

Received April 14, 2020, accepted May 4, 2020, date of publication May 11, 2020, date of current version May 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993874

Streamflow Prediction Using Deep Learning Neural Network: Case Study of Yangtze River

DARONG LIU^{1,2,3}, WENCHAO JIANG⁴, LIN MU^{2,3}, AND SI WANG^{1,2,3}

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

²Shenzhen Research Institute, China University of Geosciences, Shenzhen 518057, China

³Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

⁴School of Computer Science, Wuhan University, Wuhan 430070, China

Corresponding author: Lin Mu (moulin1977@hotmail.com)

This work was supported in part by the Discipline Layout Project for Basic Research of Shenzhen Science and Technology Innovation Committee under Grant JCYJ20170810103011913, and in part by the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) under Grant GML2019ZD0604.

ABSTRACT The most important motivation for streamflow forecasts is flood prediction and longtime continuous prediction in hydrological research. As for many traditional statistical models, forecasting flood peak discharge is nearly impossible. They can only get acceptable results in normal year. On the other hand, the numerical methods including physics mechanisms and rainfall-atmospherics could provide a better performance when floods coming, but the minima prediction period of them is about one month ahead, which is too short to be used in hydrological application. In this study, a deep neural network was employed to predict the streamflow of the Hankou Hydrological Station on the Yangtze River. This method combined the Empirical Mode Decomposition (EMD) algorithm and Encoder Decoder Long Short-Term Memory (En-De-LSTM) architecture. Owing to the hydrological series prediction problem usually contains several different frequency components, which will affect the precision of the longtime prediction. The EMD technique could read and decomposes the original data into several different frequency components. It will help the model to make longtime predictions more efficiently. The LSTM based En-De-LSTM neural network could make the forecasting closer to the observed in peak flow value through reading, training, remembering the valuable information and forgetting the useless data. Monthly streamflow data (from January 1952 to December 2008) from Hankou Hydrological Station on the Yangtze River was selected to train the model, and predictions were made in two years with catastrophic flood events and ten years rolling forecast. Furthermore, the Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Willmott's Index of agreement (WI) and the Legates-McCabe's Index (LMI) were used to evaluate the goodness-of-fit and performance of this model. The results showed the reliability of this method in catastrophic flood years and longtime continuous rolling forecasting.

INDEX TERMS Yangtze River, hydrological time series forecasting, streamflow prediction, empirical mode decomposition, deep learning.

I. INTRODUCTION

Yangtze River is the largest river in China and the third-longest river in the world. Providing accurate and reliable future streamflow information plays an important role in flood-control and disaster relief, thus developing excellent streamflow forecasting methods has attracted increased attention from hydrology researchers [1]. The flow of the Yangtze River is usually affected by numerous factors, such as rainfall, evaporation, water stage, and groundwater, etc., which are

The associate editor coordinating the review of this manuscript and approving it for publication was X. Huang¹.

usually nonlinear, complex, abrupt [2] and dependent on a large number of parameters including temporal and spatial variations. Accuracy and skill of flow prediction models can have a direct influence on management decisions of water resources. Various statistical and conceptual streamflow prediction models have been developed to help urban planners, administrators, and policymakers in better and informed decision making [3]. The accurate analysis of the evolution process of the flow at hydrological stations has received increased attention from hydrology researchers. Hydrological processes are driven by natural fluctuations over the physical scale, and the resulting variance in the underlying model input

datasets [4]. The classical run-off prediction methods could be roughly divided into two categories: numerical models and statistical prediction methods. The former is the numerical prediction model based on the study of atmospheric circulation, the evolution of long-term weather processes, and physical conditions [5]. The latter type of method is based on the fitting of the measured run-off process to establish a statistical prediction model [6], [7]. Wang *et al.* used the Hermite Projection Pursuit Regression combined with Social Spider Optimization (SSO) and Least Square (LS) to make annual maximum flood peak discharge [8]. The results showed that the SSO and LS algorithm could improve the prediction accuracy of peak streamflow. Sang *et al.* developed an Adaptive Metropolis-Markov Chain Monte Carlo-Wavelet Regression (AMMC-MC-WR) model to improve the accuracy in the hydrologic time series forecasting [9]. There are also some researchers proposed the hybrid model called Multivariate Adaptive Regression Spline to forecast streamflow pattern in semi-arid region [10].

However, there are disadvantages in these models. Firstly, the effective prediction period of traditional numerical models including physical mechanism and the atmospheric-rainfall model is about one month ahead, which is too short to be used for the hydrological application. Secondly, most classical statistical methods cannot forecast the flood peak discharge precisely, which is the most important motivation for streamflow prediction. Last but not least, most research towards river flow prediction is yearly [11], [12], seasonal [13], daily [14], [15] or even hourly [16]. However, the flood season on the Yangtze River usually lasts 2-3 months within a year. The major purpose of the monthly streamflow prediction is flood control and disaster relief. What we need is an accurate monthly flow forecast, which can help the water conservancy department to know the drastic changes in the peak flow 2-3 months or even 6 months ahead. It is very significant for hydrological disaster prevention and mitigation [17]–[19]. In view of the mentioned shortcomings, new techniques for flow forecasts need to be introduced.

The machine learning is also a statistical method with data-driven and self-adaptive features. Last decades, machine learning techniques have been employed in solving prediction problems in many domains [20] including hydrological research [21], [22]. These methods, such as Support Vector Machine (SVM) [23]–[25] and Artificial Neural Network (ANN) [19], [27], [28], have been widely utilized for river flow prediction. ANNs have good performance in dealing with nonlinear time series and there are numerous ANN applications for streamflow prediction [3], [16], [29]–[31]. They do not need the complex nature of run-off processes and do not rely on high-precision rainfall forecasting when dealing with the hydrological processes. In many cases, machine learning can provide better performance for short and intermediate-term predictions than traditional models [32]. Demirel *et al.* focused on the issue of streamflow prediction using the soil and water assessment

tool (SWAT) and ANN models [33]. The study found that the ANN model can predict peak discharge more efficiently than the SWAT model. Yaseen *et al.* introduced the Emotional Neural Network (ENN) to make hourly flow predictions [16]. Non-Linear Input Variable Selection Approach Integrated With Non-Tuned Data Intelligence Model was introduced by Hadi [34] to forecast the streamflow. Several hydrological variables including rainfalls, temperature and evapotranspiration were used to build and train the model. The performance of them was better than other machine learning models, the prediction period is too short to be used in actual application, though. Adaptive Neuro-Fuzzy Inference System (ANFIS) is a kind of artificial neural network, too. A method based on ANFIS- Particle Swarm Optimization (PSO) was proposed to forecast highly stochastic river flow in tropical environment [31]. Its main limitation is that it does not take into account the streamflow of the basins in the non-tropical regions just like the Yangtze River Basin (sub-tropical zone) or The Yellow River (temperate zone). Most of river basins in China are not in tropical zone. The ANN methods such as SVM, Radial Basis Function Network (RBFN) [11], Heuristic-Regression [14], ANFIS, etc. have the capacity of representing highly non-linear correlations between input and output that statistical models do not have. However, there are still some disadvantages when dealing with nonstationary streamflow data.

In recent years, due to sufficient observational data and increased computational power, deep neural networks have been applied to prediction studies as a sub-field of machine learning methods. Most deep learning architectures evolve from NN consisting of layers (hidden layers, input layers, and output layers) and neurons. They can readily learn temporal dependence and handle temporal structures on time series data. Consequently, deep learning algorithms are effective for analyzing non-linear data and constructing preferred predictive models. Of all deep learning architectures, the Recurrent Neural Networks (RNNs) can achieve satisfactory performance in time series tasks. They have the ability to retain important information in the previous time step due to the cycle between cells. The Long Short-Term Memory (LSTM), as a particular type of RNN, can remember information for a long time and could forget the useless information through the training process due to changes in the RNN internal structure. This extending structure makes the LSTM outperform the other deep learning architectures in long-term time series prediction [35]–[37]. Based on the LSTM framework, an Encoder Decoder LSTM (En-De-LSTM) structure was proposed for sequence-to-sequence long-term prediction. It can reconstruct the input sequence in the encoder part and forecast the next sequence in the decoder part. The LSTM-based Encoder Decoder (En-De) architecture has been adopted in many fields such as language translation [38], [39], image captioning [40], and speech recognition [41], also streamflow simulation [15]. However, a few studies focus on analyzing data sets of streamflow indices with the LSTM-based En-De architecture.

Recently, the data-driven modeling approaches combined with Empirical Mode Decomposition (EMD) are being widely used as surrogates for physically-based models. Because they overcome some limitations associated with physically-based approaches. Huang *et al.* proposed a signal decomposition method called EMD in 1998 [42]. It is usually used to preprocess the nonlinear and nonstationary data. In the hydrological time series prediction problems, monthly streamflow data contains several different frequency components. However, previous studies [23] had directly used the original series as the input variables when they built a prediction method, which may have led to missing information among different frequencies. Several researchers combined the EMD and modified-EMD with the machine learning models to forecast the trends of streamflow data. For instance Huang *et al.* [25] and Meng *et al.* [26] used the modified EMD to get better results. Wang, Xu *et al.* presented a model combined Ensembled EMD with SVM which used the PSO algorithm [24] to predict the rainfall-runoff. But these EMD-based SVM models have some drawbacks, e.g. slow learning, trap related to local minimization value and saddle points, also the over-fitting. None of them made the longtime predictions which longer than 5 years either. Besides, the M5 model tree (M5Tree) and Multivariate Adaptive Regression Spline (MARS) combined EEMD was proposed to forecast the daily river flow in Iran and South Korea [14]. However, this research is focused on daily run-off forecasts. Its guidance towards flood control is limited.

As mentioned above, the deep neural network is rarely used in flow prediction, especially the EMD-based long-time monthly prediction deep learning model. In this study, a model combining the EMD and En-De-LSTM was proposed to forecast the longtime (10 year) monthly streamflow data from Hankou Hydrological Station on the Yangtze River. The proposed model combines the advantages of EMD decomposition of different frequency data and the superiority of the LSTM neural network in processing time series data, avoiding a series of above problems that traditional numerical and physically-based models also ANNs may bring.

II. METHODOLOGY

A. EMPIRICAL MODE DECOMPOSITION

EMD is a data-driven algorithm working efficiently for data with the non-linear and non-stationary features. The decomposed signal component length is the same as the original signal without departing from the time domain. In this study, EMD was used to pre-process the original Yangtze River flow data. Generally, the original time series can be decomposed into a collection of Intrinsic Mode Functions (IMFs) and a residue by EMD. Every decomposed IMF must satisfy the following requirements:

1. The difference between the number of extreme values and the number of zero crossings must be zero or one.
2. At any time, the average of the envelope defined by the local maximum and the envelope defined by the local minimum should be zero.

For a given time series $T(t)$, the main steps of EMD can be described as follows:

1. Identifying all the local maxima $\max(t)$ and local minimal $\min(t)$ points of original time series data $T(t)$, then the mean envelope curve $m_1(t)$ can be formed by computing the average values on the $\max(t)$ and $\min(t)$

$$m_1(t) = \frac{\max(t) + \min(t)}{2} \quad (1)$$

2. Find the average of the upper and lower envelopes to obtain the mean envelope $m_1(t)$

3. The mean envelope is subtracted from the original signal $T(t)$ to obtain the first component $h_1(t)$:

$$h_1(t) = T(t) - m_1(t) \quad (2)$$

4. Check whether the $h_1(t)$ follows IMF conditions. If not, replacing $T(t)$ with $h_1(t)$ and return to Step-1 for the second-round screening:

$$h_2(t) = h_1(t) - m_2(t) \quad (3)$$

Repeat the above process for k times:

$$h_k(t) = h_{k-1}(t) - m_k(t) \quad (4)$$

until $h_k(t)$ meets with IMF conditions, and the first IMF of the $T(t)$ is obtained:

$$c_1(t) = h_k(t) \quad (5)$$

5. Subtracting $c_1(t)$ from the original signal $T(t)$ yields the remaining amount $r_1(t)$:

$$r_1(t) = s(t) - c_1(t) \quad (6)$$

6. Based on $r_1(t)$, repeat Step 1 to Step 5, up to n times until the last IMF $r_n(t)$ becomes a monotonic function, and $r_n(t)$ is the residue of the original time series. Finally, the original signal $T(t)$ can be expressed as a combination of n IMF components ($c_k(t)$) and a residue($r_n(t)$):

$$s(t) = \sum_{k=1}^n c_k(t) + r_n(t) \quad (7)$$

The advantage of EMD is that it decomposes non-linear and non-stationary original signals into a number of smooth signals without losing any information.

