

# Prediction of Runoff in Huayuankou of the Yellow River based on Autoformer-LSTM

Feifei Yu

School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou Henan, 450046, China

**Abstract:** To effectively extract information features of runoff time series and improve the accuracy of river runoff prediction methods, we integrated Autoformer with Long Short-Term Memory (LSTM) to develop an Autoformer-LSTM runoff prediction model. We simulated and validated the model using daily runoff, precipitation, and average water level data from January 1, 2002, to December 31, 2022, at Huayuankou Hydrological Station, a key station in the lower Yellow River Basin. And select Autoformer model, Transformer model, and Informer model for comparative prediction experiments. The results show that the Autoformer-LSTM and Autoformer models outperformed the Transformer and Informer models in predictive accuracy. Furthermore, compared to the Autoformer model, the Autoformer-LSTM model reduced Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) by 13.1%, 78.7%, and 53.9%. This prediction method can fully explore the inherent laws of hydrological time series data, effectively utilize the temporal nature of hydrological information, and improve prediction accuracy.

**Keywords:** Time Series Prediction; Autoformer; LSTM; Huayuankou Hydrological Station.

## 1. Introduction

As one of the core issues in the field of hydrological prediction, the improvement of runoff prediction accuracy is very important for water resources management, optimal allocation of water resources and drought and flood risk warning. The formation mechanism of runoff is affected by environmental topography, climate change and human activities, showing significant nonlinear and non-stationary characteristics, which makes the study of runoff time series face complex and fuzzy uncertainty[1].

In order to improve the accuracy of runoff prediction, domestic and foreign scholars have proposed a variety of prediction models, which can be broadly categorized into process-driven and data-driven models based on their construction bases and core logic[2]. The process-driven model relies on detailed hydrological and meteorological data and needs to consider the characteristics of spatial and temporal changes, so it may face challenges such as insufficient generalization ability and uncertainty of parameter setting[3]. However, data-driven models do not rely on hydrological physical mechanisms and directly utilize historical data of runoff and its correlates with the goal of establishing optimal mathematical relationships to build a black-box model of the relationship between inputs and outputs[4]. With the increasing maturity of data science and artificial intelligence technology, data-driven model provides new ideas and methods for runoff prediction with its excellent data processing ability and strong adaptive learning ability. For example, in the same paper[5], the LSTM model was established by using the measured daily runoff of Jingcun hydrological station from 1981 to 2010 as a sample, and the effective prediction of runoff was realized. Based on the historical observation data of Huayuankou Hydrological Station, Wang et al.[3] established a multivariate runoff prediction model of the Yellow River based on TCN-Attention by integrating Attention mechanism and time convolutional neural network. The results show that the Attention mechanism enhances the prediction accuracy of the

TCN model by optimizing the feature weights. Dou et al.[6] applied the PCA-MLP neural network model to the flood season runoff prediction of the Ningxia section of the Yellow River, which simplified the model structure of the neural network while improving the prediction effect. These models have shown good performance in predicting runoff, but the extraction of internal relations of complex runoff time series is not sufficient, and the similarity between runoff subsequences is ignored, resulting in an increase in errors.

In recent years, the Autoformer model has introduced a time series decomposition architecture with autocorrelation mechanism, which can discover and represent dependencies at the sub-series level of the same phase[7], so as to obtain the internal relationship of complex time series more accurately. Long short-term memory (LSTM) network uses multi-gate mechanism to effectively solve the common gradient disappearance problem of traditional recurrent neural network in processing long sequence data, and can effectively capture and utilize various patterns and laws in time series. It has obvious advantages in mining long-term correlation of time series data[8]. Due to the long-time dependence in hydrological data, LSTM can capture and utilize this long-term dependence more effectively than Feed Forward through the gating mechanism, which is helpful to improve the accuracy of runoff prediction. Therefore, this paper proposes a runoff prediction model based on Autoformer-LSTM. The measured data of Huayuankou hydrological station on the Yellow River from January 1, 2002 to December 31, 2022 were selected to carry out simulation verification, and the effectiveness of various runoff prediction models was compared and analyzed by ablation method. The research results have important reference value for optimizing runoff prediction technology system and scientific management of water resources.

## 2. Research Method

### 2.1. Autoformer Model

The Autoformer model adopts the design architecture of

encoder-decoder. The encoder is used to encode historical information, and the decoder generates future long sequence prediction. The two are connected by the input and output of sequence decomposition. The core mechanism sequence decomposition and Auto-correlation mechanism are embedded as independent modules.

