M ³)/Month			
	Min Pd	6	12	18	24	6	12	18	24
	Observed	77065.00/8				65758.00/7			
Stacked LSTM	1 feature	57211.82/9	53656.70/8	51463.75/8	56766.26/8	51303.70/8	48951.63/8	42252.02/6	48679.22/8
	2 features	64647.34/8	63604.23/8	57249.11/8	58708.07/8	58042.62/6	60753.18/8	55705.89/6	53397.02/6
Conv En De LSTM	1 feature	55976.55/9	57560.53/8	63218.27/8	62804.75/8	53913.72/8	55936.09/8	60353.25/8	58593.31/8
	2 features	60367.56/8	65174.24/8	70113.20/8	69740.06/9	58457.18/8	63363.18/7	63425.56/7	60643.40/7
Conv En De GRU	1 feature	54054.31/9	58217.41/9	59528.55/8	59591.22/8	52622.26/8	53630.74/7	57775.40/8	63374.66/6
	2 features	59583.14/8	67916.24/8	73878.27/8	68581.81/8	55687.05/7	60457.82/7	60273.86/6	59028.74/6

CONCLUSION

In this paper, three network structures, stacked LSTM, Conv LSTM encoder-decoder LSTM, and Conv LSTM encoder-decoder GRU, are trained using two eigenvalues datasets, ENSO and monthly streamflow, to predict the monthly streamflow of the Yangtze River in Hankou and Datong stations in 1998 and 2016. The best results were obtained with the Conv LSTM encoder-decoder GRU: the R^2 exceeded 0.80, the RMSE was less than 8000 m^3 , the WI was over 0.95, and the LMI over 0.65, indicating more accurate flood prediction.

This paper compares the prediction results of the three network structures for two flood years, 1998 and 2016, and shows that the prediction accuracy of all three network structures is improved by adding ENSO data. The improvement of Conv LSTM encoder-decoder LSTM and Conv LSTM encoder-decoder GRU is smaller than Stacked LSTM. The final prediction results achieved a maximum evaluation index of $R^2=0.88$, $LMI=0.66$, $WI=0.97$ and a minimum prediction difference of 389 m^3 for the flood peak. It can be seen that the enhancement of the dataset results in different magnitudes of improvement in the prediction for each of the three network structures; this is because Conv LSTM encoder-decoder LSTM and Conv LSTM encoder-decoder GRU already have strong feature extraction capabilities for time series, while Stacked LSTM has relatively weak feature extraction capabilities. The added deep connection between ENSO data and flow data enables the network structure to extract more information, thus compensating for the time series feature extraction deficiency to some extent and greatly improving the accuracy of prediction.

We found that the neural network model predicts the middle and lower reaches of the Yangtze River represented by the flows at Hankou and Datong stations. By adding ENSO data to the streamflow data, the prediction ability of each model on different parameters is greatly improved, which reveals that there is an implicit relationship between ENSO and flow data that can be learned by the neural network. At the same time, the discrepancy between the 1998 and 2016 forecasts shows that there is a certain discrepancy between the 1998 flood season streamflow forecast trend and the actual

streamflow trend; the 2016 flood season streamflow forecast trend is almost the same as the actual trend because the streamflow of the mainstream of the Yangtze River started to be disturbed by human interference after the construction of the Three Gorges Dam and other water conservation projects. The difference between the streamflow of the Yangtze River in the last century and the streamflow changes in the current century is due to this influence, which leads to the fact that the prediction model cannot learn similar unnatural river streamflow changes simply by adding ENSO data. We note that the number of streamflow data samples collected is only about 700, which is small for machine learning. Augmenting the model with ENSO data can be seen as augmenting the training set and compensating for this lack of data.

The variation in streamflow volume in the Yangtze River is not only related to the ENSO data but also many other variables; thus, the data can be enhanced by adding more variables, which would make the prediction more accurate. Different regions in the Yangtze River basin have different relationships with climate change, and different locations in the Yangtze River have different relationships with upstream streamflow; thus, more sites could be used for joint prediction. With the rapid development of deep learning, there may be many more powerful models that would enable us to obtain better prediction results.

Funding

This work was supported by the Key-Area Research and Development Program of Guangdong Province (Grant No. 2020B1111020005), National Natural Science Foundation of China (Grant No. U2006210) and Shenzhen Fundamental Research Program (Grant No. JCYJ20200109110220482).

References

- [1] Xiao M. “Summary of flood disasters in China in 2017” , China Flood & Drought Management, Vol.28,No.8,pp.60-66.
- [2] Huang F&Xia Z&Zhang N, et al. “Flow-Complexity Analysis of the Upper Reaches of the Yangtze River, China” , Journal of Hydrologic Engineering, Vol.16,No.11,pp.914-919.
- [3] Xiang Z&Demir I. “Distributed long-term hourly streamflow predictions using deep learning – A case study for State of Iowa” , Environmental Modelling & Software, Vol.131,pp.104761.

- [4] Xiang Z&Yan J&Demir I. “A Rainfall - Runoff Model With LSTM - Based Sequence - to - Sequence Learning” , Water Resources Research, Vol.56,No.1.
- [5] Noori N&Kalin L. “Coupling SWAT and ANN models for enhanced daily streamflow prediction”, Journal of Hydrology, Vol.533,pp.141-151.
- [6] Adnan R M&Liang Z&Trajkovic S, et al. “Daily streamflow prediction using optimally pruned extreme learning machine” , Journal of Hydrology, Vol.577,pp.123981.
- [7] Cheng M&Fang F&Kinouchi T, et al. “Long lead-time daily and monthly streamflow forecasting using machine learning methods” , Journal of Hydrology, Vol.590,pp.125376.
- [8] Kashid S S&Ghosh S&Maity R. “Streamflow prediction using multi-site rainfall obtained from hydroclimatic teleconnection” , Journal of Hydrology, Vol.395,No.1,pp.23-38.
- [9] Keteklahijani V K&Alimohammadi S&Fattahi E. “Predicting changes in monthly streamflow to Karaj dam reservoir, Iran, in climate change condition and assessing its uncertainty” , Ain Shams Engineering Journal, Vol.10,No.4,pp.669-679.
- [10] Ni Q&Wang L&Ye R, et al. “Evolutionary Modeling for Streamflow Forecasting with Minimal Datasets: A Case Study in the West Malian River, China” , Environmental Engineering Science, Vol.27,No.5,pp.377-385.
- [11] BEVEN K J&KIRKBY M J. “A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant” , Hydrological Sciences Bulletin, Vol.24,No.1,pp.43-69.
- [12] Ren-Jun Z. “The Xinanjiang model applied in China” , Journal of Hydrology, Vol.135,No.1,pp.371-381.
- [13] Box G. Box and Jenkins: Time Series Analysis, Forecasting and Control[M]//MILLS T C. A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century. London: Palgrave Macmillan UK, 2013:161-215.
- [14] Valipour M. “Long-term runoff study using SARIMA and ARIMA models in the United States” , Meteorological Applications, Vol.22,No.3,pp.592-598.
- [15] Valipour M&Banihabib M E&Behbahani S M R. “Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir” , Journal of Hydrology, Vol.476,pp.433-441.
- [16] M. Valipour&M. E. Banihabib&S. M. R. Behbahani. “Monthly Inflow Forecasting using Autoregressive Artificial Neural Network” , Journal of Applied Sciences, Vol.12,No.20,pp.2139-2147.
- [17] Haykin S. “Neural networks: A Comprehensive Foundation” .
- [18] J. K&R. E. Particle swarm optimization: Proceedings of ICNN'95 - International Conference on Neural Networks, 1995[C].1995
27 Nov.-1 Dec. 1995.
- [19] L. A Z. “Fuzzy logic” , Computer, Vol.21,No.4,pp.83-93.
- [20] Schwefel H. Numerical Optimization of Computer Models[M]. John Wiley & Sons, Inc., 1981.
- [21] Grossmann A&Morlet J. “Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape” , SIAM Journal on Mathematical Analysis, Vol.15,No.4,pp.723-736.
- [22] Wu Y&Schuster M&Chen Z, et al. “Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation” .
- [23] Cho K&Van Merriënboer B&Gulcehre C, et al. “Learning phrase representations using RNN encoder-decoder for statistical machine translation” , arXiv preprint arXiv:1406.1078.

