**Research on River Annual Runoff Prediction Model Based on EEMD-ANFIS**

**Huifang Guo,Shixia Zhang,Yuan Fang\***

Zhejiang Tongji Vocational College of Science and Technology ,Hangzhou,Zhejiang,China

\*Corresponding Author:Yuan Fang(E-mail address:z20120190801@zjtongji.edu.cn).

|  |
| --- |
| ***Abstract:***  Rivers play an important role in human production and life. On the one hand, rivers provide water for human production and life. On the other hand, when the river has too much water, it will bring flood disasters to human beings. Therefore, the prediction of river runoff is particularly important. Accurate runoff prediction can not only provide basic data for the allocation and operation of water resources, but also provide reference for flood control and waterlogging control of the basin. The formation process of runoff is affected by rainfall, underlying surface, human activities and other factors. The improvement of runoff prediction accuracy has always been a difficult problem in the hydrological field. Because the runoff is affected by many factors and contains a lot of noise, the prediction accuracy will be reduced by using the data containing noise. EEMD is a good tool to separate signal and noise. This method is used to preprocess the runoff series, decompose the runoff series into multiple IMF intrinsic modulus, and then use the ANFIS algorithm with strong nonlinear approximation ability to predict each IMF function, and then reconstruct the predicted data to improve the prediction accuracy of runoff. By comparison, the prediction accuracy of EEMD-ANFIS model is about 34% higher than that of ANFIS model.  ***Keywords****: Runoff, river, runoff prediction, prediction accuracy, EEMD-ANFIS.* |

**I. INTRODUCTION**

Water is an indispensable part of human life. Industry, agriculture and life are inseparable from water resources. However, with the shortage of water resources, the occurrence of rapid turn of drought and flood, the intensification of water pollution and other reasons, water resources have become a strategic resource [1]. The amount of water resources will limit the development of cities. The shortage of water resources will affect people's domestic water, industrial production water and agricultural production water. On the other hand, it will also lead to the gradual reduction of flood control and drainage capacity, and the economic loss will increase exponentially when encountering the same level of rainstorm. When the runoff is large, it will cause flood [2], and when the runoff is small, it will cause water shortage. Therefore, accurate runoff prediction becomes particularly important. Accurate runoff prediction can create conditions for human beings to make better use of water resources. However, with the acceleration of urbanization, the increase of urban population, the further agglomeration of production factors and the rapid development of social economy, there are fewer and fewer natural storage sites, but the floor area of hard ground such as cement roof and pavement is increasing, which makes the convergence time of flood faster, and the change process and law of flood become difficult to predict.

There are many runoff prediction methods, and the research at home and abroad can be roughly divided into two kinds [3]. One is to use mathematical and chemical models to simulate the generation process of runoff, such as Xin'anjiang model, and the other is to use black box model to simulate the law of hydrological change. With the rapid development of computer technology, the black box model is used to establish the nonlinear relationship between the factors affecting runoff and runoff, and improve the accuracy of runoff prediction with the help of the rapid calculation advantage of computer. At present, it is widely used in runoff prediction. However, in the process of runoff generation, due to the influence of many factors, it mostly contains noise. The existence of noise submerges the real change process of runoff. Using the runoff data with noise to predict directly, its accuracy is generally difficult to meet the requirements. In order to improve the accuracy of runoff prediction, EEMD model is introduced in this paper. The model has good decomposition ability and has been widely used in finance [4], power [5].physics[6,7] and other industries. The model is used to decompose the runoff data and decompose the runoff series into more stable intrinsic modulus functions IMF. The ANFIS model with strong adaptive, self-learning and nonlinear approximation ability is used to predict the IMF respectively, and the predicted IMF is reconstructed to obtain the prediction results of the original runoff series. By comparing the relative error of EEMD-ANFIS calculation results with that of ANFIS model, the applicability of this method in river runoff prediction is explored.