B. LONG SHORT-TERM MEMORY

LSTM is an artificial RNN, which was proposed by Sepp Hochreiter *et al.* in 1997 [43]. Compared to the traditional feed-forward neural network, where data move from the input layer to the output layer through one or multiple layers, the RNN with an internal hidden state (or 'memory') allowing data to cycle through the network is more efficient and stable in dealing with non-linear long-range time-varying problems than traditional methods. Extreme The Back-Propagation Through Time (BPTT) algorithm is used to train RNNs [44], which can unfold the network, calculate error and update weights over each step.

In other words, as the sequence progresses, the previously hidden layer will affect the hidden layer behind. RNNs have demonstrated that they have advantages in supervised sequential problems. However, when the matrix with relatively small values is multiplied by multiple matrices, this can eventually lead to the gradient decreasing exponentially and disappearing after a few steps. This is the problem of Vanishing Gradient [45], which can make RNNs miss valuable distant information.

Owing to the aforementioned deficiency of RNNs, LSTM was proposed to overcome the vanishing gradient issues. The LSTM consists of the cell memory that stores the summary of the past input sequence, and the gating mechanism by which the information flow among the input, output, and cell memory are controlled. Figure 1 shows the standard LSTM structure in which neurons in the hidden layer are weighted.

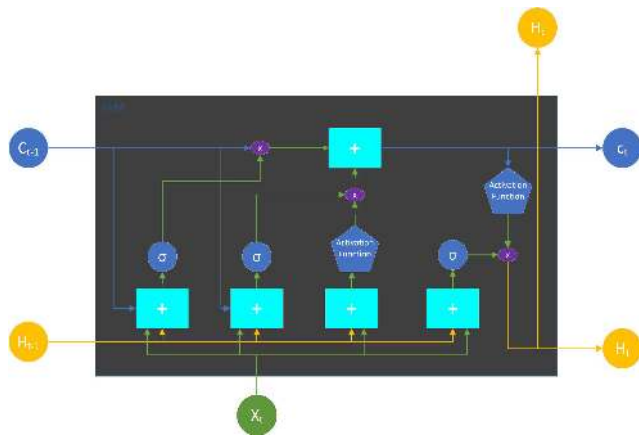


FIGURE 1. The network of LSTM memory unit.

It could learn long-term dependencies by adding the input gate, the forget gate and the output gate to the memory unit in RNNs. Because the LSTM is inherited from the RNN, it still maintains the connections inside the hidden layers and enhances the quality of connections between the back and front nodes through three gates (i.e. forget gate, input gate, output gate). The theory of LSTM can be described as equation (8)-(13).

$$f_t = \sigma (W_f [H_{t-1}, X_t] + b_f) \tag{8}$$

$$I_t = \sigma (W_i [H_{t-1}, X_t] + b_i) \tag{9}$$

$$\bar{C}_t = \tanh (W_c [H_{t-1}, X_t] + b_c) \tag{10}$$

$$C_t = f_t * C_{t-1} + \bar{C}_t \tag{11}$$

$$O_t = \sigma (W_o [H_{t-1}, X_t] + b_o) \tag{12}$$

$$H_t = O_t * \tanh (C_t) \tag{13}$$

Equation (8) shows how the forget gate f_t works. The main function of the forget gate is to decide which data should be discarded from one memory unit. The current input and previous hidden state are represented by X_t and H_{t-1} , respectively. The activation is a sigmoid function which makes the output value range from 0 to 1 in most of the LSTM neurons.

In this model, the Rectified Linear Unit (ReLU) function was chosen as the activation function of the proposed model. Through all the activation functions in machine learning, ReLU function has better gradient propagation. When compared with the sigmoid function, it has fewer vanishing gradient problem. Also, it can make computation, addition, and multiplication more efficient than other functions. All these features and advantages will make the prediction process smoother and the results more robust. Thus, the input gate I_t combines X_t and H_{t-1} and passes them through the sigmoid function in Equation (9). Then, in Equation (10), a hyperbolic tangent is usually used as the activation function (also sigmoid function or others), which creates a candidate vector \bar{C} and this vector can be added to the memory state. After the above-mentioned steps, the old memory state C_{t-1} is replaced by a new memory state C_t , as shown in Equation (11). Eventually, a memory unit will calculate how much information will be output through Equation (12) and (13). In summary, the weight matrices are $W (f, i, c, o)$ and bias vectors are $b (f, i, c, o)$ among the Equation (8)-(13), and they update iteratively in the LSTM network by the BPTT algorithm.

C. ENCODER DECODER LSTM

In this study, a hybrid neural network model, the En-De-LSTM, is used in long-term prediction problems. The En-De architecture is usually used in seq2seq (sequence to sequence) problems [46]. The Encoder part can encode a variable-length sequence into a fixed-length vector representation, and the Decoder part can decode the given fixed-length back into a variable-length sequence. An En-De-LSTM architecture is proposed and illustrated as Fig. 2.

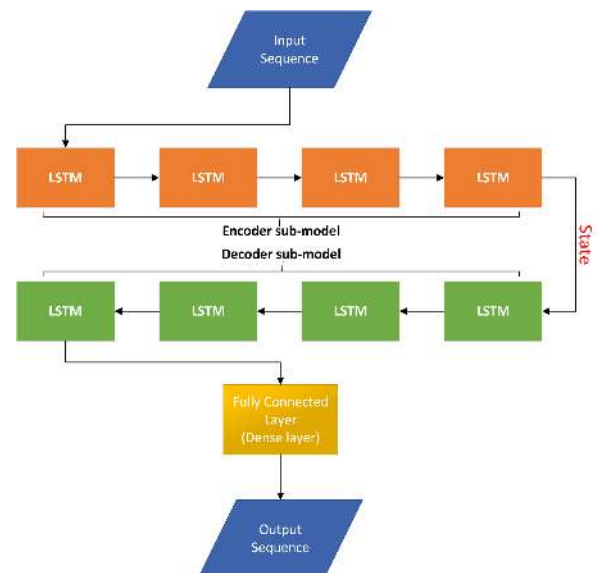


FIGURE 2. The architecture of En-De-LSTM.

It is composed of two Stacked LSTM (four-layers LSTM): Encoder stacked LSTM and Decoder stacked LSTM. When compared to the Vanilla LSTM, the structure of the network

was too simple to learn valuable information. And the performance of results was too far from the actual observations. Besides, the other LSTMs like bidirectional LSTM, CNN-LSTM, and Conv-LSTM were too complicated for this sequence prediction problem. Although they can get more precise value in normal year than the other structures, when facing the extremely high value, the model would treat the extremum as the abnormal value and forget them. The complexity becomes a limitation for extremum forecasting. Predictions of extremum is an indispensable part in this study. Above all, the encoder part and decoder part chose the stacked LSTM structure. The encoder LSTM could automatically extract the features from input sequential data and transforms them into a fixed-length internal state, which provides the context for the decoder sub-model. Then in the decoder stacked LSTM, which consists of the decoder part, the internal state has interpreted the data to a fully connected layer then generates the final sequential data prediction.

D. EMD-ENCODER DECODER-LSTM MODEL

The original hydrological time series are non-linear and non-stationary. They were used to make forecasts which recurrently led to some missing features of different resolutions. EMD is based on the theory of local scale separation. When compared to the other data decomposition technique, EMD does not need any predetermined basis functions. It is a fully self-adaptive algorithm. It is usually used to decompose the nonlinear and nonstationary data just like streamflow data. The original flow data usually contains several different frequency components. If original data is used in prediction problem directly, some important data may lose. After the training and validation process inside En-De-LSTM, all of IMFs could be reconstructed to the primary data without any loss. This is also the major advantage of EMD. In this study, the original data was decomposed into a set of IMFs and a residue, respectively. The whole procedure of EMD-Encoder Decoder LSTM (EMD-En-De-LSTM) was illustrated in Figure 3. As is mentioned above, the LSTM could predict the time series data efficiently. Because it could remember the important information in the training process and forget the useless data simultaneously. These two features of LSTM could improve the accuracy. En-De module is a powerful function in the seq2seq problem, it can make the prediction more efficient. When the decomposed process completed, several IMFs that belong to several different frequencies were imported into the En-De-LSTM structure. The combination of EMD and En-De-LSTM will give full play to their strengths and make the predictions more accurate.

Initially, the original streamflow time series was decomposed into a set of IMFs and a residue by EMD. The IMFs are normalized to a scale between 0.0 and 1.0. Furthermore, all the components are input to the En-De-LSTM model for training. The most important step of training model is the hyperparameter optimization. Grid search method was chosen to determine the hyperparameters of the model. Grid Search is an exhaustive searching through a manually speci-

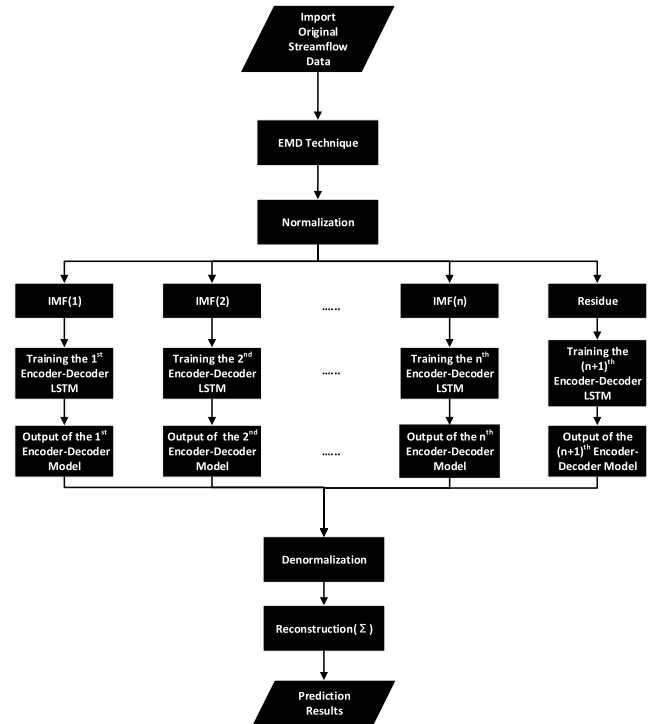


FIGURE 3. The flowchart of the proposed model.

fied subset of the hyperparameter space of a machine learning algorithm [47]. It could find the optimal hyperparameter values by checking all parameter combinations based on a given model. In most cases, it's useful and powerful because it's exhaustive and leaves no stone unturned. Finally, the forecasting values are normalized to original values, plus, the final prediction values can be obtained through linear addition of all denormalized values.

- The input sequential data is decomposed into a set of IMFs and a residue with the EMD technique. Then, the decomposed components are normalized to the scale between 0.0 and 1.0.
- Transforming the normalized data to the object characterized by features and the label, which can be input to the En-De-LSTM model.
- Determining initial values of hyper-parameters of the En-De-LSTM model with Grid Search method.
- Training and building separated predictive models to perform forecasts for each IMF and one residue.

Then, the predicted values are renormalized to the original scales. Finally, the forecasting results can be obtained by linear addition of all renormalized predicted values.