- [24] Xu H&Gao Y&Yu F, et al. End-to-end learning of driving models from large-scale video datasets, 2017[C].2017.
- [25] Xu H&Gao Y&Yu F, et al. “End-to-end Learning of Driving Models from Large-scale Video Datasets” .
- [26] Lai G&Chang W&Yang Y, et al. Modeling long-and short-term temporal patterns with deep neural networks: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018[C].
- [27] Nalley D&Adamowski J&Biswas A, et al. “A multiscale and multivariate analysis of precipitation and streamflow variability in relation to ENSO, NAO and PDO ” , Journal of Hydrology, Vol.574,pp.288-307.
- [28] Wei W&Chang Y&Dai Z. “Streamflow changes of the Changjiang (Yangtze) River in the recent 60 years: Impacts of the East Asian summer monsoon, ENSO, and human activities” , Quaternary International, Vol.336,pp.98-107.
- [29] Räsänen T A&Kummu M. “Spatiotemporal influences of ENSO on precipitation and flood pulse in the Mekong River Basin” , Journal of Hydrology, Vol.476,pp.154-168.
- [30] Zhang Q&Xu C&Jiang T, et al. “Possible influence of ENSO on annual maximum streamflow of the Yangtze River, China” , Journal of Hydrology, Vol.333,No.2,pp.265-274.
- [31] Tong J&Qiang Z&Deming Z, et al. “Yangtze floods and droughts (China) and teleconnections with ENSO activities (1470 - 2003)” , Quaternary International, Vol.144,No.1,pp.29-37.
- [32] Sharma S&Srivastava P&Fang X, et al. “Performance comparison of Adoptive Neuro Fuzzy Inference System (ANFIS) with Loading Simulation Program C++ (LSPC) model for streamflow simulation in El Niño Southern Oscillation (ENSO)-affected watershed” , Expert Systems with Applications, Vol.42,No.4,pp.2213-2223.
- [33] Danandeh Mehr A&Kahya E&Şahin A, et al. “Successive-station monthly streamflow prediction using different artificial neural network algorithms” , International Journal of Environmental Science and Technology, Vol.12,No.7,pp.2191-2200.
- [34] Hochreiter S&Schmidhuber J. “ Long Short-Term Memory ” , Neural Computation, Vol.9,No.8,pp.1735-1780.
- [35] Hochreiter S. “The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions” , International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol.06,No.02,pp.107-116.
- [36] Cho K&van Merriënboer B&Gulcehre C, et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation” .
- [37] Chung J&Gulcehre C&Cho K, et al. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling” .
- [38] Shi X&Chen Z&Wang H, et al. “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting” .
- [39] Sutskever I&Vinyals O&Le Q V. “Sequence to Sequence Learning with Neural Networks” , Advances in Neural Information Processing Systems,pp.3104-3112.
- [40] Cheng D&Minquan F. “Analysis of The Relationship Between Flood Distribution And Water Level-Discharge Characteristics In The Yangtze River”, Journal of Xi'an University of Technology.
- [41] Qingzhou W. “A Review of Flood Disasters in China in the 20th Century” , Journal Of Catastrophology, No.02, pp.64-71.
- [42] Jun W. “Characteristics and enlightenment of 2016 Changjiang River flood” , Yangtze River,

Vol.48,No.4,pp.54-57, 65.

- [43] Chai T&Draxler R R. “Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature” , Geoscientific Model Development, Vol.7,No.3,pp.1247-1250.
- [44] Carpenter R G. “Principles and procedures of statistics, with special reference to the biological sciences” , The Eugenics Review, Vol.52,No.3,pp.172-173.
- [45] Willmott C J&Robeson S M&Matsuura K. “A refined index of model performance” , International Journal of Climatology, Vol.32,No.13,pp.2088-2094.
- [46] Legates D R&McCabe Jr. G J. “Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation” , Water Resources Research, Vol.35,No.1,pp.233-241.