(1) Auto-correlation mechanism

The Auto-correlation mechanism consists of two parts, one is the cycle-based dependency discovery, and the other is the time delay aggregation. Based on the theory of random process [9], the autocorrelation coefficient of the discrete time process  $\{x_t\}$  can be obtained by the following formula.

$$R_{xx}(\tau) = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{t=1}^L X_t X_{t-\tau} \quad (1)$$

where  $L$  denotes the length of  $X_t$ , and  $R_{xx}(\tau)$  denotes the time delay similarity between the sequence  $\{X_t\}$  and its n-order lagged sequence  $\{X_{t-\tau}\}$ . The higher the autocorrelation coefficient is, the more likely it is that  $\tau$  is the period of the input, and  $R(\tau)$  can be regarded as the confidence of the unnormalized period estimation[10].

In the single-head autocorrelation mechanism, the sequence  $X$  with length  $L$  is mapped to vectors  $Q$ ,  $K$ , and  $V$ . Then, the period in the input sequence is captured by calculating the autocorrelation coefficient between  $Q$  and  $K$ , and the information of similar subsequences is aggregated based on the correlation between subsequences identified by periodic dependence, so as to realize sequence-level connection. Then, the confidence is normalized. Finally, the sequence of  $V$  is aligned by the  $Roll(\cdot)$  function according to the period, and multiplied by the normalized confidence[7]. The specific calculation process is as follows :

$$\tau_1, \dots, \tau_c = \underset{\tau \in \{1, \dots, L\}}{\text{arg Topc}} \left( R_{Q,K}(\tau) \right) \quad (2)$$

$$\hat{R}_{Q,K}(\tau_1), \dots, \hat{R}_{Q,K}(\tau_c) = \text{soft max} \left( R_{Q,K}(\tau_1), \dots, R_{Q,K}(\tau_c) \right) \quad (3)$$

$$\begin{aligned} & \text{Auto-Correlation}(Q, K, V) \\ &= \sum_{i=1}^c \text{Roll}(V, \tau_i) \hat{R}_{Q,K}(\tau_i) \end{aligned} \quad (4)$$

Among them,  $\text{Topc}(\cdot)$  is the C lag order  $\tau_1, \dots, \tau_c$  for selecting the maximum probability;  $\text{soft max}(\cdot)$  is a normalized function;  $R_{Q,K}$  represents the autocorrelation information between the sequence Q and the sequence K,  $\text{Roll}(V, \tau_i)$  represents the  $\tau_i$ -order lag processing of the sequence V, and  $\text{Auto-Correlation}(Q, K, V)$  represents the autocorrelation information obtained by this Auto-correlation mechanism[11].

(2) Series Decomp

Aiming at the complex time pattern in time series, Autoformer uses the idea of moving average to decompose the original sequence into trend term and seasonal term [12].

The calculation process is as follows :

$$X_t = \text{AvgPool}(\text{Padding}(X)) \quad (5)$$

$$X_s = X - X_t \quad (6)$$

Where  $X$  represents the original sequence to be decomposed,  $X_t$  represents the trend term generated by moving average processing,  $X_s$  is the seasonal term,  $\text{Padding}(\cdot)$  represents the filling function that maintains the length of the sequence, and  $\text{AvgPool}(\cdot)$  is the average pooling function.

## 2.2. LSTM Model

LSTM is a special recurrent neural network[13]. It achieves the ability of long-term memory by introducing three special 'gate' mechanisms. The specific process and calculation are as follows:

(1) The forgetting gate is responsible for managing the discarding of information in the memory unit, and determining which information to discard the memory unit based on the previous time step unit state  $S_{t-1}$  and the current input  $X_t$ . The formula is :

$$f_t = \sigma(W_f \cdot [S_{t-1}, X_t] + b_f) \quad (7)$$

Among them,  $\sigma(\cdot)$  is sigmoid activation function;  $W_f$  is the weight matrix of forgetting gate;  $b_f$  is the bias matrix of the forgetting gate.