III. EXPERIMENT

A. STUDY AREA AND DATA

This study focuses on the streamflow data at the Hankou Hydrological Station. The data used in this study is the monthly streamflow data at the station from January 1952 to December 2018 collected by the Yangtze River

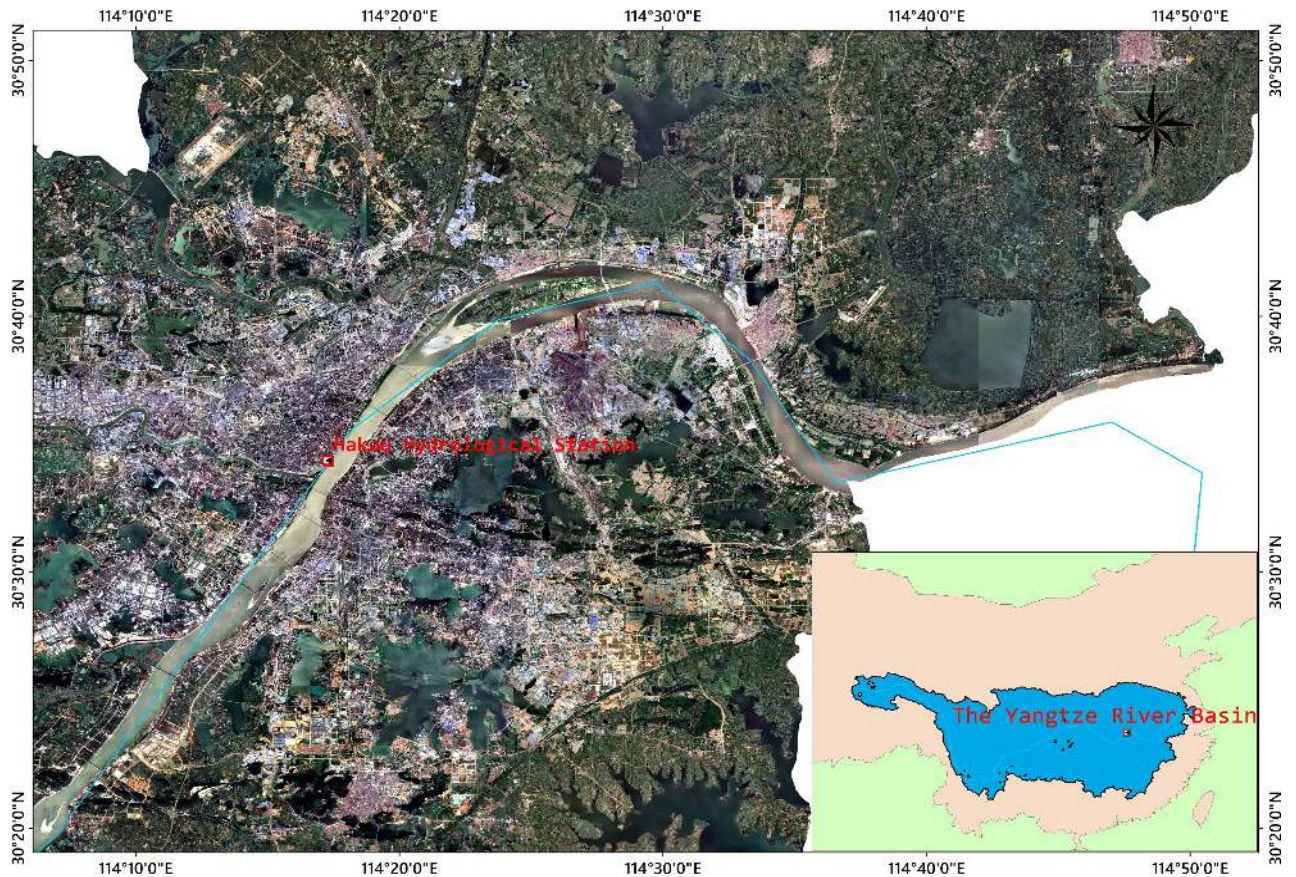


FIGURE 4. The location of the Yangtze River basin and Hankou Hydrological Station in Wuhan.

Water Resources Commission. The Hankou Hydrological Station located in Wuhan is the boundary of the middle and lower basin of the Yangtze River, as shown in Figure 4. Wuhan, located in the east of the Jiangnan Plain, is the largest city in central China and one of the biggest cities in the Yangtze River basin. It is also where the Hanjiang River, the largest tributary of the Yangtze River, flows into the Yangtze river system. About 70% of floods in China occur in the Yangtze River Basin [48]. According to the data from China Meteorological Administration, the floods of the river mostly occur during the monsoon season from June to September, and devastating floods also occurred in some years (e.g. 1931, 1954, 1998, and 2016). Due to the location of Wuhan, the catastrophic floods inundated the city nearly every time. In 1998, the strongest subtropical highs led to strong precipitation lasting several weeks and resulted in more than three times the flow in the Yangtze River basin than during the flood season. Ultimately, it was the biggest flood that occurred in the Yangtze River basin over the past 50 years. From June to August in 2016, southern China suffered from severe weather, such as heavy rainfall, thunderstorms, and hail, which triggered many deadly flood peaks. Streamflow at the Hankou Station reached the historical extreme value after the flood in 1998. Therefore, predicting the river flows effectively at the Hankou Hydrological Station

is inevitable to the economic development, water allocation, flood-control, and disaster relief of the middle and lower basin of the Yangtze River. In this study, the streamflow data from Hankou Hydrological Station in 1998 and 2016 (the two most catastrophic flood years) were chosen to perform the forecasting. And a ten-year continuous rolling prediction from January 2009 to December 2018 was made to verify the longtime prediction ability.

B. NORMALIZATION

The original streamflow data at Hankou Hydrological Station from January 1952 to December 2018 were decomposed by EMD. Normalization is conducted before all the IMFs are analyzed by the En-De-LSTM model. The decomposed components are scaled to a range between 0 and 1.0:

$$Z_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (14)$$

where Z_i is a normalized value at time i , and X_i is the value of decomposed components.

C. LSTM NETWORK CONSTRUCTION

After the normalization of the original signal, the next step is to build an LSTM neural network and determine how to train the network by selecting model hyper-parameters.

A grid search technology is applied to achieve high precision by automatically searching for hyper-parameters. A number of subsets with all the possible combinations of the hyper-parameter values are defined in the grid constructed. Models can be built for each combination, and the model with the highest score among all the validation data is considered to be the best performance. Also, this set of hyper-parameter values is the optimal value. Furthermore, for the streamflow prediction problem, the different minimum prediction periods will get quite divergent results. In this study, the 6m-min-pd (6-month-minimum-period), 12m-min-pd, 18m-min-pd and 24m-min-pd were tried to get the best performance.

D. PERFORMANCE EVALUATION

When completing the prediction of each IMF and residue, all the prediction values ranging from 0-1 need to be denormalized to the scale of original values based on Equation (15):

$$X_i = Z_i \times [\max(X) - \min(X)] + \min(X) \quad (15)$$

The final prediction results can be obtained by the linear addition of all the renormalized values without any data loss.

In order to measure the performance of the results, we introduced the following statistical criteria:

The Root Mean Square Error (RMSE) was described as Equation (16).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

RMSE represents the square root of the second sampling moment of the difference between the predicted value and the observed value or the quadratic mean of these differences. When performing calculations on data samples used for estimation, these deviations are called residuals, and the deviations outside the calculated samples are called errors (or prediction errors) RMSE is used to aggregate the size of prediction errors at different times into a single prediction capability metric. For the time series prediction problem in this article, RMSE can measure the error of the overall prediction result. It could provide information for the model's forecasting skill and quantify the goodness-of-fit relevant to high river flow values [49].

The coefficient of determination (R^2) was defined as Equation (17).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (17)$$

Coefficient of determination (R^2) is often interpreted as the proportion of response variation "explained" by the regressors in the model. It will give some information about goodness-of-fit of a model. This measurement is very useful when evaluated a specific model. It provides a measure of how well observed outcomes are replicated by the model,

based on the proportion of total variation of outcomes explained by the model [50]. Thus, $R^2 = 1$ indicates that the fitted model explains all variability in y , while $R^2 = 0$ indicates no 'linear' relationship (for straight line regression, this means that the straight line model is a constant line

Willmott's Index of agreement (WI) is formed as the Equation (18)

$$WI = 1 - \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}_i| + |y_i - \bar{y}_i|)^2} \right] \quad (18)$$

The WI is popular criteria in hydrology which calculates the ratio of Mean Square Error (MSE) and can provide an advantage over RMSE [51]. It is proposed by Nash and Sutcliffe in 1970, Watterson in 1996 and refined by Willmott in 2011 [52]. It also is dimensionless, bounded by -1.0 and 1.0 and, in general, more rationally related to model accuracy than are other existing indices. It also is quite flexible, making it applicable to a wide range of model-performance problems. It can be used to measure the differences in the real and predicted means and variances [52].

And Legates-McCabe's Index (LMI) is written as the Equation (19).

$$LMI = 1 - \left[\frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{\sum_{i=1}^n |y_i - \bar{y}_i|} \right] \quad (19)$$

The LMI criterion considers absolute values for computation and gives errors and differences in the appropriate weights [53]. This measurement varies from zero to one, and higher the LMI, better of the goodness-of-fit of the hydrological model [54]. Legates and McCabe's measure is most similar to WI. It is used to show the covariation and differences between the various indices, as well as their relative efficacies. 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}_i represents the mean of observed values.

IV. RESULTS

A. THE ORIGINAL SERIES DECOMPOSED

The whole results of the EMD decomposed were shown as Figure 5. The original streamflow data from January 1952 to December 2018 was transformed to a set of IMFs and a residue.

B. THE TEN-YEAR ROLLING PREDICTIONS

After the decomposition of EMD, all of IMFs and residue were separated into two parts: the training set from January 1952 to December 2008 which was used to train the EMD-En-De-LSTM model with 6-month-minimum-prediction-period (6m-Min-Pd). The comparison and discussion about the arguments and models will be made in part V, including the validation sets from January 2009 to December 2018.

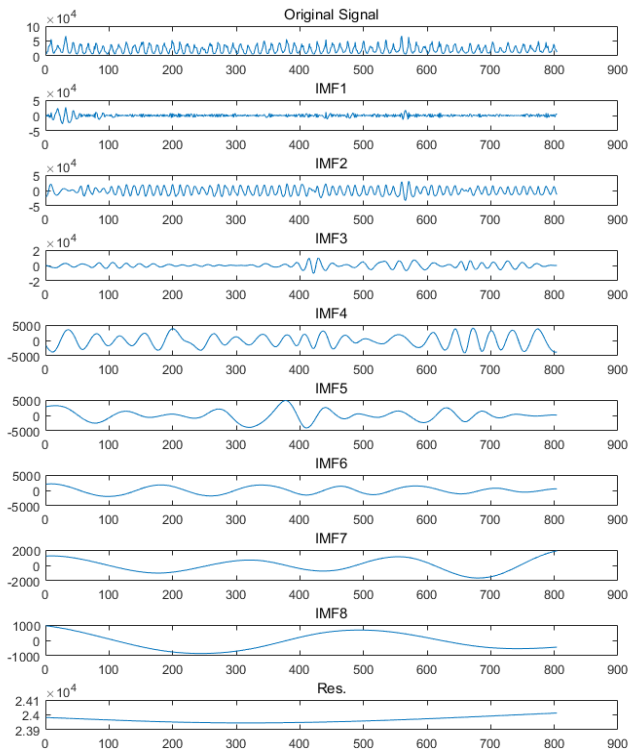


FIGURE 5. The original streamflow data and its IMFs (residue).

According to the aforementioned part III-C, the Min-Pd is an important argument for the proposed model. The different Min-Pd indicates the different minimum input variables and output variables while training. In this study, no matter which min-period was used, the validation set was set to multiples of 12, in other words, the data of a whole year, which could

make the prediction of flood season more precise. The forecasts are iteratively made using the streamflow values of the past 6 months to predict streamflow values in the following 6 months. The parameters of the predictive model are updated by the difference between predicted values and the target values.

Fig. 6 and Fig. 7 showed the results and errors in the rolling prediction. The RMSE was 1171.9 m^3 , and the R^2 was near to 1, which means the results were acceptable. The peak discharge in 2010, 2012, and 2016 reach the local maxima across 10 years. However, the predicted values in these years are close or higher to the observations. Moreover, the predictions from the other normal years also reached a high performance, which can be concluded from the value of WI and LMI. The whole results of the rolling forecasting verify the ability of long-term prediction of the proposed model perfectly.

C. THE CATASTROPHIC FLOOD EVENT YEAR PREDICTION

The most significant application for the streamflow prediction is the forecasting of the flood. As part III-A mentioned, Wuhan was the city most affected by the flood of the Yangtze River. In this section, the data from January 1952 to December 1997 (2015) was used to train the model, and the complete 12-month runoff data in 1998 (2016) was the validation set which measures the performance of the model. Fig. 8, showed the whole predictions of 1998 and 2016.

In the flood season of 1998, the Yangtze River basin experienced a catastrophic flood. The peak streamflow at the Hankou Station reached the 70000 m^3 which was 45% higher than the flow in the normal year. The prediction of flood peak values during the flood season is very close to the actual value, but the prediction of the extreme value is one month ahead of the actual.