Figures

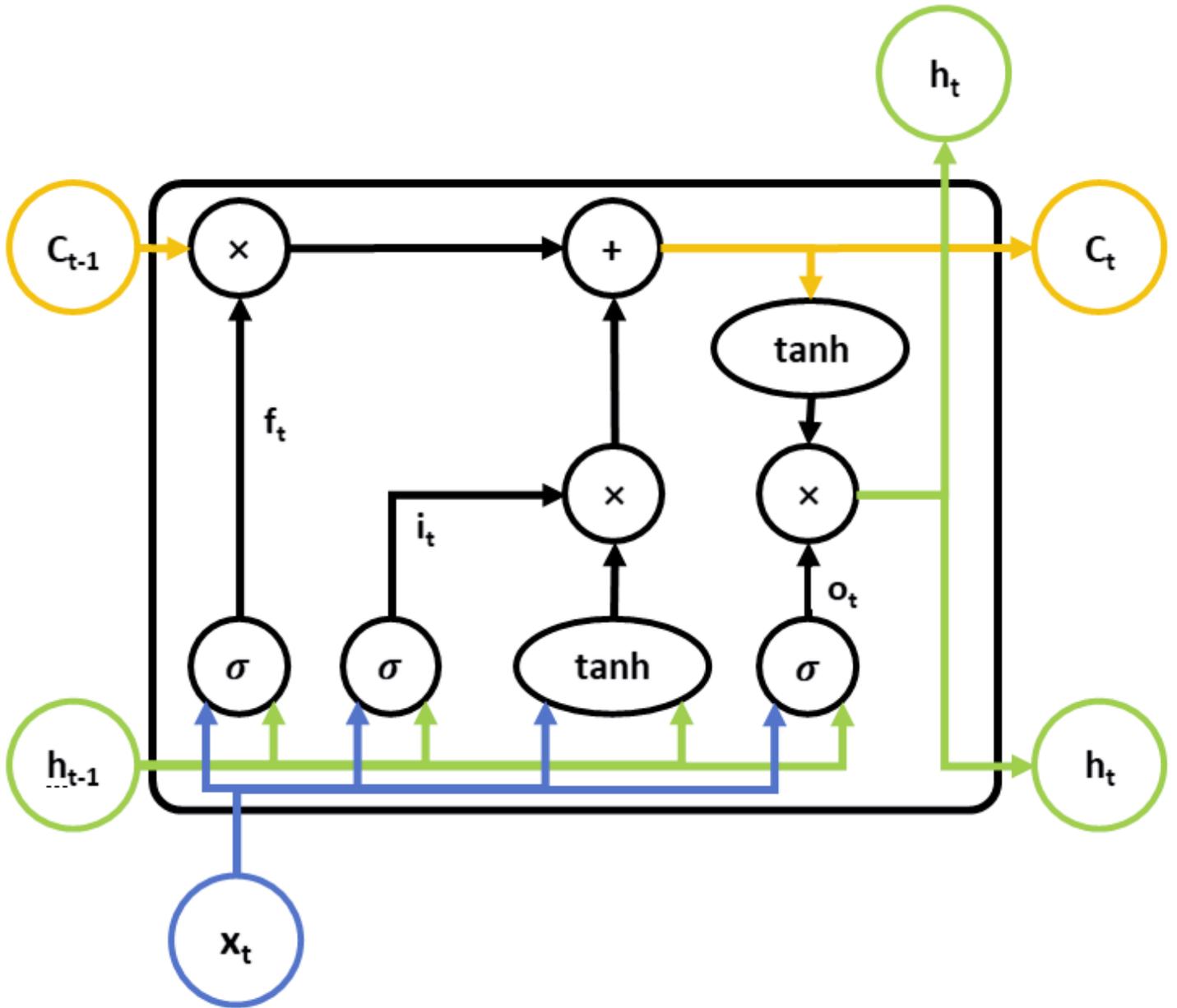


Figure 1

The networks of LSTM memory unitnet

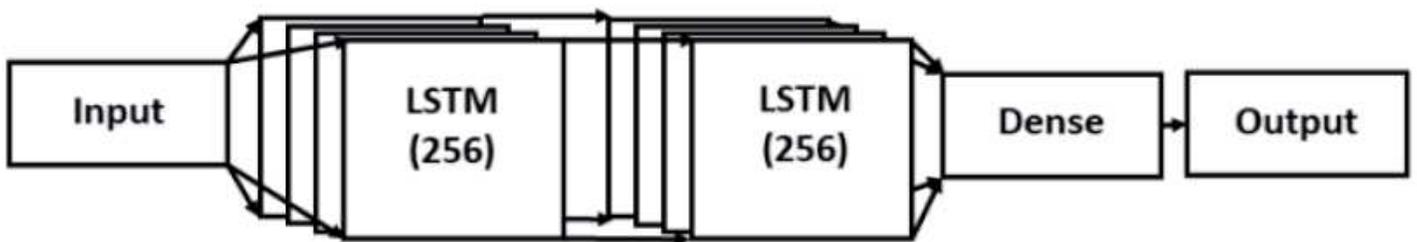


Figure 2

Stacked LSTM Structure

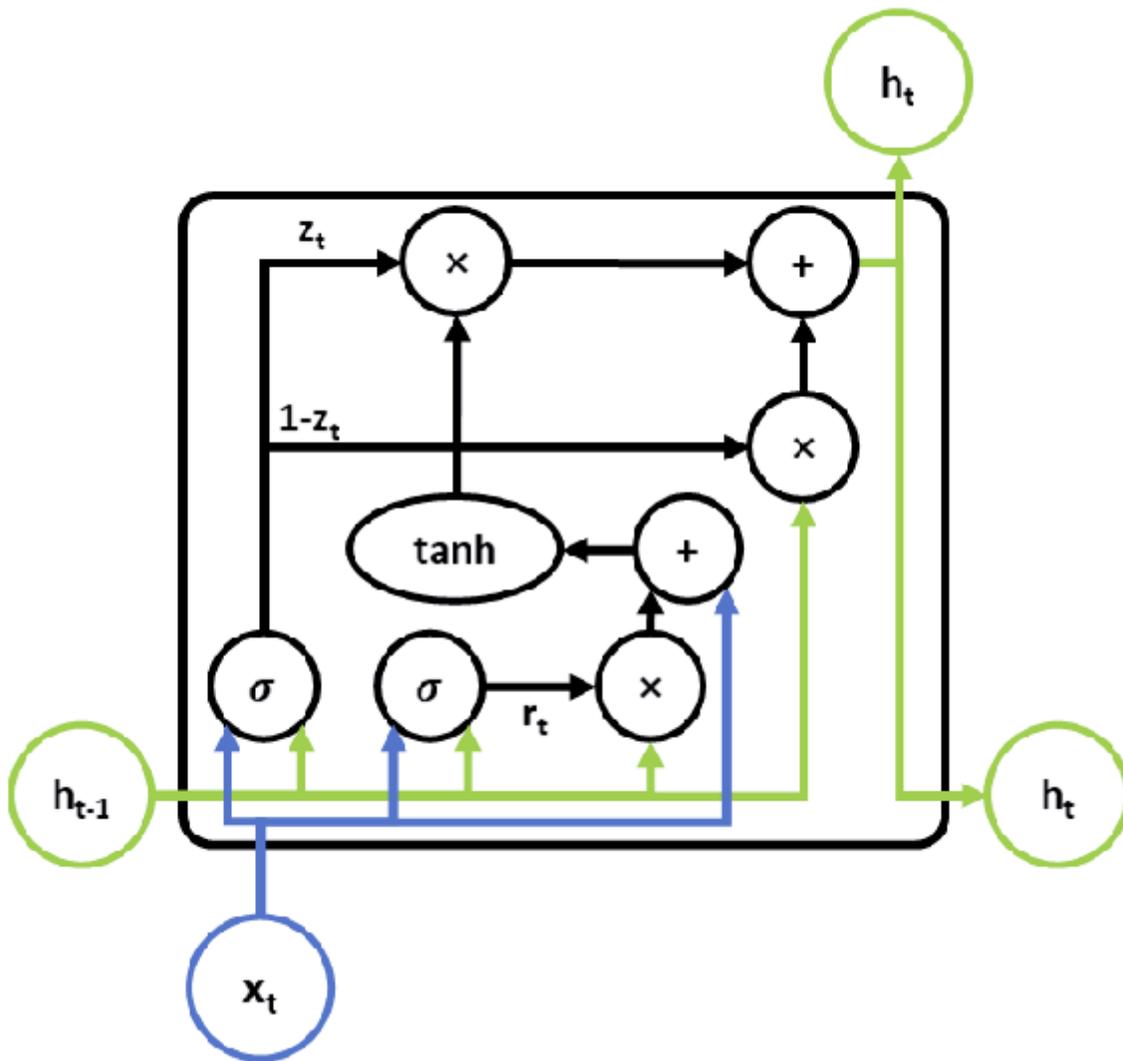


Figure 3

The networks of GRU memory unit

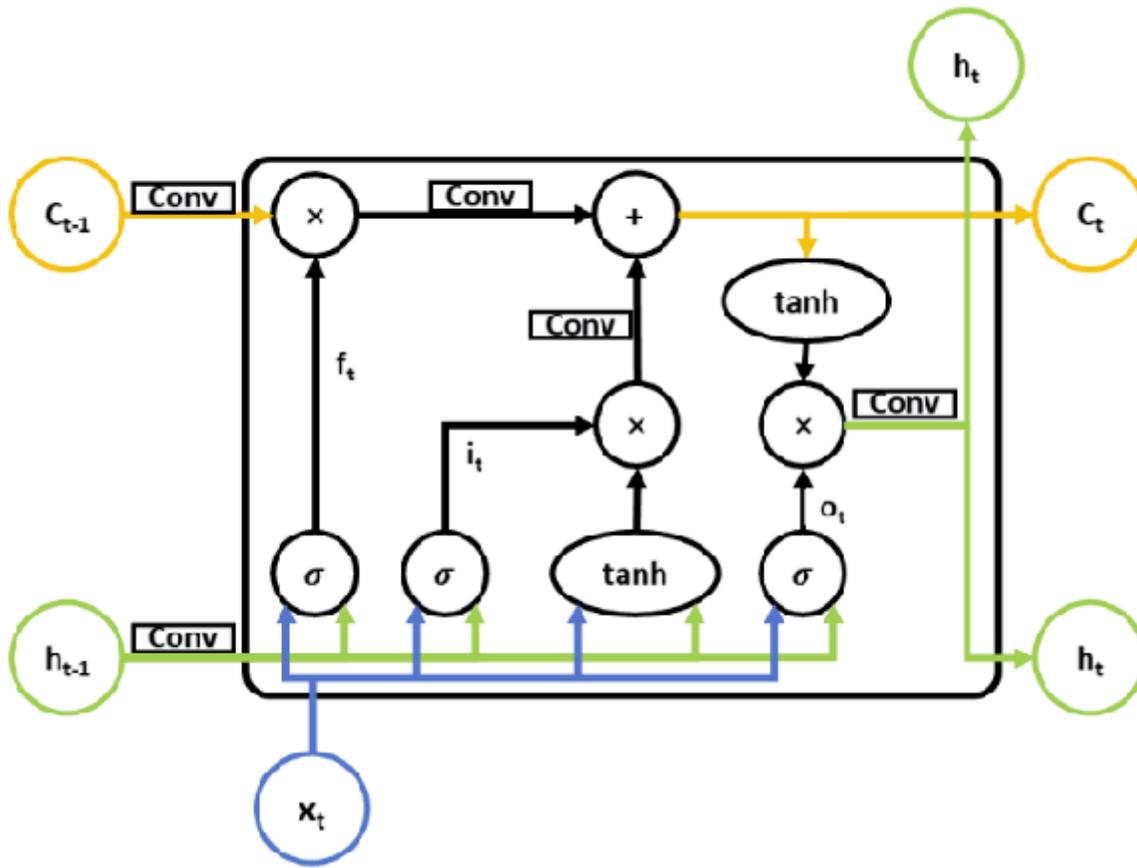


Figure 4

The networks of Conv LSTM memory unit

Image not available with this version

Figure 5

The encoder-decoder structure is shown in Fig. 5.