(2) The input of the input gate control information is based on the state  $C_{t-1}$  of the previous time step and the current input information  $X_t$ , and selectively retains part of the information to the current state  $C_t$ . The formula is :

$$i_t = \sigma(w_i \cdot [S_{t-1}, X_t] + b_i) \quad (8)$$

$$\tilde{C}_t = \tanh(W_c \cdot [S_{t-1}, X_t] + b_c) \quad (9)$$

Among them,  $w_i$  and  $W_c$  are the weight matrices of input gates and candidate vectors, respectively;  $b_i$  and  $b_c$  are input gate bias matrices.  $\tilde{C}_t$  is a candidate memory unit. Through  $f_t$ ,  $i_t$  and  $\tilde{C}_t$  combined with long-term memory and current data, the current state  $C_t$  is obtained by weighted calculation. The formula is :

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (10)$$

Where  $\otimes$  denotes the multiplication of matrices by elements.

(3) The output of the gate control data information determines how much information is output from the current state  $C_t$  to the current unit state  $S_t$ . The formula is :

$$O_t = \sigma(W_o \cdot [S_{t-1}, X_t] + b_o) \quad (11)$$

Where  $W_o$  is the weight matrix of the output gate.

Finally, the output  $S_t$  of the hidden layer at the current time is obtained by  $O_t$  and  $C_t$ , and the calculation formula is :

$$S_t = O_t \otimes \tanh(C_t) \quad (12)$$

### 2.3. Autoformer-LSTM Model

In order to deal with the periodicity, seasonality and non-stationarity of runoff time series, this paper constructs a runoff prediction model based on Autoformer-LSTM. The overall structure of the model is shown in Figure 1. Its core mechanism autocorrelation mechanism, sequence decomposition and long short-term memory neural network are embedded into the encoder and decoder as three

independent modules.

The Autoformer model has significant advantages in capturing long-term dependencies and multi-scale patterns of time series, while LSTM effectively optimizes the extraction of local features through its gating mechanism. The Autoformer-LSTM model replaces the Feed Forward feed-forward neural network in Autoformer with Long Short-Term Memory (LSTM) neural network, which further enhances the feature extraction ability of complex time series data. At the same time, the noise suppression ability of LSTM effectively reduces the potential interference of data noise on the prediction performance of the model, and enhances the robustness of the model while improving the prediction accuracy.

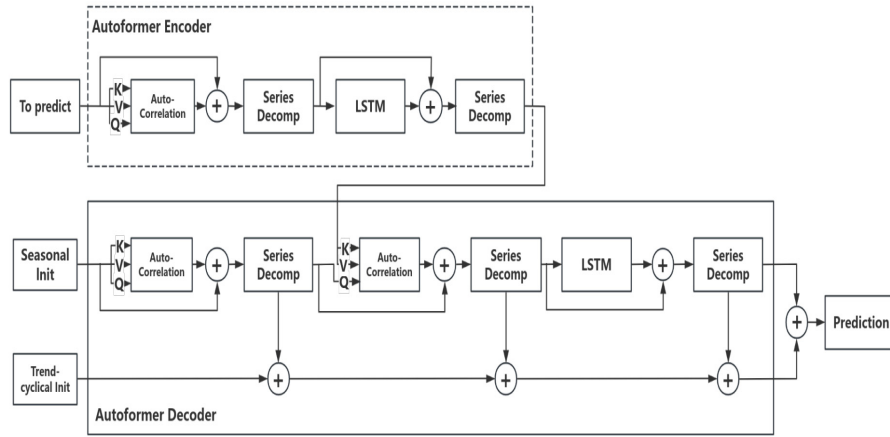


Fig 1. Autoformer-LSTM model structure diagram

## 3. Application Example

### 3.1. Research Object

The Huayuankou hydrological station is the key control station of the main stream of the Yellow River and the standard station for flood control in the lower reaches. The control basin area is 730,000 Km<sup>2</sup>, accounting for 97 % of the total area of the Yellow River basin. It is an important hydrological station at the national level [14]. The monitoring flow of the station has always been an important basis for downstream flood control. In this paper, the daily runoff and daily average water level data of Huayuankou hydrological station from January 1,2002 to December 31,2022 are

selected, and the meteorological data are selected from Zhengzhou station (site No.57083) from January 1,2002 to December 31,2022. Daily average temperature, daily precipitation, relative humidity and daily average wind speed data. The hydrological station data comes from the Yellow River Water Conservancy Commission of the Ministry of Water Resources of the Yellow River Network, and the meteorological data comes from the National Meteorological Information Center. The random forest model and the mean substitution method are used to deal with some missing and abnormal data.