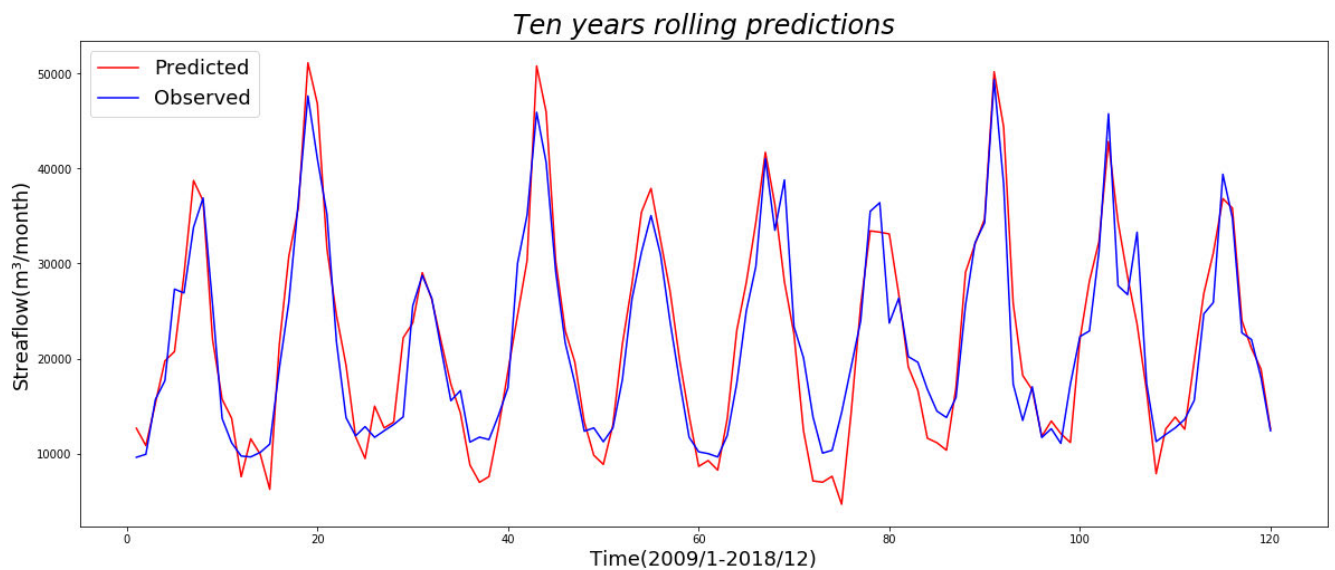


FIGURE 6. Observed and predicted data of ten years rolling prediction from 2009/1 to 2018/12.

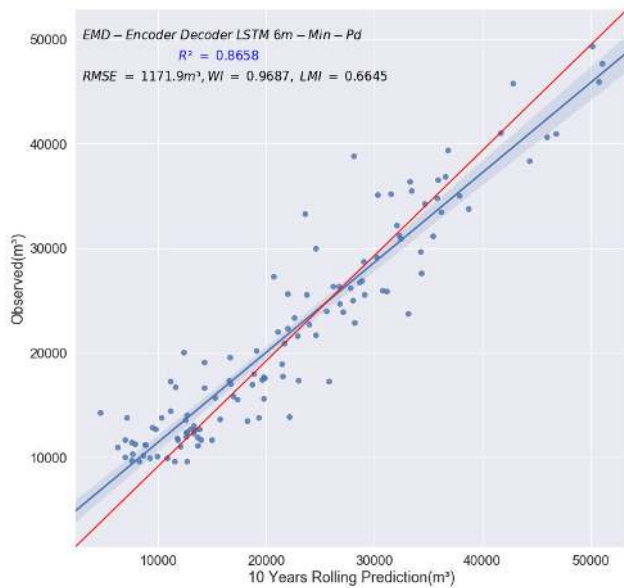


FIGURE 7. Validation scatter plot of the rolling prediction.

For the results in 2016, the prediction from January to August is extremely precise, especially in flood season, the prediction of the flood season are nearly the same as the actual ones. The R^2 of both two years are higher than 0.8 and the LMIs are close to 0.65. Figure 9 is the validation scatter plots for the predictions. The blue area is the confidence interval and the red line is the centerline which formed $y = x$ (observed = predicted). Also, the prediction in 2016 is more precise than forecasting in 1998, From figure 8 we can see that the observed peak discharge in 1998 is higher about 40% than prediction in 2016. In figure 9-A, the confidence interval in 1998 predictions is larger than the 2016 predictions, owing to the historical peak value. From the part B-ten years rolling prediction, we can conclude the forecasting in normal year is closer to the actual value than the prediction in abnormal flood year. Thus, abnormal maxima like peak discharge in 1998 is hard to predict for the current prediction model. Nearly all of the points are close to the regression line,

which means the power of flood forecasting of the proposed model is acceptable.

V. COMPARISONS AND ANALYSIS

The structure of the proposed model combines EMD and En-De-LSTM with 6-month Min-Pd. In order to get the model with the best performance and effectiveness, the comparisons among the different models and different months with Min-Pd were made by the 12-month streamflow data in 2018. Firstly, the proposed approach is compared between the Vanilla LSTM, Stacked LSTM, En-De-LSTM, and these models combined with EMD. From Vanilla LSTM to En-De-LSTM, the structure of the neural network become more and more complex. The amounts of the layers gradually increase. Making comparisons among these different LSTMs, the influence of the number of layers could be concluded from the results. The structure with the best performance can be found by this process. Besides, the reliability and the robustness of the final model can be strengthened. Then, the comparisons of different months-Min-Pd with the above models were made. The different minimum prediction period influenced the way of training dataset on how to be trained Through all training process, for the first time, the neural network read the first-round data and predict the next round data recurrently. The 6m-Min-Pd, 12m-Min-Pd, 18m-Min-Pd, and 24m-Min-Pd were chosen to find the best Min-Pd. The 12m-Min-Pd use the data of last year to predict the data in the next year. In this model the complete annual flood fluctuation process can be kept and could learn the trends of flood more efficiently. Similarly, the 6m-Min-Pd, 18m-Min-Pd and 24m-Min-Pd learn the flood trends every half year, every 18 months and every two year separately. As for the other possible alternatives, the period of them is too long, which will make the major trends less intuitive and make the training process shorter and less accurate. All of choices used in the paper, they have retained the complete annual flood fluctuation process, or a single fluctuation process.

- (1) Figure 10 shows the predictions among different LSTM structures with the 6m-Min-Pd. For each model,

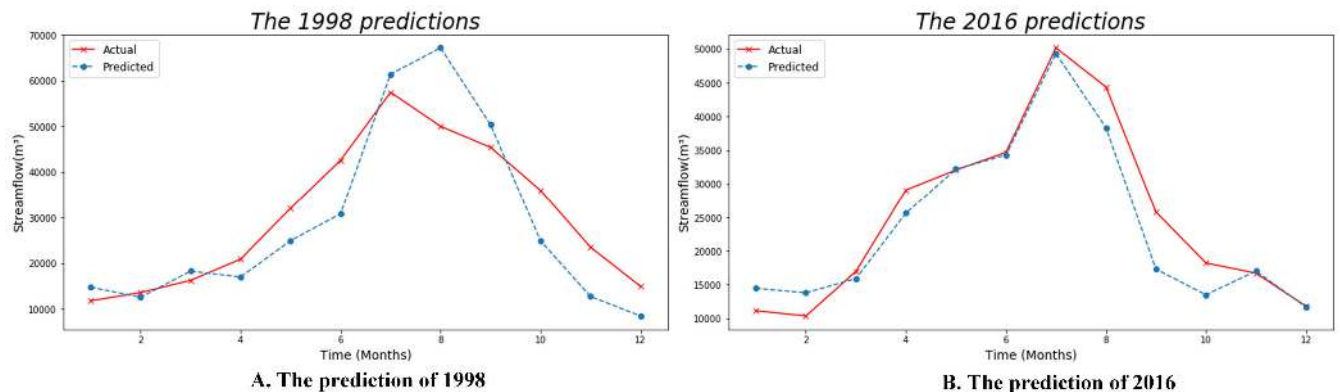


FIGURE 8. Forecasted and Actual streamflow in 1998 and 2016.

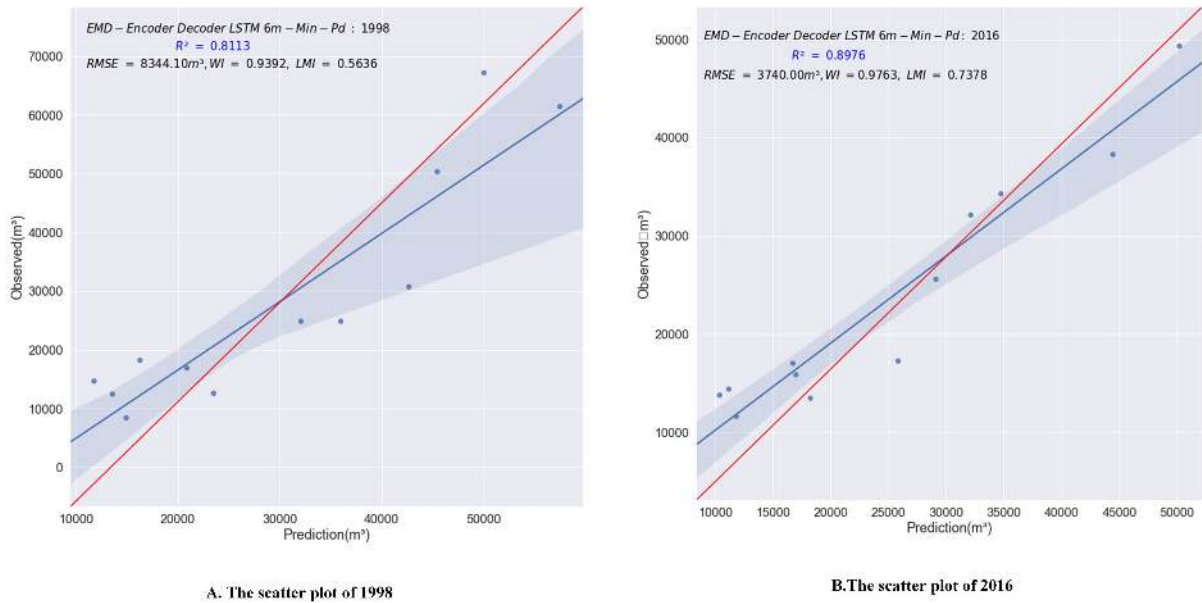


FIGURE 9. The validation scatter plot of 1998 and 2016.

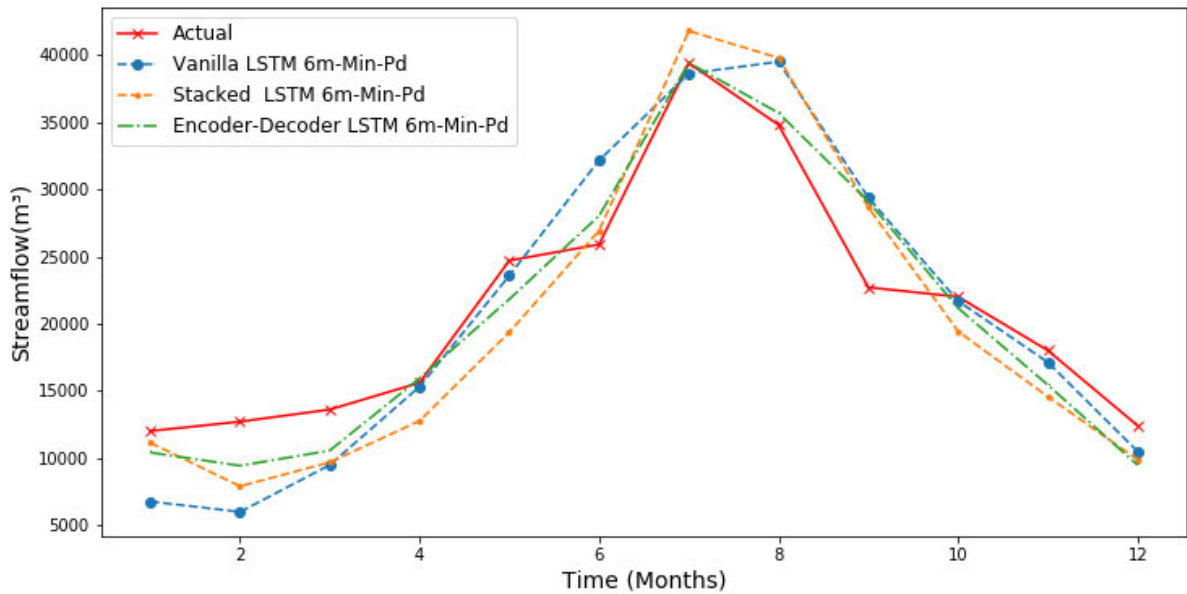


FIGURE 10. The comparison among the different LSTM model with 6m-Min-Pd.

the first 6-month streamflow data is the input and the following 6-month data is the output, as mentioned in part IV-B. The validation set was the complete 12-month dataset in 2018. We can see that the predictions in the second half-year of 2018 are better than the forecasting results in the first half-year of 2018. As the layers adding from the Vanilla LSTM (a-layer LSTM) to the Stacked LSTM (multi-layer LSTM) and finally to the En-De-LSTM, the goodness-of-fit is increasing gradually.