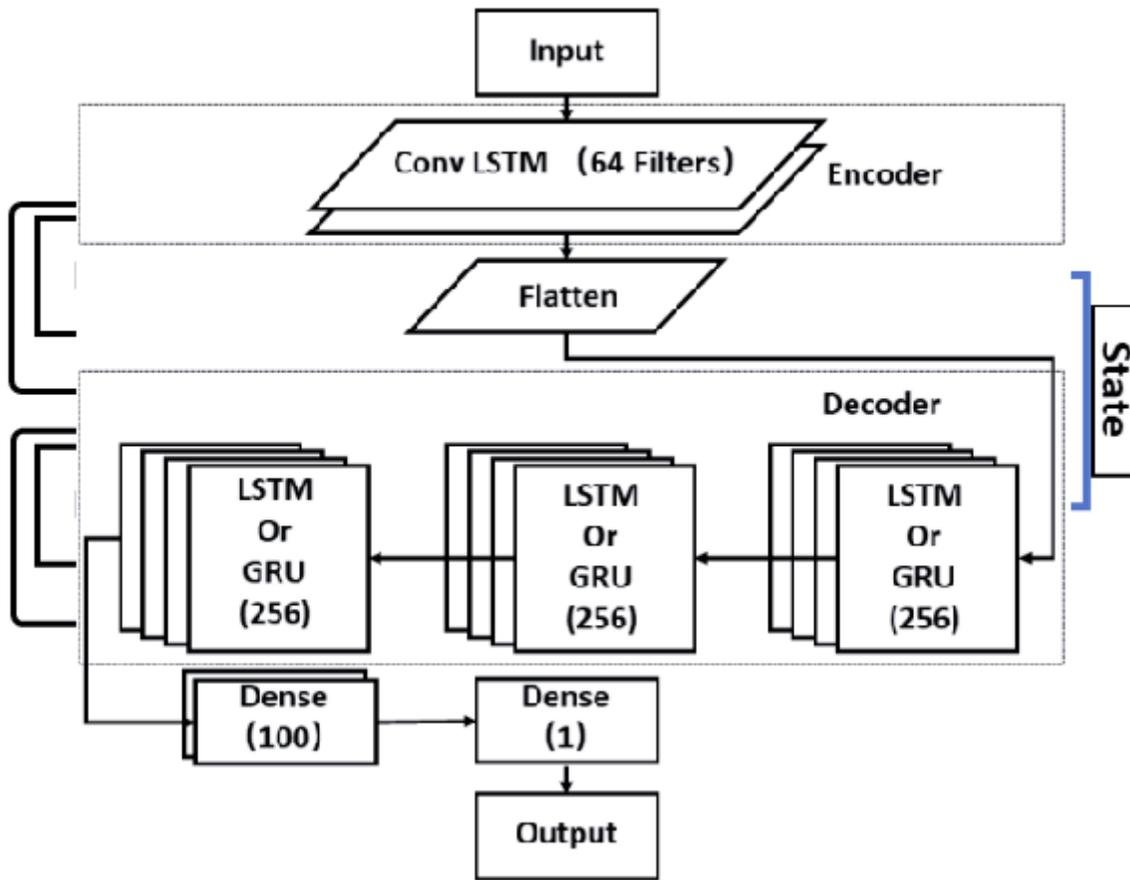


Figure 6

Conv LSTM Encoder decoder

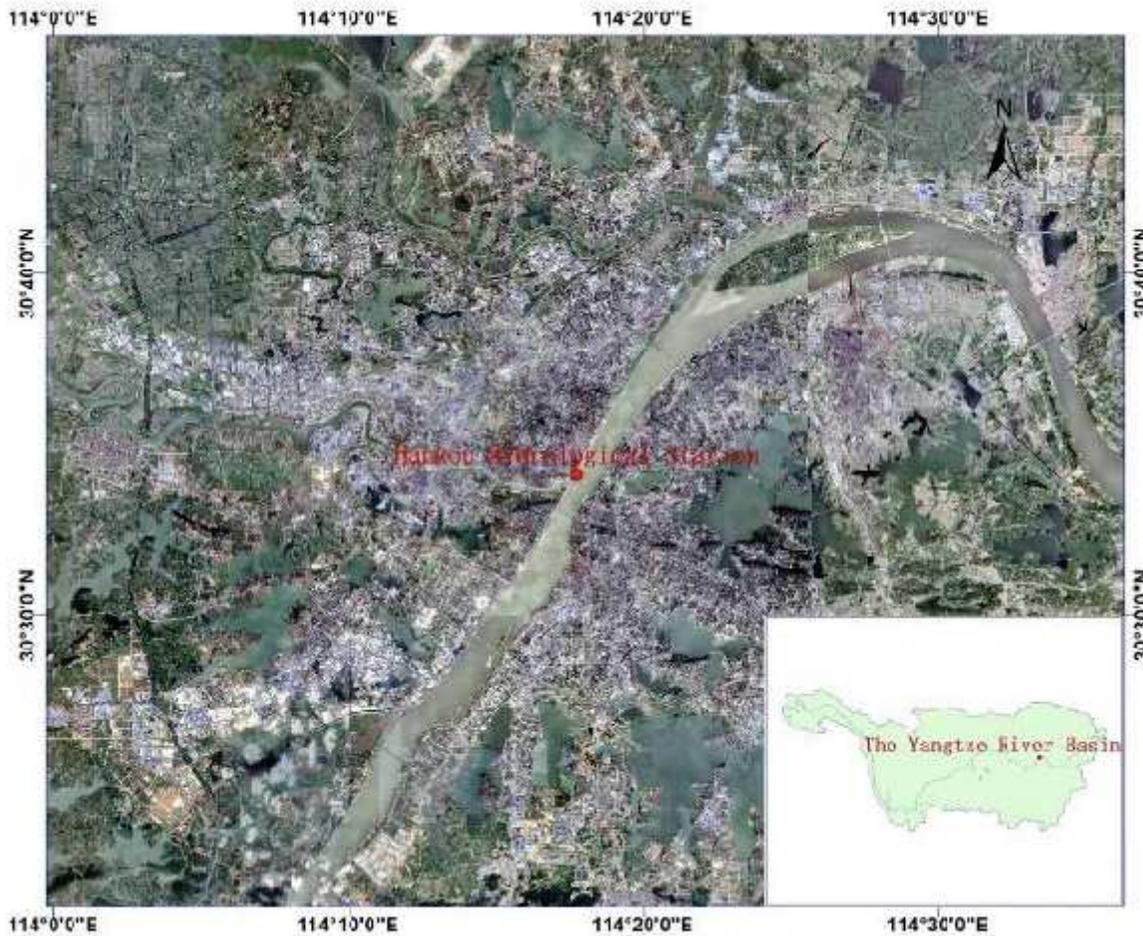


Figure 7

The location of the Yangtze River basin and Hankou Hydrological Station in Wuhan. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

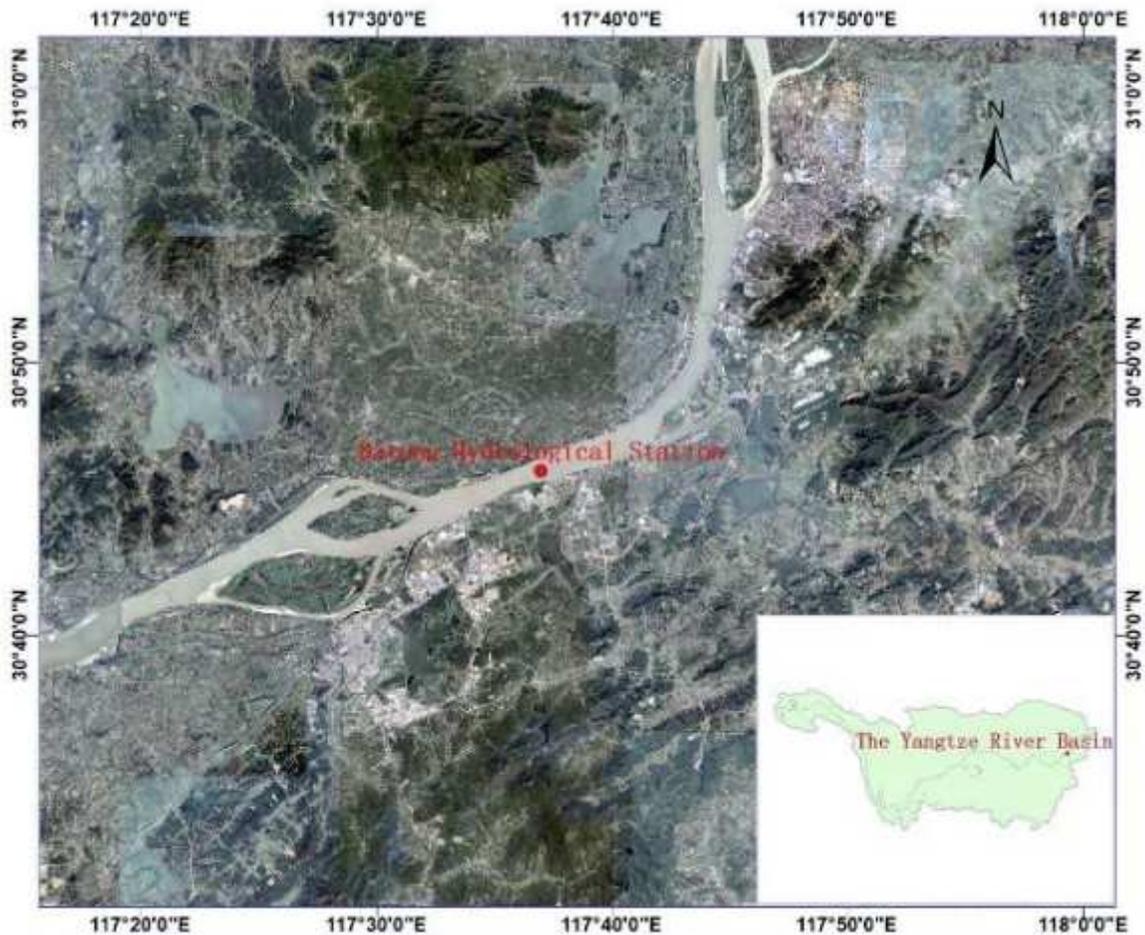


Figure 8

The loc location of the Yangtze River basin and Datong Hydrological Station in Chi Zhou. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