### 3.2. Analysis of Relationship

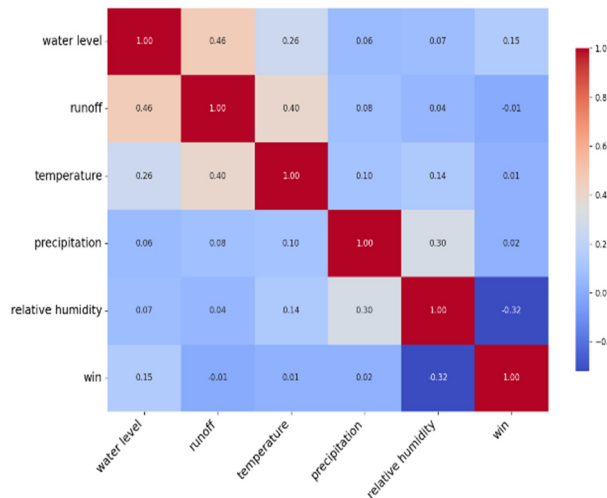


Fig 2. Related heat maps

In this paper, six factors of daily average flow, daily average water level, daily precipitation, daily average temperature, relative humidity and average wind speed of Huayuankou station are selected to explore the prediction of runoff. In order to better explore the influence degree of influencing factors on runoff prediction research, the Pearson correlation coefficient test between each factor is first carried out [15], and the correlation coefficient heat map between the obtained factors is shown in figure 2.

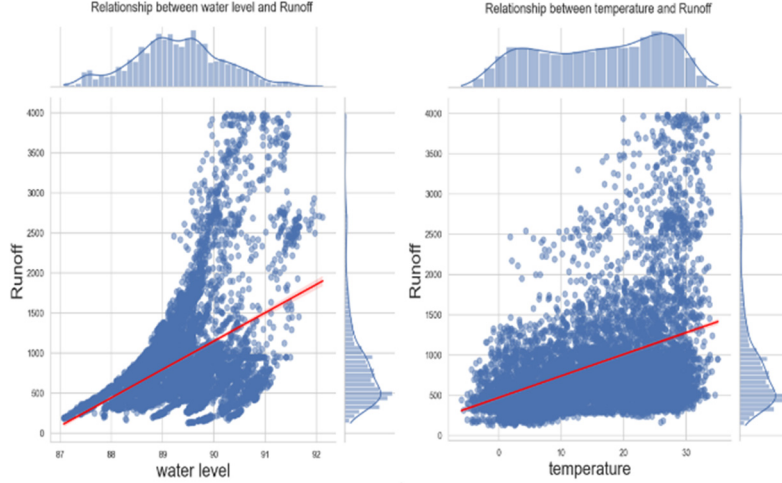


Fig 3. Scatter plot of correlation between water level, temperature and runoff

### 3.3. Model Parameter Settings

In order to evaluate the performance of multivariate Autoformer-LSTM model in daily average flow prediction, Autoformer model, Transformer model and Informer model are set as comparative prediction models. The data is divided into training set, validation set and test set according to 7:1:2. In the model parameter setting, the number of nodes in the input and output layers of the model is equal to the number and 1 of the input variables, and the iterative update is performed on the divided training set. The model includes two encoder layers and one decoder layer. MSE is used as the loss function, and Adam optimizer is selected. The initial learning rate is set to 0.001, the activation function is gelu, the training rounds are 50 epochs, and the remaining hyperparameters are obtained by trial-and-error method.

### 3.4. Evaluating Indicator

In order to accurately evaluate the prediction effect of each model, this paper selects mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE) and fitting coefficient ( $R^2$ ) to evaluate the prediction accuracy of various prediction models. The calculation formulas of MAE, MSE, RMSE and  $R^2$  are as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - y'_t| \quad (13)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - y'_t)^2 \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - y'_t)^2} \quad (15)$$

By observing the Pearson correlation coefficient results between the characteristics and the target variables, it can be seen that the correlation between the daily average water level and the daily average temperature and the runoff data is strong. The specific correlation scatter plot is shown in Figure 3, and a large number of data points are close to the regression line. The second is the daily precipitation. The correlation between relative humidity and average wind speed and runoff data is weak, and the influence degree is small, so it is not used in the final prediction.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y'_t - y_t)^2}{\sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (16)$$

In the formula:  $n$  is the number of time series data,  $y_t$  is the observed value,  $y'_t$  is the predicted value, and  $\hat{y}_t$  is the arithmetic mean value of the observed value.