**II. Research progress of runoff prediction at home and abroad**

In 1995, Cai Yudong [8] proposed an artificial neural network method for long-term runoff prediction, which can be widely used in runoff prediction. In 1998, Feng Guozhang [9] proposed a forward multi-layer artificial neural network runoff prediction model based on runoff formation mechanism and taking period precipitation and early runoff as prediction factors. In 1999, Xie Xinmin [10] applied artificial neural network technology to the real-time prediction of river runoff, and optimized the model by using conjugate gradient optimization and error back propagation training algorithm. In 1999, Quan Xianzhang [11] established a local prediction method based on chaotic dynamics and applied it to the runoff prediction of multiple rivers. The results show that the prediction result has obvious advantages. In 2006, Lin Jianyi [12] explored the application of support vector machine in medium and long-term runoff prediction, and compared it with ANN model. It is considered that support vector machine model can improve the prediction accuracy. In 2012, Zhao tongtiegang [13] proposed a runoff prediction method based on random forest, and the results show that the prediction accuracy is good. In 2018, Li Wenjing [14] proposed the improved quantum particle swarm optimization (IQPSO), and used the algorithm to realize the automatic optimization of support vector machine parameters. The results show that the prediction accuracy of the model is good. In 2019, Fang Ruiming [15] proposed a combined runoff prediction model based on wavelet transform and correlation vector machine. The results show that the prediction accuracy of the model can meet the requirements of monthly runoff prediction. In 2021, Feng Rui [16] proposed an ln-lstm-pso model and applied it to runoff prediction. The results show that the prediction accuracy is significantly improved. In 2021, Chen Shu [17] proposed a runoff prediction model based on two-stage decomposition and support vector machine. The results show that the combination of machine learning method based on two-stage decomposition can effectively improve the prediction accuracy of annual runoff. In 2022, Xiao Lu [18] applied the regression tree ensemble model to the daily runoff prediction. The maximum absolute value of the relative error of the model is 8.11%. The ensemble model is outstanding in Hydrological Prediction. In 2022, Han heechan, [19] used the deep learning method to make error prediction and improve the performance of runoff prediction. The results show that the model is effective for hydrological modeling results with short prediction lead time. In 2022, Jing Xin [20] applied variational modal decomposition and artificial neural network to runoff prediction and obtained better prediction results.in 2022, Xiao Wenjing [21] et al. predicted the flow of Pingshan Hydrologic Station, Yichang Hydrologic Station and Cuntan Hydrologic Station on the Yangtze River by using the short-term memory network and support vector machine model, and evaluated the prediction results. The results show that the short-term memory network is better than the support vector machine model.In 2022, Li Fuwei [22] et al. proposed a new medium range runoff forecasting method based on multi model fusion. This method uses information entropy, BP neural network and support vector machine to forecast the medium term runoff. After the evaluation of the average absolute error, root mean square error, prediction qualification rate and other indicators, it shows that the calculation results of this method are better than single prediction models such as BP neural network, multiple regression and support vector machine, and the prediction results of the combined model are more consistent with the reality.In 2022, Zhang Xianqi [23] et al. forecast the monthly runoff data of Cuntan Hydrometric Station by combining the autoregressive comprehensive moving average model and EEMD model. The research shows that the prediction accuracy of this method has been effectively improved, and the problem of time series data interruption or data fluctuation disappearance due to the extension of time series has been solved.In 2022, Li Bao Jian [24] et al. combined the ensemble empirical mode decomposition, sparrow search method and short and long term neural network for the study of monthly runoff forecast. The results show that the performance evaluation of the calculation results indicators of the combined model is excellent and can meet the needs of monthly runoff forecast.

**III.EEMD-ANFIS model**

3.1EEMD

The set empirical mode decomposition algorithm [25,26] was proposed by professors of Central University (Huang and Wu) in 2004. It is improved on the basis of traditional EMD. The traditional EMD algorithm was proposed by NASA scientists in 1998. It is a time-domain filtering algorithm for denoising. This method does not need to input frequency parameters in advance, and is effective for the denoising of nonlinear and non-stationary data sequences. This method can extract the eigenmodulus functions with different periodic characteristics and realize the stabilization preprocessing of the analysis data. At present, it has been widely used in finance, petroleum and other industries. The flow of EMD signal processing is as follows:

(1) Analyze the size and location of local minimum and maximum of data sequence y (T).

(2) The cubic spline curve is used as the envelope to fit the maximum and minimum respectively, and form the upper and lower envelope U (T) and l (T) of the minimum of the maximum set.

(3) Calculate the mean m (T) = [u (T) + L (T)] / 2 at each time point of the upper and lower envelope, and calculate the difference H1 (T) between the mean m (T) and the original value Y (T), to judge whether it meets the requirements. If so, the value is the intrinsic modulus function IMF. If it does not meet the assumed conditions, replace the original data y (T) with the difference H1 (T), and repeat the above steps until the requirements are met.

(4) After the nth iteration:

Then, if the judgment conditions are met, the first intrinsic modulus imf1 (T) = H1N (T).