Fig. 11 and Table 1 show the validation scatter plots and statistical criteria of the different LSTM models separately. The En-De-LSTM model has the best performance among three models. The RMSE decreased by 30%, and the LMI improved by nearly 20%. Also, the regression line is the closest to the center as the model complexifies.

(2) Figure 12 shows the predictions among the different LSTM model with 12m-Min-Pd. Figure 13 and Table 2 showed the validation scatter plots and criteria among the three models. For each model, the first

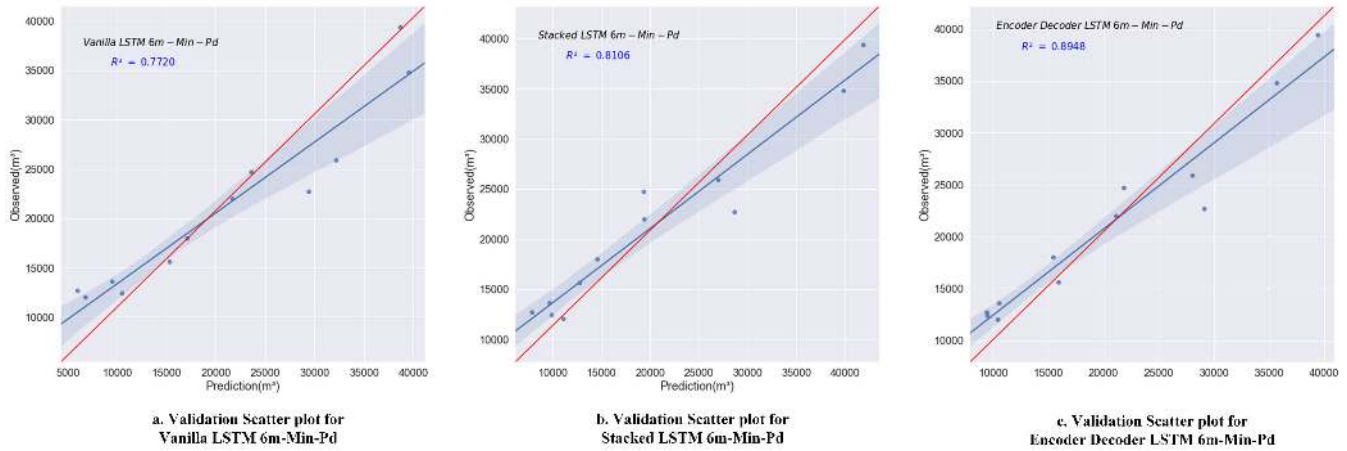


FIGURE 11. Validation scatter plots for different LSTM models with 6m-Min-Pd.

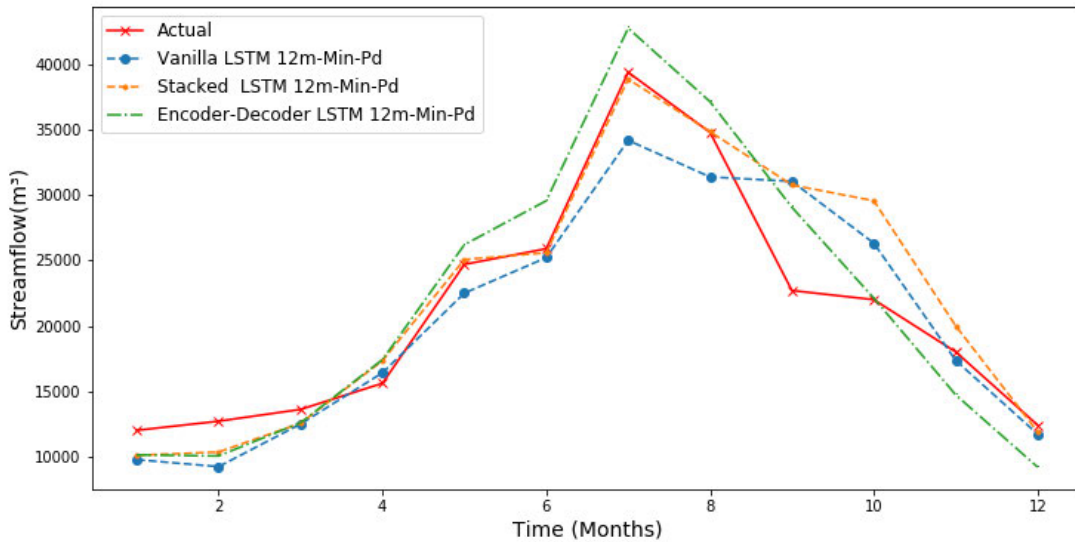


FIGURE 12. The comparison among the different LSTM model with 12m-Min-Pd.

TABLE 1. Comparison of criteria in different LSTM model with 6m-Min-Pd.

| | WI | LMI | RMSE(m ³) | R2 |
|---------------------|---------------|--------|-----------------------|--------|
| Vanilla LSTM | 0.9572 | 0.5417 | 4105.0003 | 0.7720 |
| Stacked LSTM | 0.9641 | 0.5220 | 3741.6487 | 0.8106 |
| En-De-LSTM | 0.9773 | 0.6838 | 2789.2669 | 0.8948 |

12-month streamflow data is the input and the next 12-month data is the output. Subsequently, the forecasts are recurrently made, and the validation set is still the streamflow of 12-month in 2018. As the number of layers of LSTM and the En-De part were added, the accuracy of the model improved as discussed above. The overall accuracy is still relatively high, and the forecast accuracy in the first half of the year is higher

than in the second half. Compared with the accuracy of data from August to December has slightly decreased. The best of WI is the En-De-LSTM, the best of LMI is stacked LSTM and the best of the R2 is also the En-De-LSTM. It can be concluded that the 12m-Min-Pd has some shortage when compared to the 6m-Min-Pd.

- Figure 14-15 depict the comparisons among the different LSTM models with 18m-Min-Pd and the validation scatter plots of these methods. Table 3 contains the values of error among them. In this section, the input is the 18-month streamflow data and output is the next 18-month data, and these models will be trained recurrently. Finally, the output is the last 18-month data from June 2017 to December 2018. We took the whole data of 2018 to validate the model. The 18m-Min-Pd predictions are obviously less accurate than the two situations above. From figure 14, the best of these models

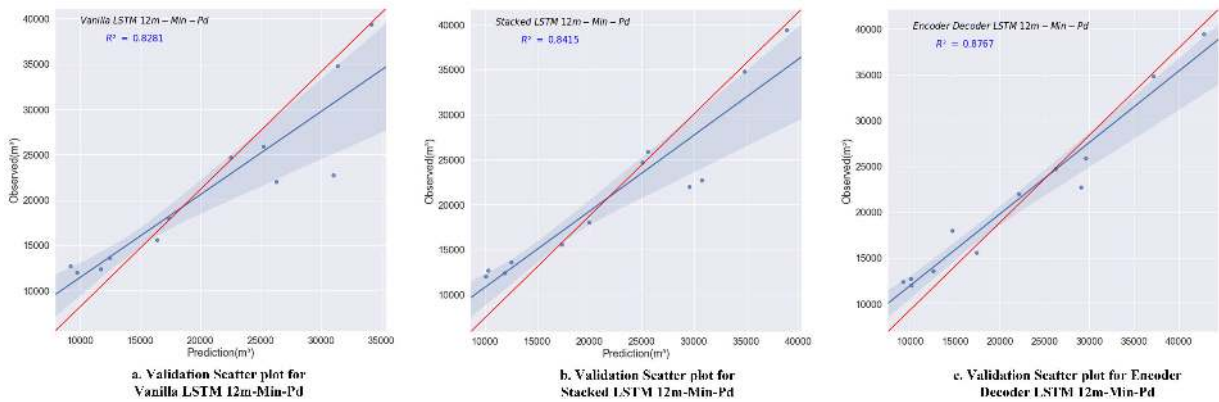


FIGURE 13. Validation scatter plots for different LSTM models with 12m-Min-Pd.

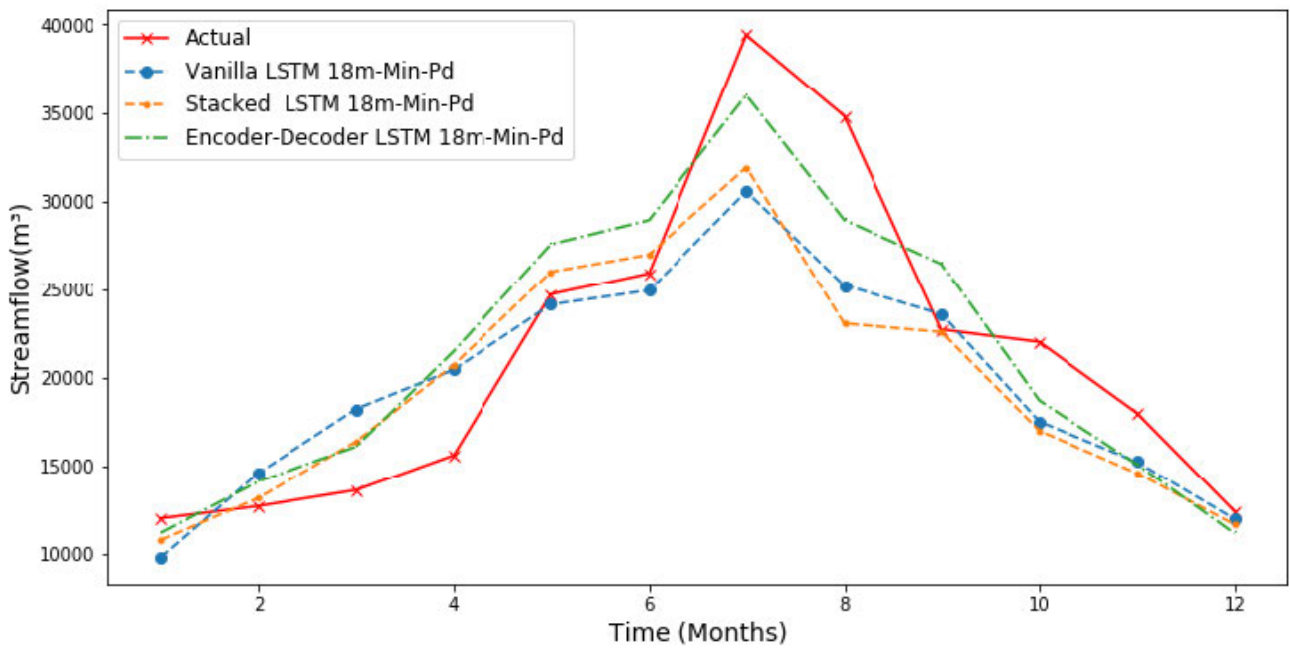


FIGURE 14. The comparison among the different LSTM model with 18m-Min-Pd.