### 3.5. Model Prediction Results and Analysis

After the training of the runoff prediction model of Huayuankou hydrological station established by Autoformer-LSTM method is completed, the comparison line chart between the predicted results and the corresponding real values on the test set is shown in Figure 4. It can be seen that the trend line of the predicted value of the model is generally consistent with the fluctuation trend of the measured value, and the predicted value is in good agreement with the real value. The overall prediction can basically fit the change trend of runoff.

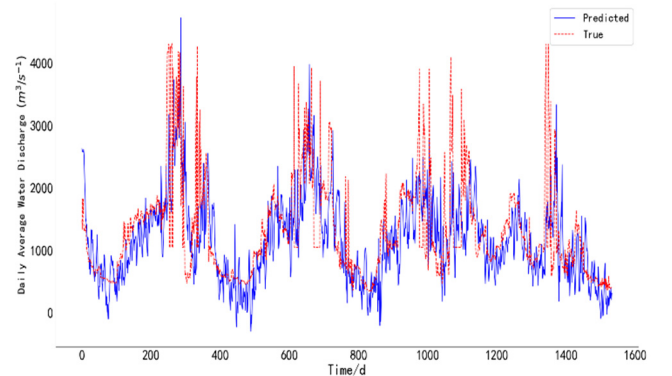


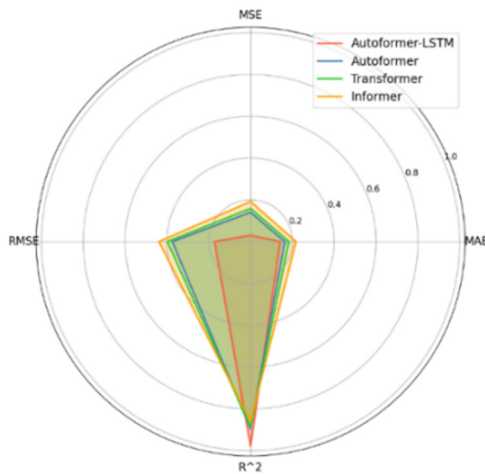
Fig 4. Autoformer-LSTM model fitting effect diagram

In order to fully evaluate the prediction effect of Autoformer-LSTM runoff prediction model, this paper selects Autoformer, Transformer and Informer models for comparative verification, and uses MAE, MSE, RMSE and  $R^2$  to evaluate the prediction performance of each model. In order to reduce the contingency that may occur during the training process, multiple training methods are used, and the results with the best training performance of each model are selected for recording. The results are shown in Table 1.

**Table 1.** Comparison of results from multiple prediction models

MODEL	MAE	MSE	RMSE	$R^2$
<b>Autoformer-LSTM</b>	0.1423	0.0298	0.1728	0.9777
<b>Autoformer</b>	0.1638	0.1403	0.3746	0.8969
<b>Transformer</b>	0.1833	0.1584	0.3980	0.8880
<b>Informer</b>	0.2170	0.1918	0.4380	0.8605

It can be seen from Table 1 that the main accuracy evaluation indexes of the Autoformer-LSTM combined prediction model proposed in this paper are better than those of Autoformer, Transformer and Informer models. The MAE value of Autoformer-LSTM model is 0.1423, which is 13.1%, 22.3% and 34.4% lower than that of Autoformer, Transformer and Informer respectively. The MSE value of Autoformer-LSTM model is 0.0298, which is 78.7%, 81.1% and 84.3% lower than that of Autoformer, Transformer and Informer respectively. The Autoformer-LSTM model is also improved compared with the other three models.



**Fig 5.** Comparison radar chart of model prediction evaluation indicators

As shown in Figure 5, both the Autoformer-LSTM model and the Autoformer model show better overall performance than the Transformer model and the Informer model in terms of prediction performance. In particular, because LSTM is good at dealing with complex nonlinear relationships and effectively capturing long-term dependencies in time series, the Autoformer-LSTM model combined with LSTM significantly exceeds the single Autoformer model in prediction accuracy, further improving the prediction effect. The above analysis fully verifies that the Autoformer-LSTM model has higher applicability than other models in the runoff prediction task of multivariate time series. This method can not only effectively capture the internal relationship of complex runoff time series data, but also significantly improve the prediction accuracy of runoff prediction, thus

proving its superiority and effectiveness.