(5) Calculate the residual sequence R (T):

R(t)=Y(t)-IMF1(t)

Judge whether it meets the EMD decomposition conditions. If so, end the decomposition. If not, let H2 (T) = R (T), X (T) = R (T), and return to step (1).

(6) Through the above steps, gradually calculate the 1st ~ nth intrinsic modulus imfi (T) and a residual R (T), and the combination of intrinsic modulus and residual is the original data.

A major defect of EMD algorithm is that it assumes that the original data has no noise, but the actual data sequence contains noise. The existence of noise will cause mode aliasing in the process of decomposition. Based on the above problems, Huang improved the EMD algorithm, that is, in the process of decomposition, evenly distributed white noise is added for many times, and the white noise is used to mask the noise existing in the data itself, so as to solve the problem of mode aliasing. The calculation steps of EEMD are as follows:

(1) In the original data, the white noise is n times, and a new data sequence y\*(T) containing noise is obtained.

(2) EMD decomposition is performed on y\*(T) to calculate the intrinsic modulus IMFn.

(3) Repeat the above steps until all eigenmoduli and residuals are calculated.

The key parameters of EEMD algorithm are the number of adding noise and setting the standard deviation of adding noise. According to relevant literature research, the standard deviation is generally set as 0.005-0.2, and the calculation result is better when the number of noise additions is 100.

3.2 ANFIS model

In the early 1990s, J. - S.R. Jang proposed ANFIS model, which includes three basic processes: fuzziness, fuzzy reasoning and anti fuzziness. The model not only has the advantages of self-learning, self-adaptive and nonlinear approximation of artificial neural network, but also can use fuzzy rules. Therefore, the model is widely used in solving nonlinear models. It has two key parameters for learning: pre component parameters and post component parameters. Suppose a multi input single output layer fuzzy neural network, the output of  nodes in layer K is , the expected output is , and the actual output is 

3.2.1. Learning of front part parameters

Assuming that the membership function used in the second layer is Gaussian type:

|  |  |  |
| --- | --- | --- |
|  |  | （1） |

Where and are parameters related to nodes.

Assuming that there are training samples of  group, the training error function of group is defined as:

|  |  |  |
| --- | --- | --- |
|  |  | （2） |

The error function of the whole network is:

|  |  |  |
| --- | --- | --- |
|  |  | （3） |

The gradient descent method is used to improve the learning speed. Firstly, the first derivative of the  group input vector to each output node is calculated:

|  |  |  |
| --- | --- | --- |
|  |  | （4） |

The error change rate ofnode is:

|  |  |  |
| --- | --- | --- |
|  | , | （5） |

Convert it

|  |  |  |
| --- | --- | --- |
|  |  | （6） |

Among them,  is a parameter of the adaptive network,  is the set of nodes related to  output, then the error index function of the whole network about  is:

|  |  |  |
| --- | --- | --- |
|  |  | （7） |

Taking  as an example, the parameter update formula is:

|  |  |  |
| --- | --- | --- |
|  |  | （8） |

 is learning efficiency.

|  |  |  |
| --- | --- | --- |
|  |  | （9） |

 is the step size, and the speed of can be changed by changing the speed of .

3.2.2. Learning of subsequent parameters

The expected output  and actual output  have the following relationship with the corresponding expected input :

|  |  |  |
| --- | --- | --- |
|  |  | （10） |

In which,

|  |  |  |
| --- | --- | --- |
|  |  | （11） |

, is the  sample data pair.

According to the least square method, its performance index is defined as:

|  |  |  |
| --- | --- | --- |
|  |  | （12） |

Among them,  is the forgetting factor, and the value is 0 ~ 1. The smaller the , the weaker the influence of historical data. Bring  into :

|  |  |  |
| --- | --- | --- |
|  |  | （13） |

From ,then

|  |  |  |
| --- | --- | --- |
|  |  | （14） |

The recursive least square method is used to solve the problem, and the corresponding estimation is:

|  |  |  |
| --- | --- | --- |
|  |  | （15） |

Where:

|  |  |  |
| --- | --- | --- |
|  |  | （16） |

 becomes the covariance matrix. Usually, the initial piece is selected as , , is the unit matrix, is a large positive number, and a positive number of 103 can be taken. When , the recursive least square method is close to the real value.

3.3 EEMD-ANFIS model

The calculation steps of EEMD-ANFIS are:

(1) The runoff series is decomposed into several IMF functions by EEMD.

(2) ANFIS model is used to predict each IMF function.