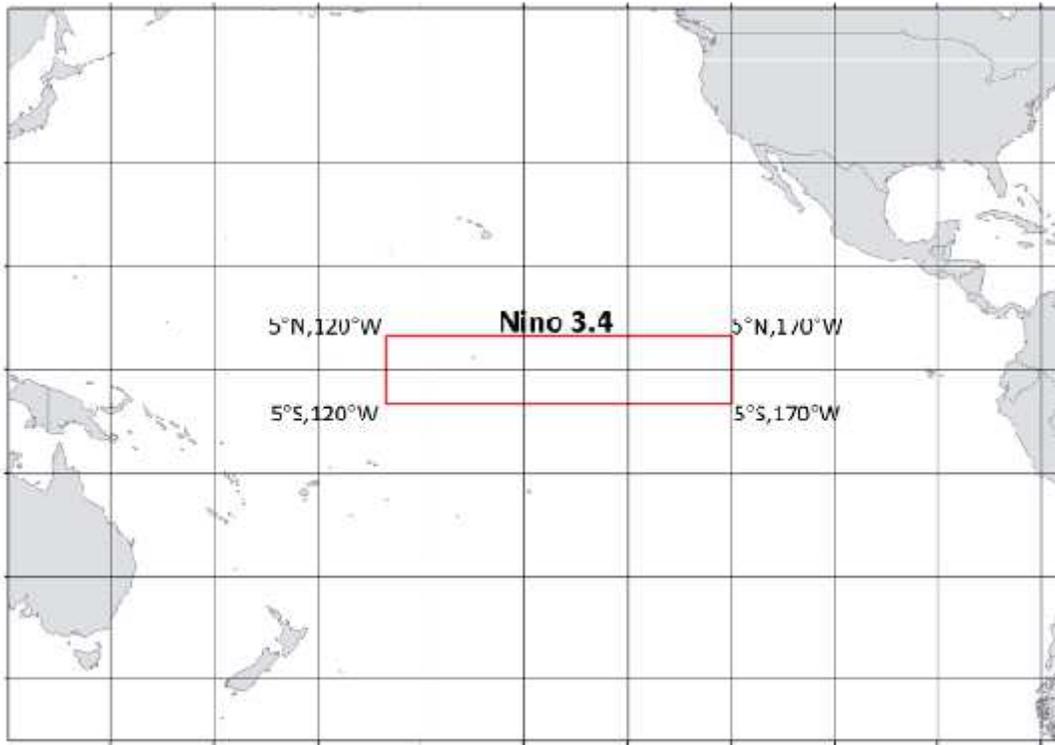
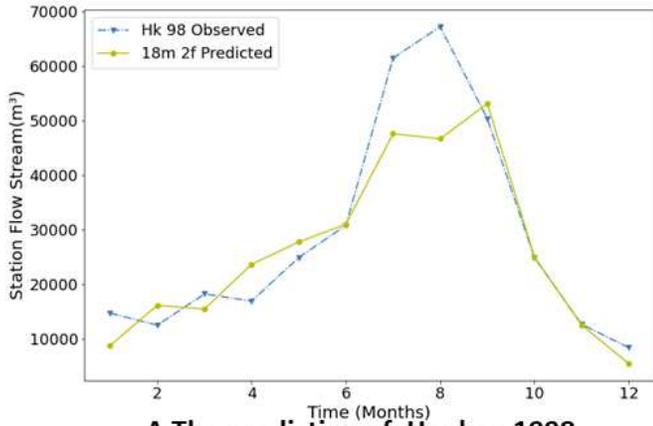
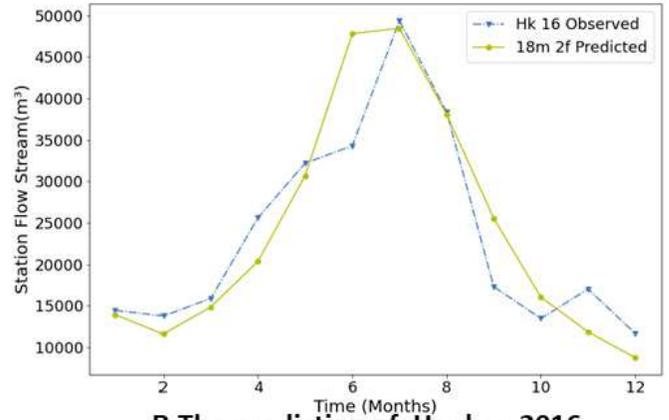


Figure 9

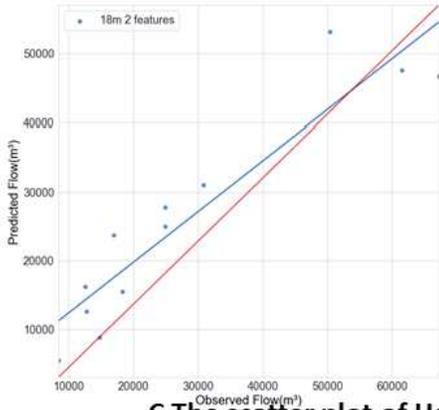
The Nino 3.4 area . Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



A The prediction of Hankou 1998

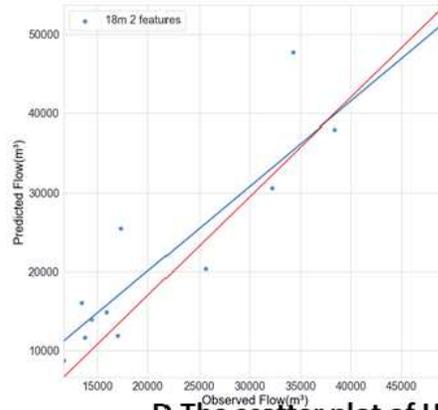


B The prediction of Hankou 2016



C The scatter plot of Hankou 1998

**$R^2=0.83$
LMI=0.67
WI=0.95
RMSE=7849.91**

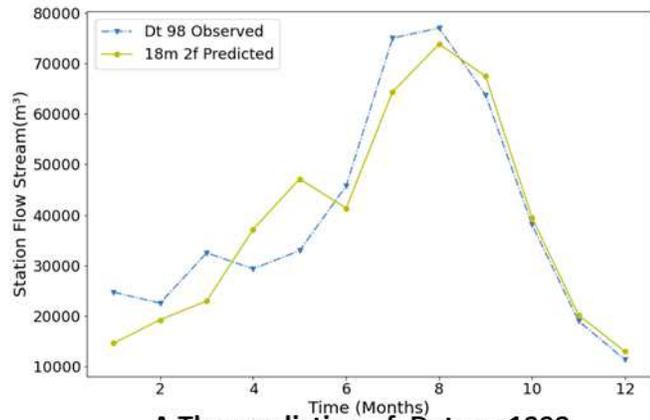


D The scatter plot of Hankou 2016

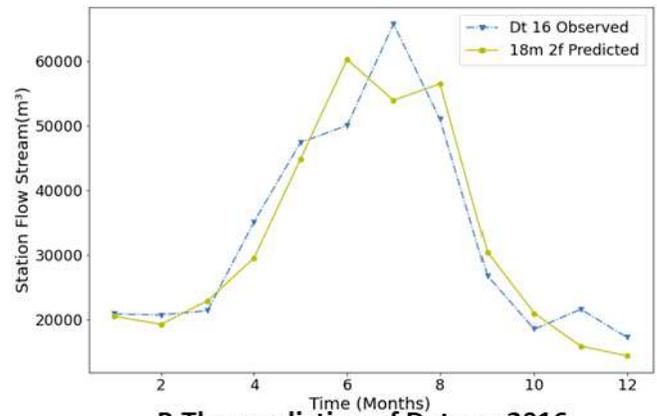
**$R^2=0.80$
LMI=0.64
WI=0.96
RMSE=5228.97**

Figure 10

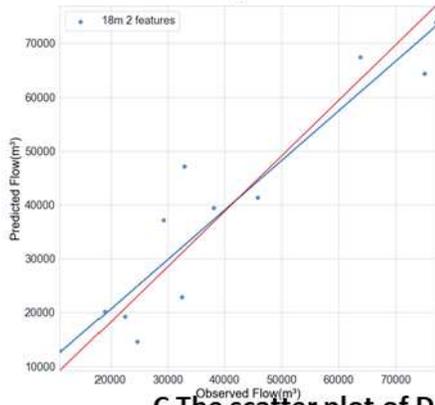
The result of Hankou Station



A The prediction of Datong 1998

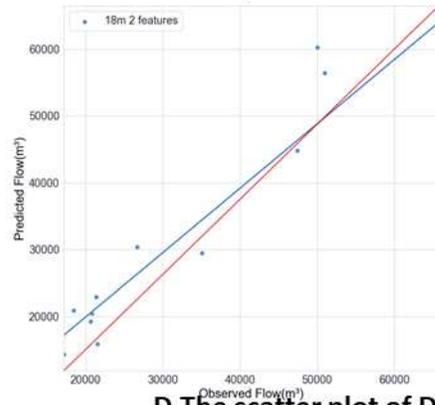


B The prediction of Datong 2016



C The scatter plot of Datong 1998

$R^2=0.88$
 $LMI=0.66$
 $WI=0.97$
 $RMSE=7249.87$

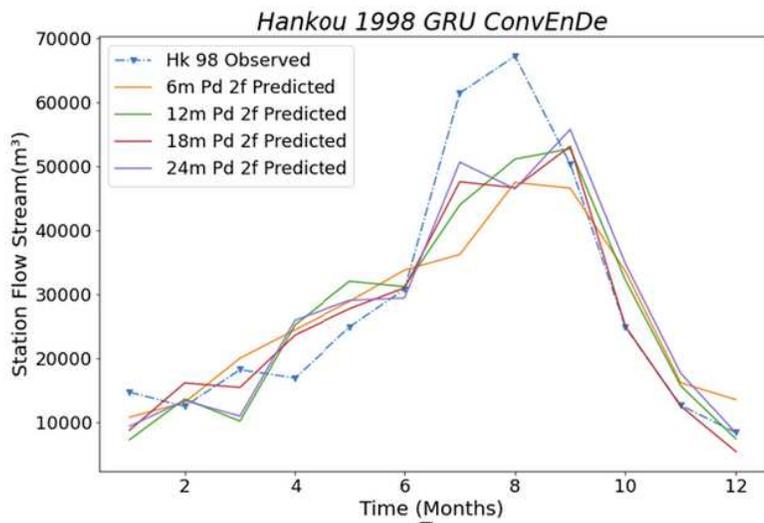


D The scatter plot of Datong 2016

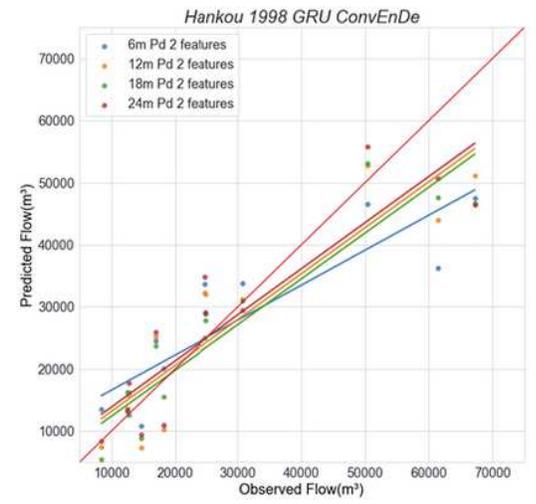
$R^2=0.87$
 $LMI=0.68$
 $WI=0.97$
 $RMSE=5610.82$