## 4. Conclusion

Aiming at the problem of river runoff prediction, this paper proposes a multivariate runoff prediction model based on Autoformer-LSTM. The model effectively captures and utilizes various patterns and laws in runoff time series data, and excavates the long-term correlation in runoff time series data, so as to more accurately obtain the internal relationship of complex runoff time series data. The runoff data of Huayuankou Hydrological Station of the Yellow River from January 1, 2002 to December 31, 2022 were taken as an example for verification and analysis. The main conclusions are as follows:

(1) According to the results of correlation characterization based on Pearson correlation coefficient, it can be found that for the hydrological data of Huayuankou hydrological station, daily average water level and daily average temperature are the key related characteristic factors in runoff prediction, while other characteristic factors such as relative humidity and daily average wind speed have little effect on runoff prediction. The results show that the Autoformer-LSTM model and the Autoformer model are significantly better than the Transformer model and the Informer model in prediction performance.

(2) From the analysis of the case verification results of the Huayuankou hydrological station on the Yellow River, although the Autoformer, Transformer, Informer and Autoformer-LSTM models can effectively capture the general trend of the daily average flow, the results show that the Autoformer-LSTM model and the Autoformer model are significantly better than the Transformer model and the Informer model in prediction performance. In addition, compared with the Autoformer model, the values of MAE, MSE and RMSE of the Autoformer-LSTM model decreased by 13.1%, 78.7% and 53.9%, respectively. The  $R^2$  of the Autoformer-LSTM model is also 8.3% higher than Autoformer. In summary, the Autoformer-LSTM model can better capture the complex internal relationship and long-term correlation between runoff time series, thus effectively improving the accuracy of runoff prediction.

## References

- [1] Yang Qiongbo, Cui Dongwen, Comparison of WPD-RSO-ESN and SSA-RSO-ESN models in runoff time series prediction [J]. China Rural Water Conservancy and Hydropower, 2022,(02):61-67+75.
- [2] Wang Wen, Ma Jun. Summary of several hydrological forecasting methods [J]. Technologic progress of water conservancy and hydropower, 2005,(01):56-60.
- [3] Wang Jun, Gao Zixun, Shan Chunyi. Multivariate Yellow River runoff prediction based on TCN-Attention model [J]. People's Yellow River, 2022,44 (11):20-25.
- [4] Qian Lipeng, Liu Changzheng, Chen Cuizhong, et al. A runoff prediction method based on deep belief network [J]. Journal of Shihezi University (Natural Science Edition), 2021,39 (02): 259-264. DOI:10.13880/j.cnki.65-1174/n.2020.21.047.
- [5] Tong TaoWen. Research on river runoff and sediment concentration prediction method based on LSTM deep learning[J]. Journal of Irrigation and Drainage, 2021,40(S1):1-4. DOI:10.13522/j.cnki.gggs.2021146.
- [6] Do Miao, Ho XiangNing. Application of PCA-MLP neural network model in runoff prediction of Ningxia section of the

- Yellow River [J]. Water conservancy informatization, 2024, (04):49-53.DOI:10.19364/j.1674-9405.2024.04.009.
- [7] Wu Haixu, Xu Jiehui, Wang Jianmin, et al. Autoformer: Decomposition transformers with Auto-Correlation for long-term series forecasting [J]. Advances in Neural Information Processing Systems, 2021, 34:22419-22430.
- [8] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997,9(8):1735-1780.
- [9] Athanasios Papoulis and H Saunders. Probability, random variables and stochastic processes.1989.
- [10] Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In ICLR, 2020.
- [11] Shiyang Li, Xiaoyong Jin, Yao Xuan, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In NeurIPS, 2019.
- [12] Reza Asadi and Amelia C Regan. A spatio-temporal decomposition based deep neural network for time series forecasting. Appl. Soft Comput., 2020.
- [13] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323 (6088): 533-536.
- [14] Kang Miaoye, Xiao Weihua, Lu Fan, et al. Analysis of multi-time scale runoff evolution law of Huayuankou hydrological station [J]. People's Yellow River, 2022, 44(05): 25-29.
- [15] Jiang Wenqian, Xu Da, Lin Xiuqing, et al. Nonnegative Matrix Factorization and Improved Correlation Analysis for Low Voltage Area Topology Identification [J]. Journal of Power Systems and Automation, 2024, 36(07): 133-139. DOI: 10.19635/j.cnki.csu-epsa.001351.