(3) The simulation and prediction structure of each IMF is reconstructed to obtain the simulation sequence results of the original runoff series.

1. Analyze the error of the prediction model to see whether it meets the requirements of runoff prediction.

**IV Case study**

In this paper, Nanxi River Basin is used for example analysis. The hydrological station used for runoff is Shizhu station

4.1 Basin overview

Nanxi River Basin is located in the eastern coast of Zhejiang Province. It is an important tributary of the lower reaches of Oujiang River. It belongs to Yongjia County, Jinyun County, Huangyan District, Yueqing City and Qingtian County. It is the largest basin in Yongjia. The main stream of Nanxi River has a total length of 142.2km, a drainage area of 2436km2 and an average gradient of 6.0‰. The main stream is a mountain stream above Shatou. The gradient of the river channel 61km above the mouth of the stream is 10.46‰. The gradient of the river channel 48km from the mouth of the stream to Shatou is 1.12‰. The river channel below the head of the stream is affected by the tide of the Oujiang Estuary and is a tidal reach.

4.2 Selected hydrological stations

Shizhu hydrological station was established by Zhejiang Hydrological Bureau on June 1, 1955, with a catchment area of 1273km2. It observed the water level, water surface gradient and precipitation, and began to observe the flow on July 19 of the same year. The flow section is located in the downstream section of the 400m relatively straight river section. In normal water period, the river width is about 100m, the beach land is about 90m, the high water river is about 200m wide, the deep water line is inclined to the right bank, and there is a small stream about 800m downstream. The high water is sometimes affected by its backwater jacking. There is a large bend about 1km upstream, a sharp beach 200m downstream of the section, and the confluence of Da and Xiao Nan streams is 4km downstream.

4.3 Runoff characteristics

The annual average flow of Shizhu station from 1961 to 2018 was 45.2m3/s, the highest year was 84.9m3/s (1990), the lowest year was 23.2m3/s (2003), and the ratio of runoff in high and low water years was 3.66 times. The annual water distribution usually presents a bimodal type. Among them, the front peak is located in June and is generally formed by Meiyu; The back peak occurred from August to September, mainly due to Taiwan Wind and rain. The dry season is mostly from November to February of the next year, and the total runoff in these four months accounts for only 11.1% of the total annual volume. The runoff in December accounts for only 2.3% of the annual runoff. The monthly distribution of multi-year average runoff of Shizhu station is shown inTable 1(Table 1:monthly distribution of annual average runoff of Shizhu station).

**TABLE I. monthly distribution of annual average runoff of Shizhu station**

|  |  |  |
| --- | --- | --- |
| **month** | Average flow (m3/s) | Percentage(%) |
| **January** | 13.1 | 2.5 |
| **February** | 19.8 | 3.6 |
| **March** | 37.9 | 7.2 |
| **April** | 44.5 | 8.1 |
| **May** | 52.8 | 10 |
| **June** | 90.9 | 16.6 |
| **July** | 65.8 | 12.4 |
| **August** | 90.3 | 17 |
| **September** | 71.8 | 13.1 |
| **October** | 26.5 | 5 |
| **November** | 14.9 | 2.7 |
| **December** | 12 | 2.3 |
| **annually average** | 45 | 100 |

4.4 Calculation results

In view of the information obtained,the annual runoff data of Shizhu station from 1957 to 2019 are used for calculation. The EEMD decomposition results are shown in Figure 1(Fig 1:EEMD decomposition results).