TABLE 2. Comparison of criteria in different LSTM model with 12m-Min-Pd.

| | WI | LMI | RMSE(m ³) | R ² |
|--------------|--------|--------|-----------------------|----------------|
| Vanilla LSTM | 0.9552 | 0.6107 | 3564.7153 | 0.8281 |
| Stacked LSTM | 0.9636 | 0.6894 | 3423.1601 | 0.8415 |
| En-De-LSTM | 0.9755 | 0.6331 | 3018.4774 | 0.8767 |

still cannot reach high performance. Only the data forecasted in January, June, September, and December are close to the observed data. In the flow prediction study, what we considered is not the standalone values, but the whole trends of the predictions. Hence, LMIs only

reach nearly 0.5. The R² values are lower than all of the above models, and the best WIs are lower than 0.9. Also, the data points on the validation plot show more scattered spread. It can be concluded from the criteria and the scatter plot that the 18m-Min-Pd is not a suitable choice for the streamflow predictions. The main reason for this is because the whole flood season from June to September is separated while splitting the dataset. The neural network cannot learn the effective trends of the advent and leaving from the data during the model building and training.

(4) Figure 16-17 illustrates the results and scatter plots for three different LSTM models with 24m-Min-Pd, and Table 4 shows the statistical criteria among these approaches. Similar to the 18m-Min-Pd, the Min-Pd of

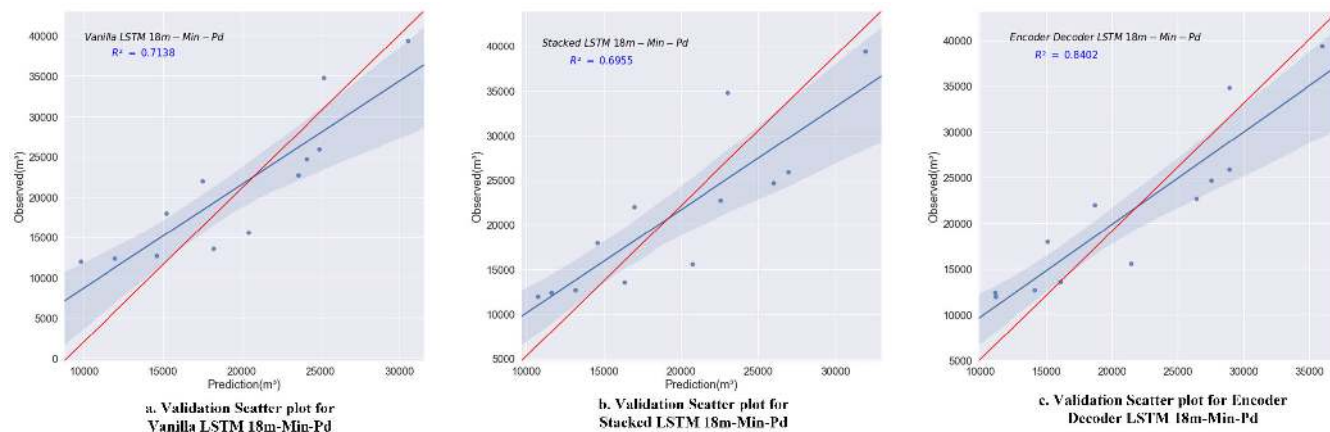


FIGURE 15. Validation scatter plots for different LSTM models with 18m-Min-Pd.

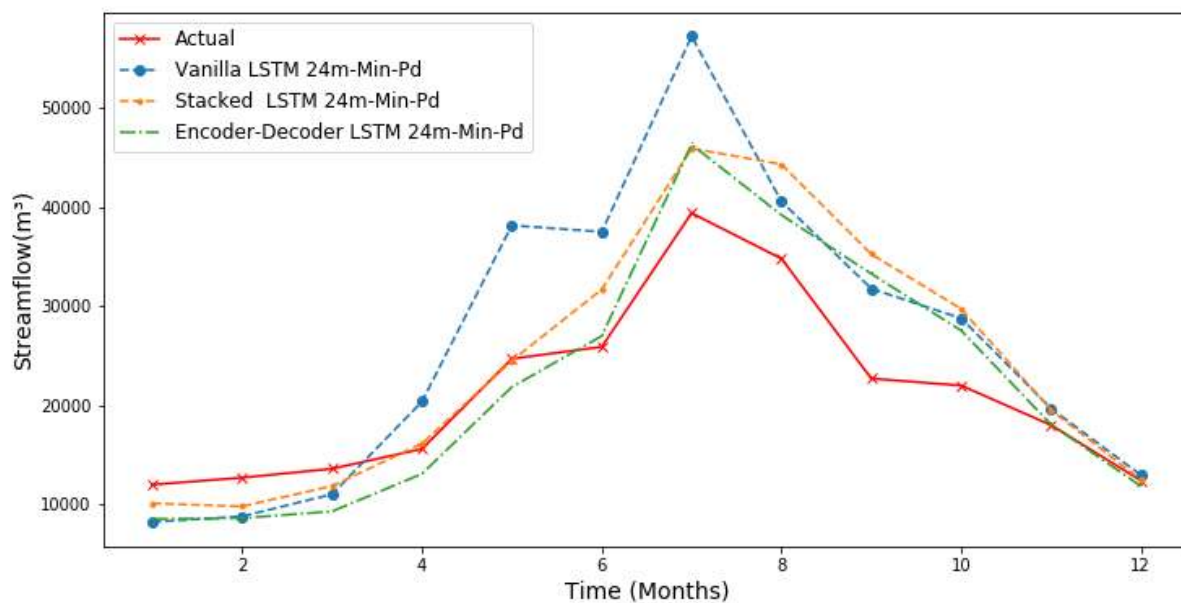


FIGURE 16. The comparison among the different LSTM model with 24m-Min-Pd.

TABLE 3. Comparison of criteria in different LSTM model with 18m-Min-Pd.

| | WI | LMI | RMSE(m ³) | R ² |
|--------------|--------|--------|-----------------------|----------------|
| Vanilla LSTM | 0.8954 | 0.5052 | 4599.5510 | 0.7138 |
| Stacked LSTM | 0.8947 | 0.5247 | 4744.1984 | 0.6955 |
| En-De-LSTM | 0.9543 | 0.5661 | 3436.9514 | 0.8402 |

the input and output is set to 24-months and the model was trained subsequently. From Figure 16, all three predicted flood peak values are higher than the actual values. However, their trends are somewhat similar to the observed trends except for the Vanilla LSTM. One-third of the points are far from the regression line

and the R² value is lower than 0.7. All criteria are smaller than 12m-Min-Pd and 6m-Min-Pd, especially the LMIs are less than 0.5. and the minima R² and LMI are less than 0.1.

Among all the above-mentioned circumstances (Table 1-4) we can simply conclude that the En-De-LSTM has the best performance It has the lowest RMSE = 2789.2669m³ and the highest WI = 0.9773, LMI = 0.6838 and R² = 0.8948. While the 12m-Min-Pd has the closest value of which WI = 0.9755, LMI = 0.5661, RMSE = 3018.47 m³ and R² = 0.8767. In other words, this model can predict the streamflow precisely. Furthermore, we still need to determine which value of months-Min-Pd is the best between the 6m-Min-Pd and 12m-Min-Pd.

The first two models were compared to the last several models. The first model is the En-De-LSTM which

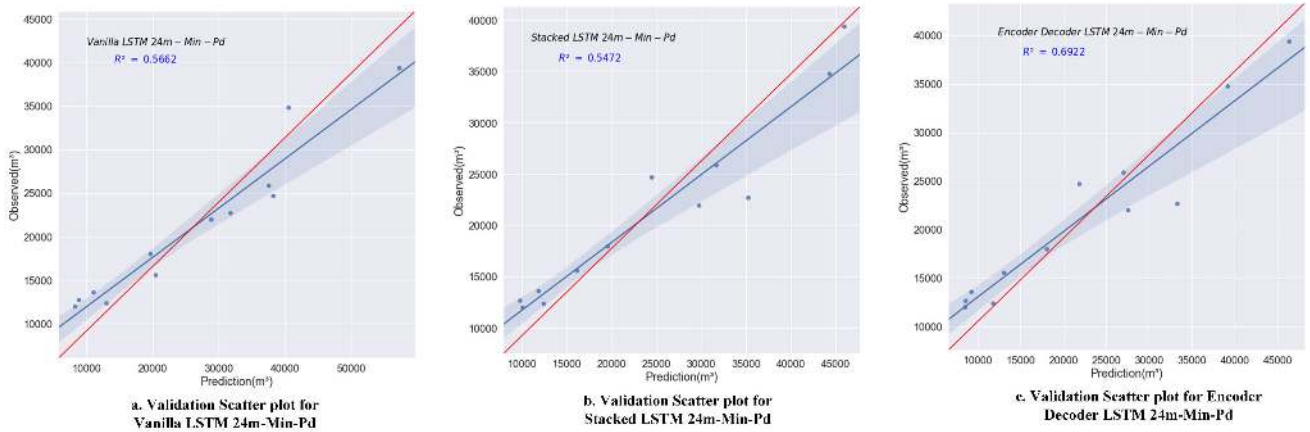


FIGURE 17. Validation scatter plots for different LSTM models 24m-Min-Pd.

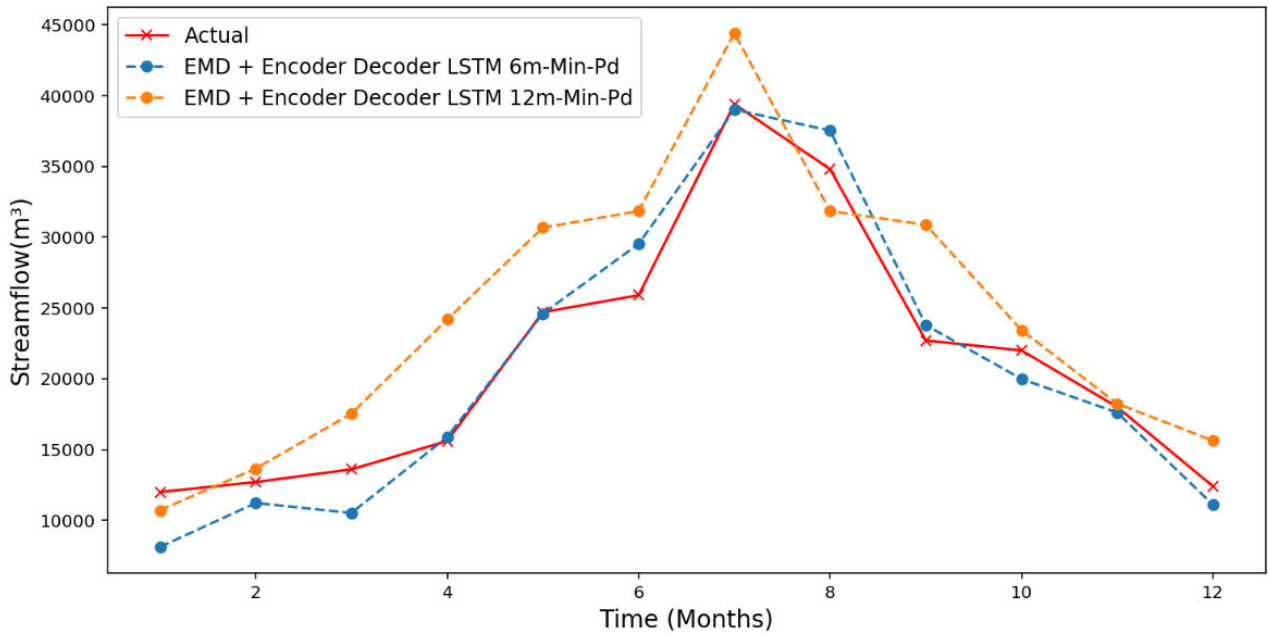


FIGURE 18. The comparison between 6m-Min-Pd and 12m-Min-Pd in EMD-En-De-LSTM.

TABLE 4. Comparison of criteria in different LSTM model with 24m-Min-Pd.

| | WI | LMI | RMSE(m ³) | R ² |
|--------------|--------|--------|-----------------------|----------------|
| Vanilla LSTM | 0.8723 | 0.0403 | 8452.7873 | 0.0335 |
| Stacked LSTM | 0.9248 | 0.4018 | 5785.2988 | 0.5472 |
| En-De-LSTM | 0.9465 | 0.4556 | 4770.0359 | 0.6922 |

combines EMD with 6m-Min-Pd. The second model is the En-De-LSTM with 12m-Min-Pd which combines EMD and. EMD combines all other LSTM models in order to be compared with the first two models mentioned. The comparison

of the first two models (6m-Min-Pd and 12m-Min-Pd) is necessary to choose the best performance model. The following comparison of the two models with the other several LSTM models is necessary in order to enhance the reliability of the study.