Figure 11

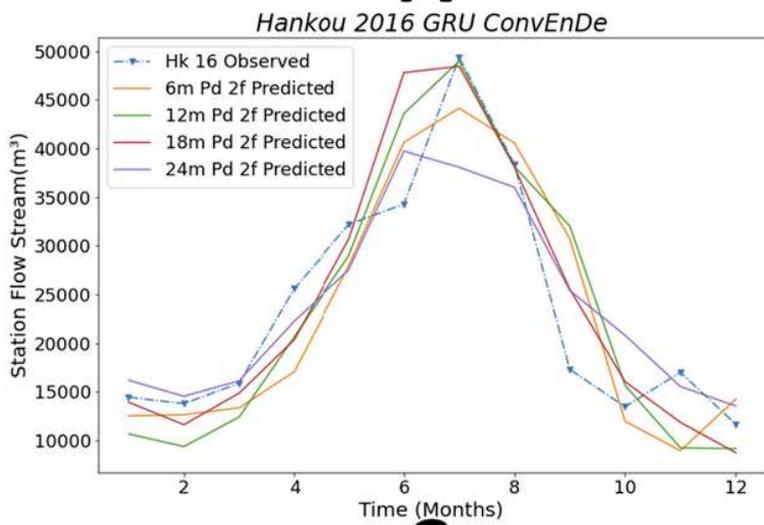
The result of Datong stationt



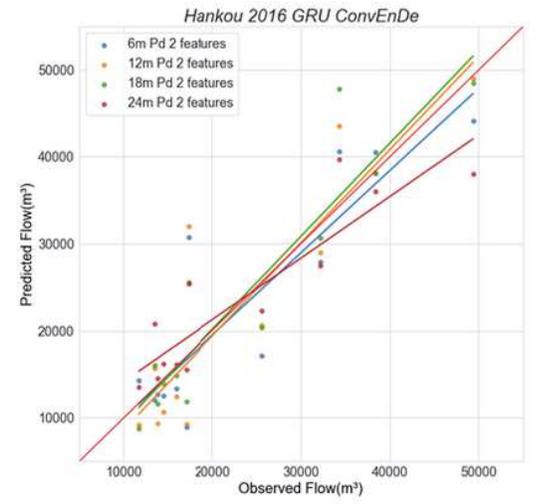
A



B



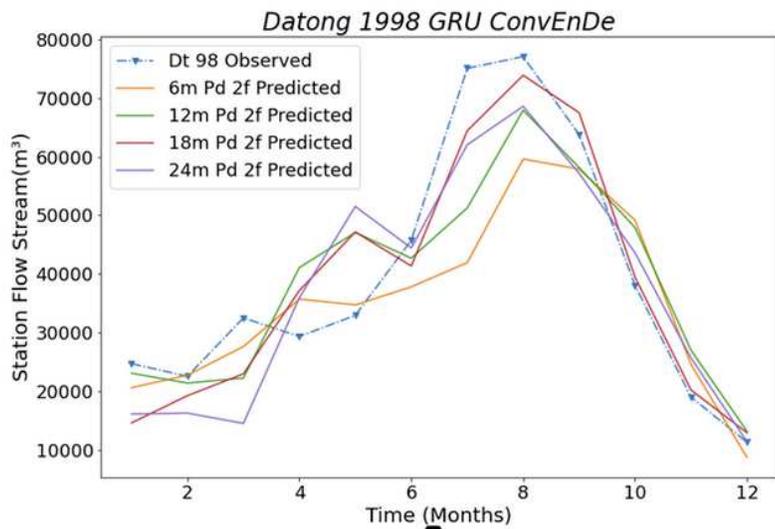
C



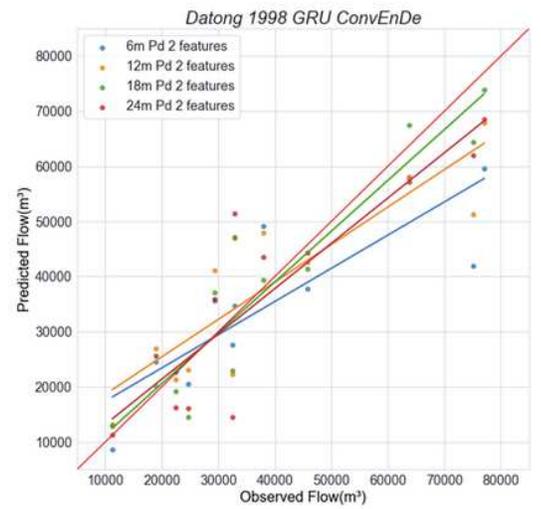
D

Figure 12

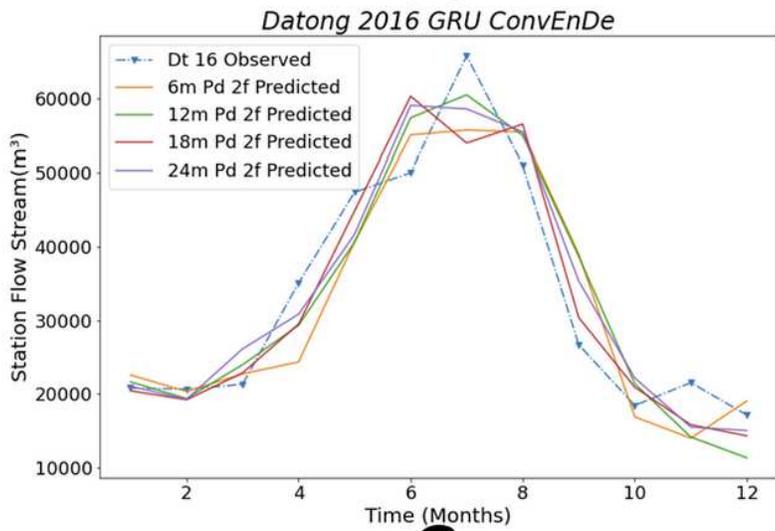
Comparison among the Encoder decoder GRU with different period in Hankou station