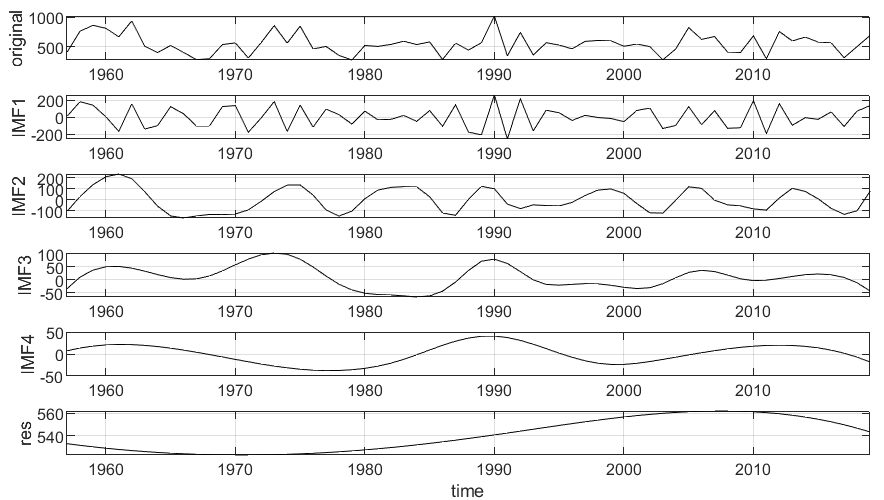


Fig 1:EEMD decomposition results

After that, AFNSI was used to calculate each IMF component. After observing the error change process, the number of iterations was set to 500. Through consulting the research results of relevant literature, the calculation function was set to be a beta function. And compare the EEMD afis model simulation prediction results with the ANFIS results under the same conditions, the comparison diagram is shown in Figure 2(Fig. 2 comparison results of simulation and prediction).

Fig. 2 comparison results of simulation and prediction

The relative errors of EEMD ANFIS model and ANFIS model are decomposed and calculated. The calculation results are shown in Table 2(TableII. relative errors of EEMD ANFIS model and ANFIS model).

**TableII. relative errors of EEMD ANFIS model and ANFIS model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **relative error** | | **<55%** | **<30%** | **<20%** |
| EEMD-ANFIS | number | 60 | 60 | 55 |
| Proportion | 98.36% | 98.36% | 90.16% |
| ANFIS | number | 60 | 51 | 41 |
| Proportion | 98.36% | 83.61% | 67.21% |

From the above table, we can see that EEMD-ANFIS model can get better simulation results. The relative error is about 98% when it is less than 30% and 90% when it is less than 20%, which meets the accuracy requirements. The average error calculated by EEMD afis is 9.24%. It can be seen that the calculation results of EEMD-ANFIS model are obviously better than the prediction results of ANFIS.

**V Conclusion**

River runoff not only affects human production and domestic water, but also affects urban flood control and drainage. Accurate runoff prediction can not only provide basic data for the allocation and operation of water resources, but also provide reference for flood control and waterlogging control of the basin. The formation process of runoff is affected by rainfall, underlying surface, human activities and other factors. It usually contains a lot of noise. Using the data containing noise for prediction will reduce the accuracy of prediction. In this paper, EEMD is used to decompose the process of runoff series, and the runoff series is decomposed into multiple IMF functions. The ANFIS model is used to predict each IMF function, and then the predicted data are reconstructed. The results show that EEMD-ANFIS model can get better prediction results than ANFIS model. The relative error of EEMD-ANFIS model is less than 20%, which is about 90%, which is about 34% higher than that of ANFIS model.

**ACKNOWLEDGEMENTS**

This research was supported by Provincial cultivation project of Natural Science Foundation of Zhejiang Province (Grant No.LZJWY22E090005) &Zhejiang Tongji Vocational College of science and Technology (Grant No. FRF21PY001)&Science and technology project of Zhejiang Water Resources Department in 2020 (Grant No.RC2031,RA2008).

**Conflict of Interest**

The authors declare no potential conflict of interest.

**REFERENCES**

[1] Wang Yao Research on Evaluation of urban comprehensive carrying capacity of Shanghai based on Sustainable Development [D] East China Normal University, 2016

[2] Li Qiong Research and improvement of flood disaster risk analysis and evaluation method [D] Huazhong University of science and technology, 2012

[3] Guo Huifang Research on medium and long term runoff prediction model based on wavelet analysis [D] Zhengzhou University, 2007

[4] Chen Yue Volatility Prediction of CSI 300 index based on Wavelet and EEMD noise reduction [D] Overseas Chinese University, 2019

[5] Cheng Qiming, Chen Lu, Cheng yinman, Zhang Qiang, Gao Jie Short term prediction method of wind power based on EEMD and LS-SVM model [J] Power automation equipment, 2018,38 (05): 27-35 DOI:10.16081/j.issn. 1006-6047.2018.05.004.

[6] Maharishi Markandeshwar,etc.Short Term Solar Irradiation Prediction Framework Based on EEMD-GA-LSTM Method[J],[Strategic Planning for Energy and the Environment](https://www.journal.riverpublishers.com/index.php/index),2022: Vol 41 Iss 3

[7] Anuj Gupta,etc.Solar Irradiation Forecasting Technologies: A Review[J].[Strategic Planning for Energy and the Environment](https://www.journal.riverpublishers.com/index.php/index),[2020: Vol 39 Iss 3-4 2020](https://www.journal.riverpublishers.com/index.php/SPEE/issue/view/339).