(5) Figure 18 illustrates the predictions of En-De-LSTM with 6m-Min-Pd and 12m-Min-Pd. Figure 19 depicts the validation scatter plot for the above two models.

From Figure 18, the whole trend of the 6m-Min-Pd is closer than the 12m-Min-Pd. In Figure 19, the points in 6m-Min-Pd are more gathered and closer to the centerline while the points in 12m-Min-Pd are more scattered and several points are out of the confidence interval. It can be drawn from the figures that the EMD-En-De-LSTM with 6m-Min-Pd has the best performance throughout all of the models in this study.

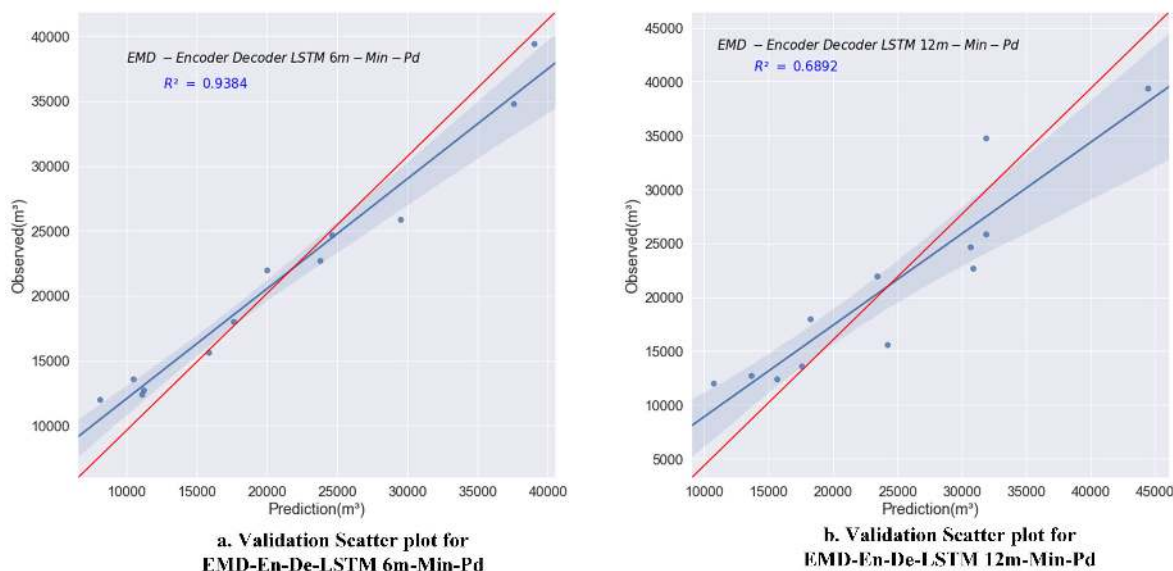


FIGURE 19. Validation scatter plots for EMD-En-De-LSTM with 6m-Min-Pd and 12m-Min-Pd.

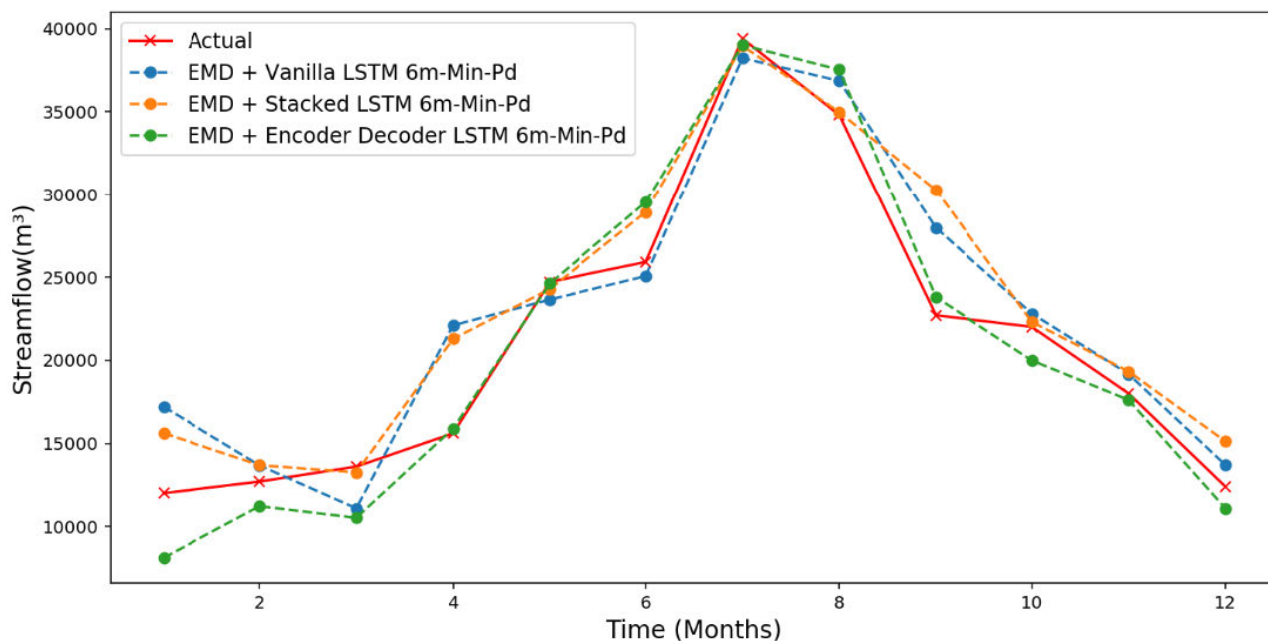


FIGURE 20. The comparison among Vanilla, Stacked and En-De LSTM combined with EMD.

(6) Comparisons among the EMD-En-De-LSTM, vanilla LSTM and stacked LSTM combine EMD with 6m-Min-Pd. These comparisons are depicted in Figure 20, and the validation scatter plot of EMD-vanilla LSTM and EMD-Stacked LSTM are illustrated in Figure 20-21. Plus, Table 5 represents the statistical criteria in different EMD-LSTM structures.

According to the Tables 1-5 and Figures 20-21, it can be concluded that EMD-En-De-LSTM with 6m-Min-Pd model has the best performance across all models. After being

combined with the EMD model, the results of the Vanilla LSTM and Stacked LSTM models also improved greatly. Their RMSE values are close to 3000 m³ and the LMIs are close to 0.65. This reflects to an extent that EMD's optimization of the LSTM model is very successful. For the EMD-En-De-LSTM model, the LMI is the only one higher than 0.75 and its RMSE is the only one near to 2000 m³, which indicates the perfect goodness-of-fit of this model. When compared with other 6m-Min-Pd methods, it contains more gathered and closer points. Nearly half of the points

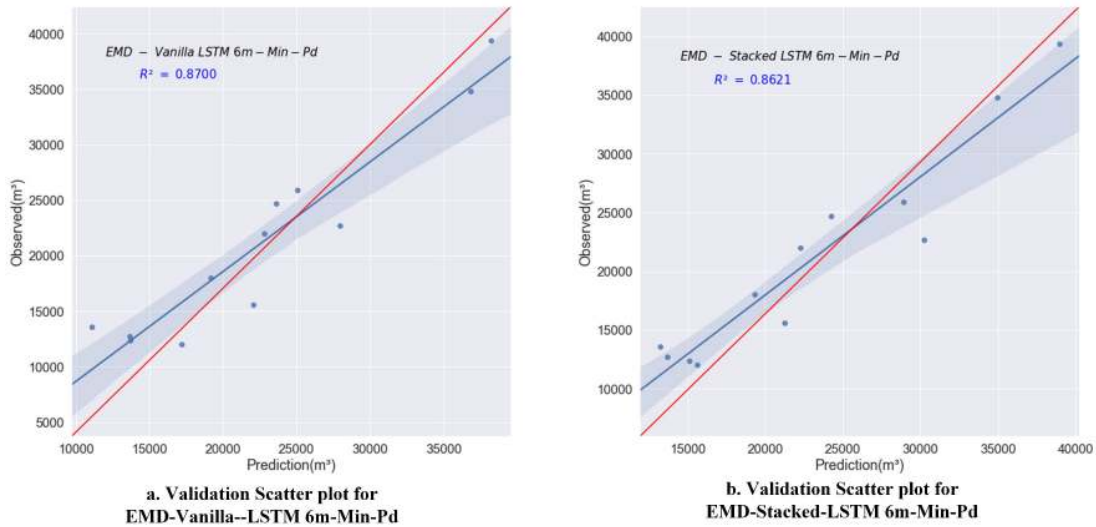


FIGURE 21. Validation scatter plots for EMD-Vanilla LSTM and Stacked LSTM with 6m-Min-Pd.

TABLE 5. Accuracy evaluation of different LSTM models combines with EMD.

| Models | EMD-Vanilla LSTM | EMD-Stacked LSTM | EMD-En-De-LSTM | |
|-----------------------|------------------|------------------|------------------|-----------|
| | 6m | 6m | 6m | 12m |
| <i>WI</i> | 0.9657 | 0.9636 | 0.9868 | 0.93 |
| <i>LMI</i> | 0.6614 | 0.688 | 0.7611 | 0.4403 |
| RMSE(m ³) | 3099.9039 | 3192.4513 | 2133.7488 | 4793.5363 |
| <i>R</i> ² | 0.87 | 0.8621 | 0.9384 | 0.6892 |

are out of the confidence interval in the scatter plot for EMD-Vanilla-LSTM and EMD-Stacked-LSTM.

VI. CONCLUSION

In this paper, a novel hybrid Encoder Decoder LSTM model based on EMD with 6m-Min-Pd was proposed to forecast the monthly streamflow for the Yangtze River. Two predictions were made from two aspects.

One is the flood prediction, in other words, the flood season prediction. In this part, the whole data in 1998 and 2016 was used to validate the performance of the model. The study area was suffered from floods in these two years. Final results of the flood prediction were acceptable, *R*² values in both years were higher than 0.8, and the values of *LMI* were close to 0.65, which means the models could be used in flood forecasting, and the accuracy will reach the 70% at least.

The other prediction is long-term forecasting. In this paper, the most recent 10-year (from 2009 to 2018) prediction was made. *WI* *LMI* and *R*² values of this experiment showed perfect reliability and goodness-of-fit. Moreover, the RMSE was only 1171.9 m³. More importantly, the local maxima in each year, which also is the flood peak discharge, were

very precise to the observed value or higher. This will make accurate yearly water allocation, timely flood disaster relief, and economic development possible.

To improve the reliability and accuracy, several comparisons were made from the different LSTM models and different month Min-Pd parameters. It can be concluded that:

1. The added layers of LSTM could improve performance (comparison between the Vanilla LSTM and Stacked LSTM). When adding the Encoder Decoder technique, the results can be further enhanced.
2. Combining the EMD algorithm with the LSTM model can produce a better result (comparison between the EMD-LSTM model and single LSTM model).
3. The 6-month-Min-Pd has the best performance through all other months-Min-Pd parameters.

The 18-Min-Pd and 24m-Min-Pd will split the complete period of flooding, which could make the deep neural network unable to learn valuable and useful information. The 12m-Min-Pd predictions contain a whole process of flood from the minima to the maxima then back to the minima. These courses of events and the results indicate that the deep learning model has some difficulties and drawbacks in such a period.

For the decision makers in water resource management, the major motivation of this paper is to provide a continuous longtime prediction and large flood forecast. When deep neural networks used in time series forecasting, we should concentrate more on the longtime trends. This is meaningful for hydrology applications like water allocation. Besides, the forecasting ability of large floods in this article is reflected by the accurate prediction of historical large floods. However, in the real world, we don't know whether the flood will occur in the next year or not. If the peak discharge prediction from the model is much higher than any other recent years' and the confidence interval of this prediction is about 80%,

this will be a warning for water resource management. This model is not giving a simple binary flood prediction, but just some prediction about the streamflow in the next few years. The catastrophic flood may come or not, but there is no harm in taking preventive measures in advance.

There is still some limitation of this study. Only one hydrological station dataset was used, the streamflow data from several hydrological stations on the Yangtze River could be processed with the spatial homogeneity of stations and considered the influence of other physical conditions to tune the model. Also, the model did not include the other hydrological variables including rainfalls, temperature, and evapotranspiration data, they could be used to build and train the model in future research. As for the deep neural network, the random search method could be used to search the best hyperparameter and make comparisons with the grid search method.