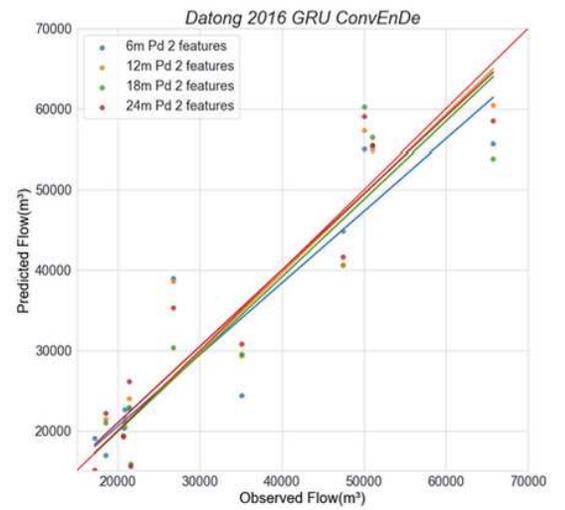
A



B



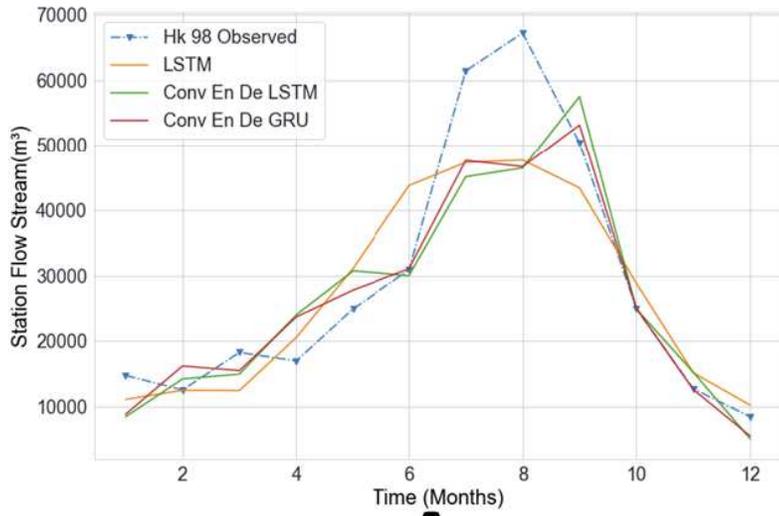
C



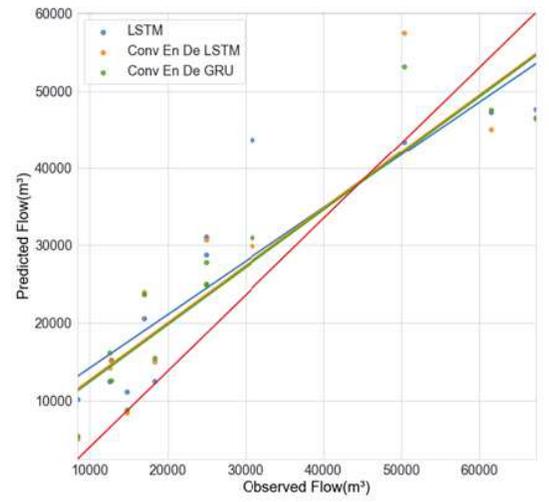
D

Figure 13

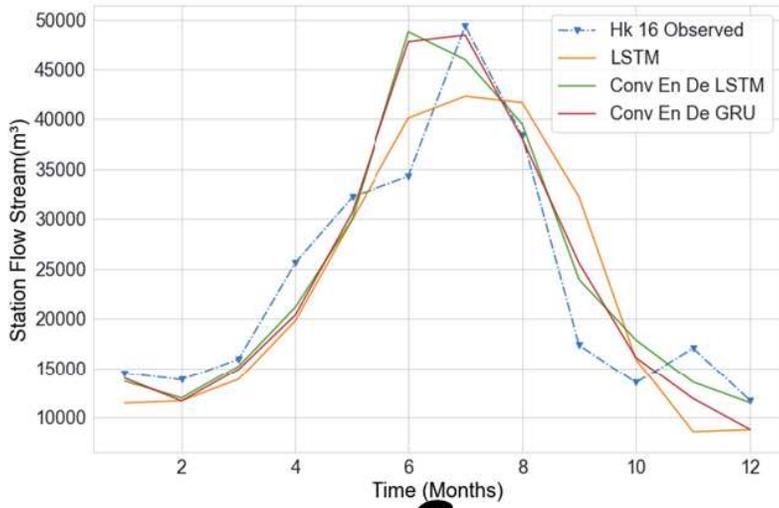
Comparison among the Cony Encoder Decoder GRU with different period in Datong station



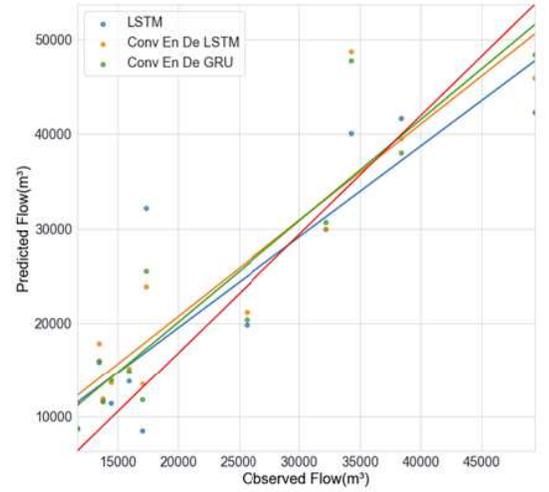
A



B



C



D

Figure 14

Comparison with different models in Hankou station

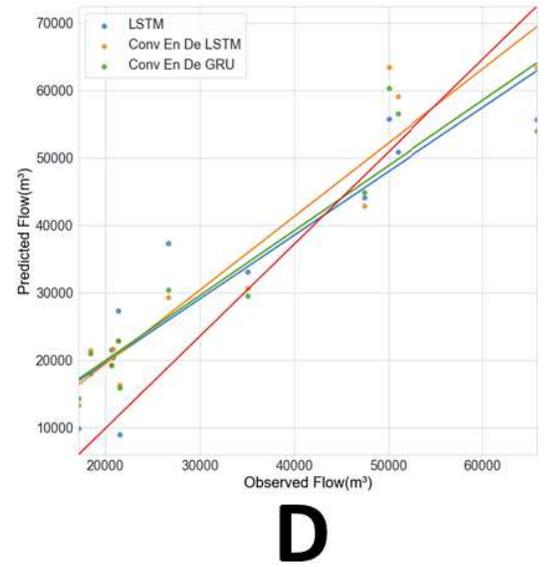
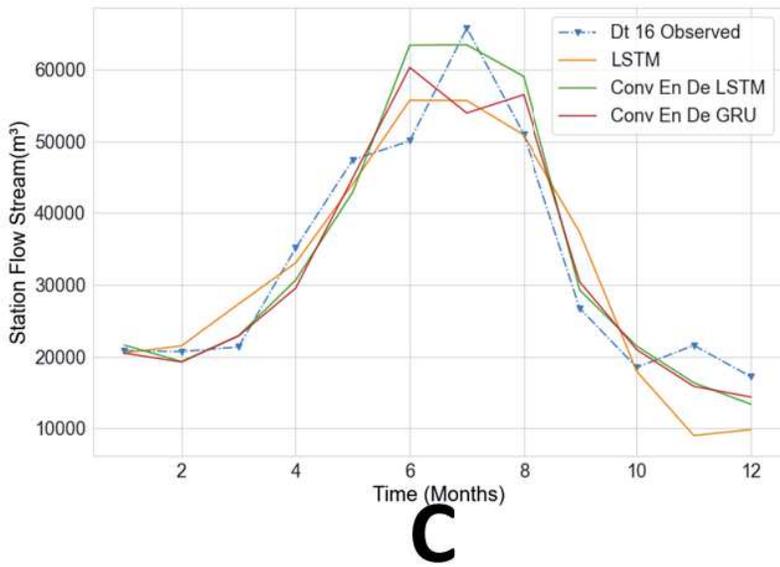
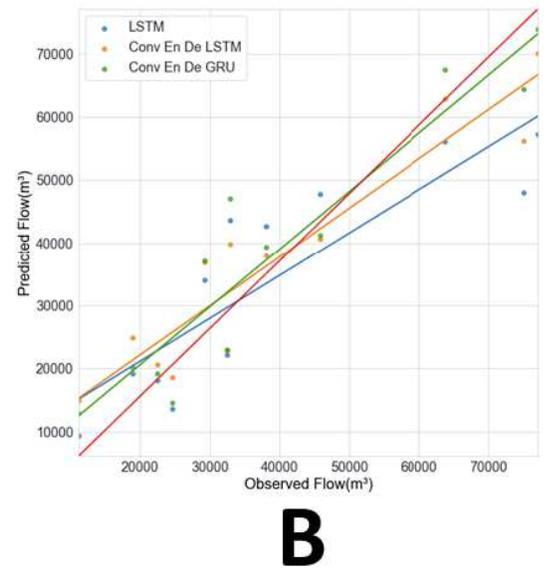
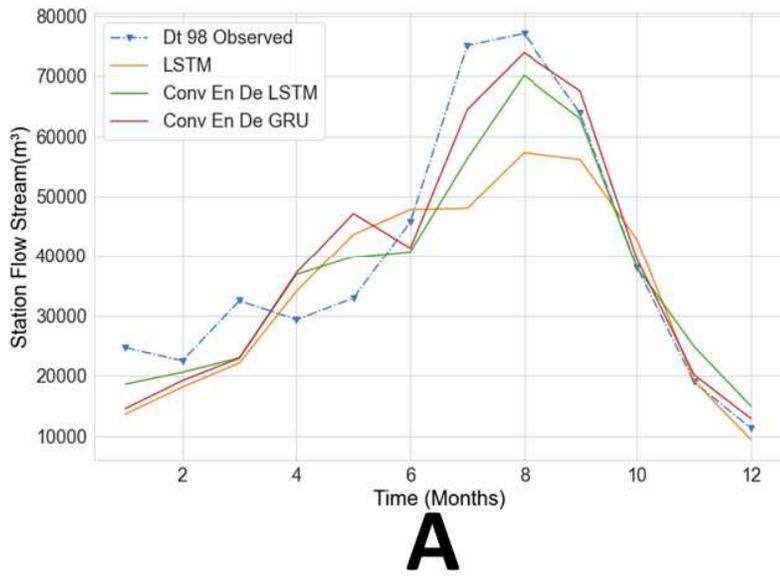
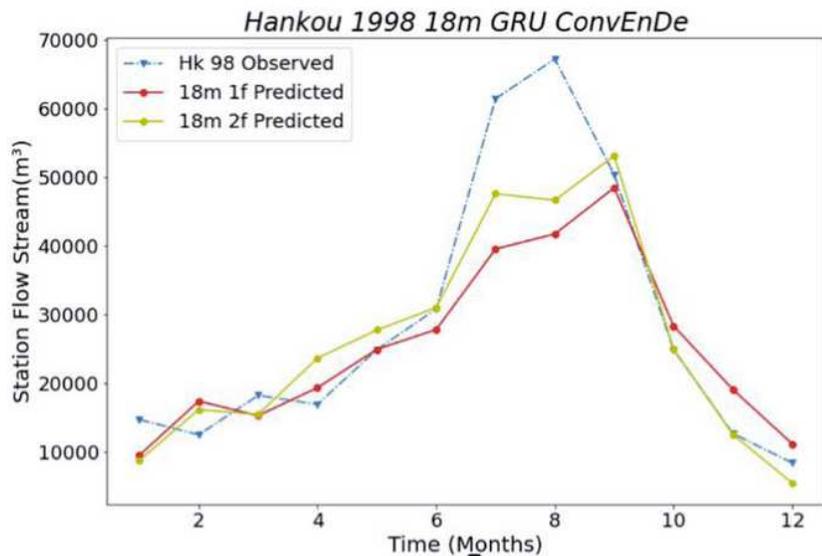
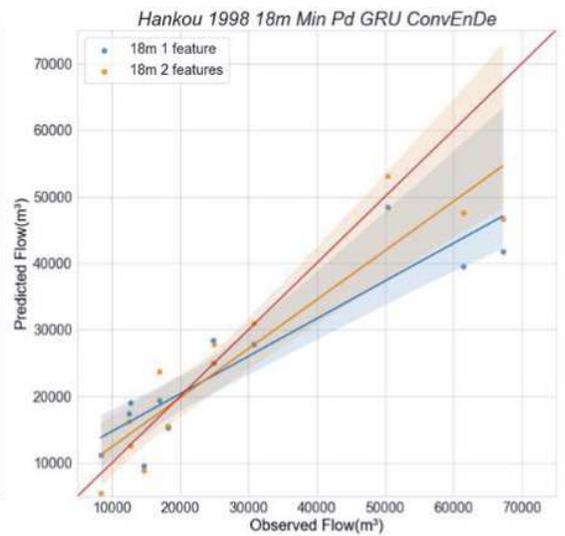