[8] Cai Yudong, Yao Linsheng Artificial neural network method for long-term runoff prediction [J] Progress in water science, 1995 (01): 61-65

[9] Feng Guozhang, Li Peicheng Influence analysis of artificial neural network structure on runoff prediction accuracy [J] Journal of natural resources, 1998 (02): 73-78

[10] Xie Xinmin, Jiang Yunzhong, Shi Yubo, Dang Yong, Yang Chunxiao Research on real-time prediction of river runoff based on artificial neural network [J] Water conservancy and hydropower technology, 1999 (09): 1-4

[11] Quan Xianzhang, Wen Quan, Zhang Yongchuan Application of chaotic prediction technology in runoff prediction [J] Journal of Huazhong University of technology, 1999 (12): 41-43 DOI:10.13245/j.hust. 1999.12.015.

[12] Lin Jianyi, Cheng Chuntian Application of support vector machine in medium and long term runoff prediction [J] Journal of water conservancy, 2006 (06): 681-686

[13] Zhao Tongtie Gang, Yang Dawen, Cai Ximing, Cao Yong Study on Runoff Prediction in dry season in the upper reaches of the Yangtze River Based on Stochastic Forest model [J] Journal of hydropower, 2012,31 (03): 18-24 + 38

[14]Li Wenjing, Li Peiwu Application of optimized SVM based on IQPSO in runoff prediction [J] Journal of Nanchang Institute of engineering, 2018,37 (03): 54-59 + 90

[15]Fang Ruiming. Wavelet based relevance vector machine model for monthly runoff prediction[J]. Water Quality Research Journal,2019,54(2).

[16]Feng Rui,Fan Guangwei,Lin Jianyi,Yao Baozhen,Guo Qinghai. Enhanced Long Short-Term Memory Model for Runoff Prediction[J]. Journal of Hydrologic Engineering,2021,26(2).

[17]Chen Shu,Ren Miaomiao,Sun Wei. Combining two-stage decomposition based machine learning methods for annual runoff forecasting[J]. Journal of Hydrology,2021,603(PB).

[18]Xiao Lu,Zhong Ming,Zha Dawei. Runoff Forecasting Using Machine-Learning Methods: Case Study in the Middle Reaches of Xijiang River [J]. Frontiers in Big Data,2022,4.

[19]Han Heechan,Morrison Ryan R. Improved runoff forecasting performance through error predictions using a deep-learning approach[J]. Journal of Hydrology,2022,608.

[20]Jing Xin,Luo Jungang,Zhang Shangyao,Wei Na. Runoff forecasting model based on variational mode decomposition and artificial neural networks.[J]. Mathematical biosciences and engineering : MBE,2022,19(2).

[21]Xiao Wenjing, Zhou Jianzhong, Yang Jianhua, Molly, Xu Zhanxing, Yang Yuqi. Medium and long-term runoff forecast of the Yangtze River mainstream based on machine learning [J]. Hydropower and Energy Science, 2022,40 (09): 31-34+26. DOI: 10.20040/j.cnki.1000-7709.2022.20212442

[22]Li Fuwei, Sun Kaixin, Ding Wei. Medium term runoff forecast based on multi model fusion [J/OL]. China Rural Water Resources and Hydropower: 1-14 [2022-09-29] <http://kns.cnki.net/kcms/detail/42.1419.TV.20220915.1104.002.html>

[23]Zhang Xianqi,Lu Yaohui,Zhu Guoyu,Wu Xilong,Zhao Dong,Duan Bingsen. Annual runoff forecast based on a combined EEMD-ARIMA model[J]. Water Supply,2022,22(8).

[24]Li Bao Jian,Yang Jing Xin,Luo Qing Yuan,Wang Wen Chuan,Zhang Tai Heng,Zhong Ling,Sun Guo Liang. A Hybrid Model of Ensemble Empirical Mode Decomposition and Sparrow Search Algorithm-Based Long Short-Term Memory Neural Networks for Monthly Runoff Forecasting[J]. Frontiers in Environmental Science,2022.

[25] Li Chenliang. Based on EEMD\_ Research on prediction of Shanghai and Shenzhen 300 index based on LSTM model [D] Beijing Jiaotong University, 2019 DOI:10.26944/d.cnki. gbfju. 2019.001439.

[26] Wang Shaojun. Research on weak fault feature extraction method of rolling bearing based on EEMD [D] Shijiazhuang Railway University, 2016.