REFERENCES

- [1] S. Zhu, J. Zhou, L. Ye, and C. Meng, "Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of yangtze river, China," *Environ. Earth Sci.*, vol. 75, no. 6, p. 531, Mar. 2016.
- [2] F. Huang, Z. Xia, N. Zhang, Y. Zhang, and J. Li, "Flow-complexity analysis of the upper reaches of the yangtze river, China," *J. Hydrologic Eng.*, vol. 16, no. 11, pp. 914–919, Nov. 2011.
- [3] N. Noori and L. Kalin, "Coupling SWAT and ANN models for enhanced daily streamflow prediction," *J. Hydrol.*, vol. 533, pp. 141–151, Feb. 2016.
- [4] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, Feb. 2019.
- [5] W. Collischonn, R. Haas, I. Andreolli, and C. E. M. Tucci, "Forecasting river uruguay flow using rainfall forecasts from a regional weather-prediction model," *J. Hydrol.*, vol. 305, nos. 1–4, pp. 87–98, Apr. 2005.
- [6] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models Part I—A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, Apr. 1970.
- [7] F. Fotovatikhah, M. Herrera, S. Shamsirband, K.-W. Chau, S. F. Ardabili, and M. J. Piran, "Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 411–437, Jan. 2018.
- [8] W.-C. Wang, K.-W. Chau, D.-M. Xu, L. Qiu, and C.-C. Liu, "The annual maximum flood peak discharge forecasting using Hermite projection pursuit regression with SSO and LS method," *Water Resour. Manage.*, vol. 31, no. 1, pp. 461–477, Jan. 2017.
- [9] Y. Sang, L. Shang, Z. Wang, C. Liu, and M. Yang, "Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting," *Chin. Sci. Bull.*, vol. 58, no. 31, pp. 3796–3805, Nov. 2013.
- [10] Z. Abdulelah Al-Sudani, S. Q. Salih, A. Sharafati, and Z. M. Yaseen, "Development of multivariate adaptive regression spline integrated with differential evolution model for streamflow simulation," *J. Hydrol.*, vol. 573, pp. 1–12, Jun. 2019.
- [11] C. Ma and Y. Li, "Improving forecasting accuracy of annual runoff time series using RBFN based on EEMD decomposition," in *Proc. DEStech Trans. Eng. Technol. Res. (EMME)*, Lancaster, PA, USA, 2016, pp. 211–216.
- [12] Y. H. Yu Guorong, X. Ziqiang, and Z. Xiaoyong, "Application of projection pursuit auto regression model in predicting runoff of Yangtze River," *J. Hohai Univ. (Natural Sci.)*, vol. 37, 3, pp. 263–266, May 2009.
- [13] H. Badrzadeh, R. Sarukkalige, and A. W. Jayawardena, "Improving ann-based short-term and long-term seasonal river flow forecasting with signal processing techniques," *River Res. Appl.*, vol. 32, no. 3, pp. 245–256, Mar. 2016.
- [14] M. Rezaie-Balf, S. Kim, H. Fallah, and S. Alaghmand, "Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in iran and South Korea," *J. Hydrol.*, vol. 572, pp. 470–485, May 2019.
- [15] M. Fu, T. Fan, Z. Ding, S. Q. Salih, N. Al-Ansari, and Z. M. Yaseen, "Deep learning data-intelligence model based on adjusted forecasting window scale: Application in daily streamflow simulation," *IEEE Access*, vol. 8, pp. 32632–32651, 2020.
- [16] Z. M. Yaseen, S. R. Naganna, Z. Sa'adi, P. Samui, M. A. Ghorbani, S. Q. Salih, and S. Shahid, "Hourly river flow forecasting: Application of emotional neural network versus multiple machine learning paradigms," *Water Resour. Manage.*, vol. 34, no. 3, pp. 1075–1091, Feb. 2020.
- [17] S. Yin, D. Tang, X. Jin, W. Chen, and N. Pu, "A combined rotated general regression neural network method for river flow forecasting," *Hydrolog. Sci. J.*, vol. 61, no. 4, pp. 669–682, Mar. 2016.
- [18] T. S. Hu, K. C. Lam, and S. T. Ng, "A modified neural network for improving river flow prediction/un réseau de neurones modifié pour améliorer la prévision de l'écoulement fluvial," *Hydrolog. Sci. J.*, vol. 50, no. 2, pp. 299–318, Apr. 2005.
- [19] A. M. Kalteh, "Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform," *Comput. Geosci.*, vol. 54, pp. 1–8, Apr. 2013.
- [20] M. Khashei and M. Bijari, "An artificial neural network (p,d,q) model for timeseries forecasting," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479–489, Jan. 2010.
- [21] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, p. 1536, 2018.
- [22] S. Shamsirband, S. Hashemi, H. Salimi, S. Samadianfard, E. Asadi, S. Shadkani, K. Kargar, A. Mosavi, N. Nabipour, and K.-W. Chau, "Predicting standardized streamflow index for hydrological drought using machine learning models," *Eng. Appl. Comput. Fluid Mech.*, vol. 14, no. 1, pp. 339–350, Jan. 2020.
- [23] J. Guo, J. Zhou, H. Qin, Q. Zou, and Q. Li, "Monthly streamflow forecasting based on improved support vector machine model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13073–13081, Sep. 2011.
- [24] W.-C. Wang, D.-M. Xu, K.-W. Chau, and S. Chen, "Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD," *J. Hydroinformatics*, vol. 15, no. 4, pp. 1377–1390, Oct. 2013.
- [25] S. Huang, J. Chang, Q. Huang, and Y. Chen, "Monthly streamflow prediction using modified EMD-based support vector machine," *J. Hydrol.*, vol. 511, pp. 764–775, Apr. 2014.
- [26] E. Meng, S. Huang, Q. Huang, W. Fang, L. Wu, and L. Wang, "A robust method for non-stationary streamflow prediction based on improved EMD-SVM model," *J. Hydrol.*, vol. 568, pp. 462–478, Jan. 2019.
- [27] M. Z. Moshfegh, "Analysis and prediction of river flow based on evolutionary programming and artificial intelligence," *Soil Struct. Interact. J.*, vol. 3, pp. 12–22, Dec. 2018.
- [28] R. Taormina and K.-W. Chau, "ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS," *Eng. Appl. Artif. Intell.*, vol. 45, pp. 429–440, Oct. 2015.
- [29] H. Yoon, S.-C. Jun, Y. Hyun, G.-O. Bae, and K.-K. Lee, "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer," *J. Hydrol.*, vol. 396, nos. 1–2, pp. 128–138, Jan. 2011.
- [30] R. Noori, A. Khakpour, B. Omidvar, and A. Farokhnia, "Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5856–5862, Aug. 2010.
- [31] Z. M. Yaseen, W. H. M. W. Mohtar, A. M. S. Ameen, I. Ebtehaj, S. F. M. Razali, H. Bonakdari, S. Q. Salih, N. Al-Ansari, and S. Shahid, "Implementation of univariate paradigm for streamflow simulation using hybrid data-driven model: Case study in tropical region," *IEEE Access*, vol. 7, pp. 74471–74481, 2019.
- [32] M. Y. A. Khan, F. Hasan, S. Panwar, and G. J. Chakrapani, "Neural network model for discharge and water-level prediction for ramganga river catchment of ganga basin, india," *Hydrolog. Sci. J.*, vol. 61, no. 11, pp. 2084–2095, Aug. 2016.
- [33] M. C. Demirel, A. Venancio, and E. Kahya, "Flow forecast by SWAT model and ANN in pracana basin, portugal," *Adv. Eng. Softw.*, vol. 40, no. 7, pp. 467–473, Jul. 2009.
- [34] S. J. Hadi, S. I. Abba, S. S. Sammen, S. Q. Salih, N. Al-Ansari, and Z. M. Yaseen, "Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation," *IEEE Access*, vol. 7, pp. 141533–141548, 2019.
- [35] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, 2017.

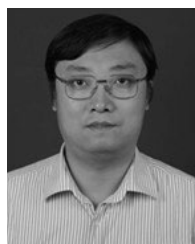
- [36] J. Zhang, Y. Zhu, X. Zhang, M. Ye, and J. Yang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas," *J. Hydrol.*, vol. 561, pp. 918–929, Jun. 2018.
- [37] T. Li, B. Wang, M. Zhou, and J. Watada, "Short-term load forecasting using optimized LSTM networks based on EMD," 2018, *arXiv:1809.10108*. [Online]. Available: <http://arxiv.org/abs/1809.10108>
- [38] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [39] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: <http://arxiv.org/abs/1409.0473>
- [40] J. Lu, C. Xiong, D. Parikh, and R. Socher, "Knowing when to look: Adaptive attention via a visual sentinel for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 375–383.
- [41] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 4960–4964.
- [42] N. E. Huang, Z. Shen, and S. R. Long, "A new view of nonlinear water waves: The Hilbert spectrum," *Annu. Rev. Fluid Mech.*, vol. 31, no. 1, pp. 417–457, Jan. 1999.
- [43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [44] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network language model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 5528–5531.
- [45] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, Apr. 1998.
- [46] J. Chorowski and N. Jaitly, "Towards better decoding and language model integration in sequence to sequence models," 2016, *arXiv:1612.02695*. [Online]. Available: <http://arxiv.org/abs/1612.02695>
- [47] D. Chicco, "Ten quick tips for machine learning in computational biology," *BioData Mining*, vol. 10, no. 1, p. 35, Dec. 2017.
- [48] X. Mei, Z. Dai, P. H. A. J. M. van Gelder, and J. Gao, "Linking three gorges dam and downstream hydrological regimes along the yangtze river, China," *Earth Space Sci.*, vol. 2, no. 4, pp. 94–106, Apr. 2015.
- [49] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [50] J. H. Torrie, *Principles and Procedures of Statistics: With Special Reference to the Biological Sciences*. New York, NY, USA: McGraw-Hill, 1960.
- [51] C. J. Willmott, "On the evaluation of model performance in physical geography," in *Spatial Statistics and Models*, vol. 40. Dordrecht, The Netherlands: Springer, 1984, pp. 443–460. [Online]. Available: https://link.springer.com/chapter/10.1007/978-94-017-3048-8_23
- [52] C. J. Willmott, S. M. Robeson, and K. Matsuura, "A refined index of model performance," *Int. J. Climatol.*, vol. 32, no. 13, pp. 2088–2094, Nov. 2012.
- [53] D. R. Legates and G. J. McCabe, "Evaluating the use of 'goodness-of-fit' Measures hydrologic hydroclimatic model validation," *Water Resour. Res.*, vol. 35, no. 1, pp. 233–241, Jan. 1999.
- [54] S. Kim, Y. Seo, M. Rezaie-Balf, O. Kisi, M. A. Ghorbani, and V. P. Singh, "Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: Peninsula vs continent," *Theor. Appl. Climatol.*, vol. 137, nos. 1–2, pp. 693–712, Jul. 2019.



DARONG LIU received the B.S. degree in engineering from the China University of Geosciences, Wuhan, China, where he is currently pursuing the Ph.D. degree. During the undergraduate study, his research interests include computer science, three-dimensional modeling, and numerical simulation in geoscience. From 2019, his research turns to the numerical simulation in physical oceanography, machine learning, and neural networks.



WENCHAO JIANG received the B.S. degree in engineering from the China University of Geosciences, Wuhan, China. He is currently pursuing the master's degree with Wuhan University. From 2017, his research interests include neural networks and evolutionary computing.



LIN MU received the B.S., M.S., and Ph.D. degrees in physical oceanography from the Ocean University of China, Qingdao, China, in 2000, 2002, and 2007, respectively. He is currently a Professor of physical oceanography with Shenzhen University and Shenzhen Research Institute, China University of Geosciences, Guangdong, China. He has authored or coauthored over 20 scientific articles and four books. His research interests include physical oceanography: prevention and

mitigation of marine disasters, maritime search and rescue, and emergency response management of offshore oil spills.



SI WANG received the M.S. degree in computer science from Wuhan Polytechnic University, Wuhan, China, in 2016. He is currently pursuing the Ph.D. degree in marine science and technology with the China University of Geosciences, Wuhan. His research interests include signal processing, time-series data analysis, and machine learning.

• • •