Figure 15

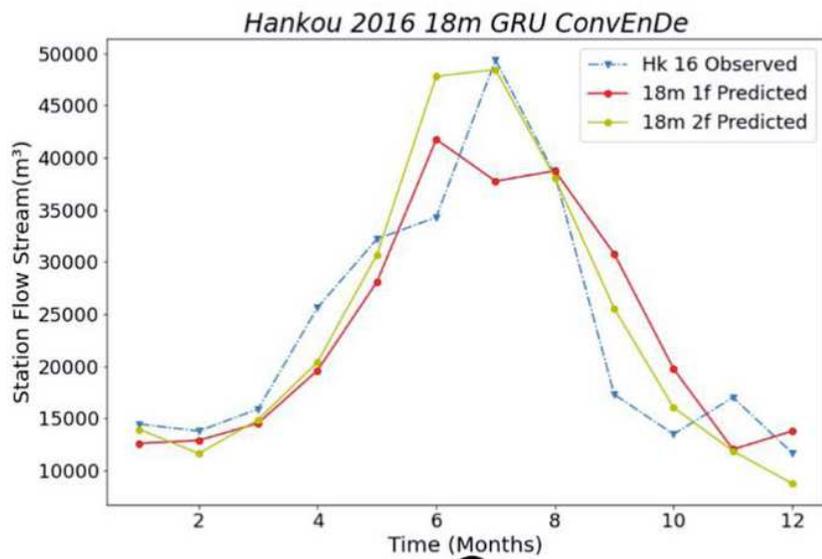
Comparison with different models in Datong station



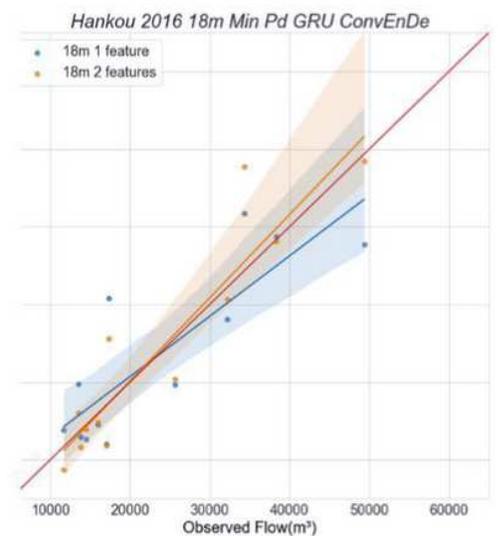
A



B



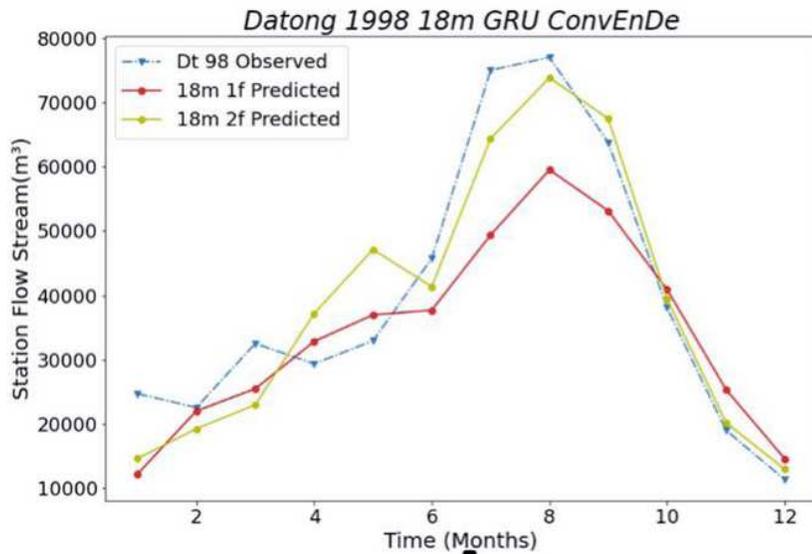
C



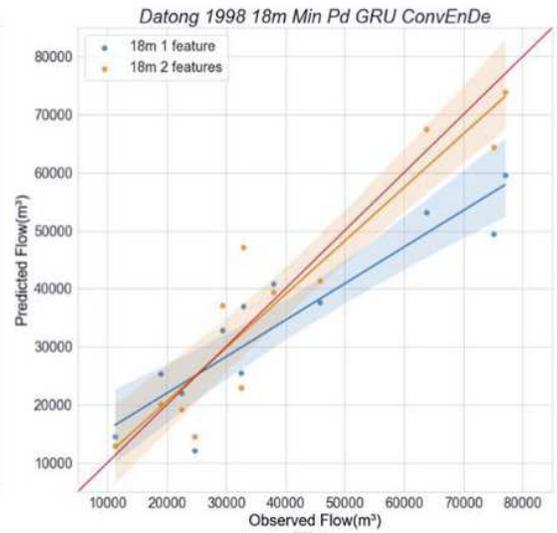
D

Figure 16

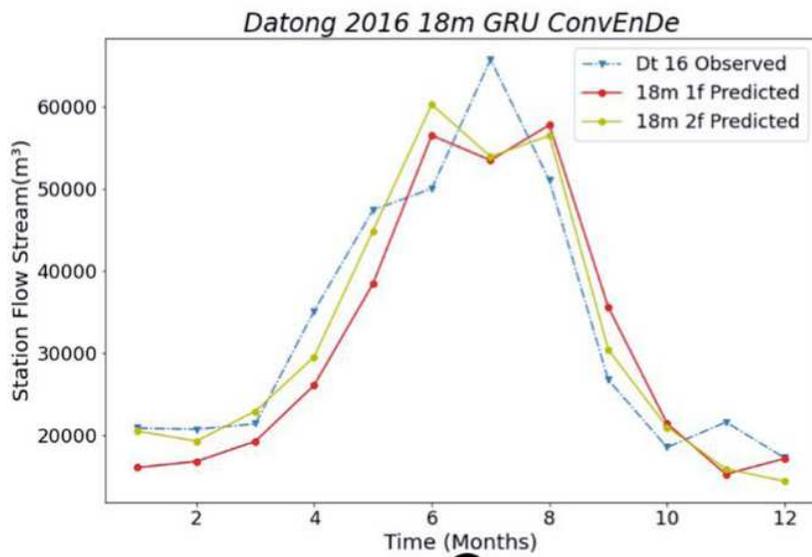
Comparison with different features in Hankou station



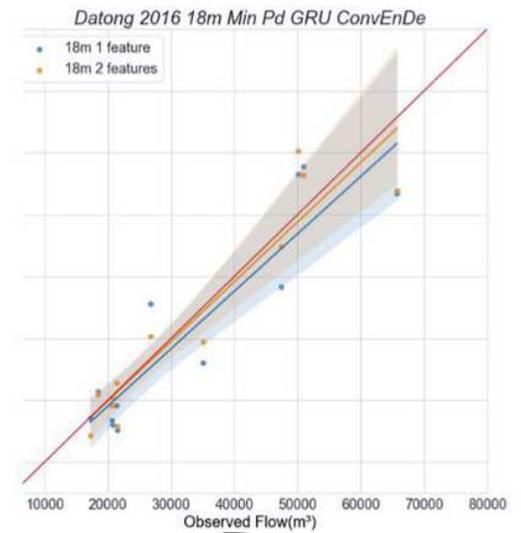
A



B



C



D

Figure 17

Comparison with different models in Datong station