International Core Journal of Engineering Volume 8 Issue 7, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

Time Series Analysis of Precipitation in Zhengzhou based on Wavelet Theory

Jingcheng Liu[1], Mengdi Zhang[1], Shixiao Chang[1], Yingying Hou[1], Zixuan Cheng[1], Yuanhui Cao[2]

[1] College of Mechanical Engineering, Tianjin University of Science and Technology, Tianjin, 300222, China

[2] School of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, 300222, China

Abstract

In July 2021, the city of Zhengzhou in Henan province was hit by a series of heavy rain, which attracted the attention of netizens. Daily precipitation data from 1983 to 2021 were collected in Zhengzhou. Firstly, we integrate the annual precipitation in Zhengzhou and process it anomaly. Secondly, periodicity and mutation point of annual precipitation in Zhengzhou was obtained by Morlet wavelet transform analysis and MK mutation test. Finally, short-term time series prediction of annual precipitation in Zhengzhou was carried out by a wavelet neural network, and the fitting results of BP neural network and SVM support vector machine were compared. According to the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (), the wavelet theory has high accuracy and high stability.

Keywords

Precipitation; Morlet Wavelet Transform; MK Mutation Test; Correlation Analysis; Wavelet Neural Network Prediction.

1. Introduction

The climate in Zhengzhou has a characteristic [5], that is, the monsoon is significant, the severe weather is frequent, and it is close to the Eurasian continent in the west and the Pacific Ocean in the east. The precipitation varies greatly with age, and disasters such as drought and waterlogging are prone to occur, which seriously affects the development of the city's agricultural industry.

Wavelet analysis is developed in view of the shortcomings of Fourier transform. Wavelet analysis highlights the characteristics of signals in certain aspects through time-frequency changes, and stretches and translates the original signals to be processed, thereby eliminating noise and boundary effects. In this paper, we use the 1983-2021 precipitation data observed by the Zhengzhou meteorological observation station to carry out Morlet wavelet analysis and MK mutation test analysis to obtain the climatic characteristics of the annual precipitation in Zhengzhou, and then use the wavelet neural network to analyze the annual precipitation in Zhengzhou. time series prediction, and compared with other time series prediction methods, highlighting the high precision and high stability of wavelet analysis.

2. Analysis Model

2.1 Wavelet Analysis

The basic idea of wavelet analysis is to represent or approximate a signal or function with a set of wavelet function systems. Therefore, the wavelet function is the key to wavelet analysis. It refers to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

a type of function that has oscillation and can rapidly decay to zero, that is, the wavelet function satisfies:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \quad (1)$$

where $\psi(t)$ is the base wavelet function. It can form a family of functions by scaling the scale and shifting on the time axis:

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) \quad a, b \in R, a \neq 0 \quad (2)$$

where $\psi_{a,b}(t)$ is the subwavelet, a is the scale factor, reflecting the period length of the wavelet, b is the translation factor, reflecting the translation in time.

2.1.1. Wavelet Transform

For a given energy-limited signal, its continuous wavelet transform is:

$$W_f(a, b) = |a|^{-1/2} \int_R f(t)\bar{\psi}\left(\frac{t-b}{a}\right) \mathrm{dt} \quad (3)$$

where $W_f(a, b)$ is the wavelet transform coefficient, $f(t)$ is a signal or square integrable function, a is the scaling scale, b is the translation parameter, $\bar{\psi}\left(\frac{t-b}{a}\right)$ is the complex conjugate function of $\psi\left(\frac{t-b}{a}\right)$. Most of the time series data observed in geosciences are discrete. Let the function f(kΔt), (k=1,2,…,N, Δt be the sampling interval), the discrete wavelet transform form is:

$$W_f(a, b) = |a|^{-1/2}\Delta\mathrm{t}\sum_{k=1}^{N} \mathrm{f(k\Delta t)}\,\bar{\psi}\left(\frac{\mathrm{k\Delta t}-b}{a}\right) \quad (4)$$

2.1.2 Wavelet Variance

Integrating the square value of the wavelet coefficients in the b domain, the wavelet variance can be obtained as:

$$\mathrm{Var(a)} = \int_{-\infty}^{+\infty} \left|W_f(a, b)\right|^2 db \quad (5)$$

2.2 MK Mutation Test

For a time series X with n sample size, construct a first-order sequence:

$$S_k = \sum_{i=1}^{k} r_i \quad (6)$$

$$\mathrm{r} = \begin{cases} 1 & X_i > X_j \\ 0 & else \end{cases} \quad j = 1,2, \dots, i \quad (7)$$

It can be seen that the order column sk is the cumulative number of the number of values at the i-th time greater than the j time. Under the assumption of random independence of the time series, define the statistics:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

$$\mathrm{UF}_k = \frac{S_k - E(S_k)}{\sqrt{Var(S_k)}} \quad k = 1,2,3,\ldots,n \tag{8}$$

where UF1=0, $E(S_k)$, $Var(S_k)$ are the mean and variance of the cumulative number sk. When x1, x2, ..., xn are independent of each other and have the same continuous distribution, they can be calculated by the following formula:

$$Var(S_k) = \frac{n(n-1)(2n+5)}{72} \tag{9}$$

where UF_i is a standard normal distribution, which is a series of statistics calculated in the order of x1, x2,...,xn in the time series x. Given a significance level α, check the normal distribution table. If $\mathrm{UF}_i > U\alpha$, it means that there is a clear trend change in the series.

2.3 Wavelet Neural Network

The wavelet neural network is a neural network based on the topology of the BP neural network, and the wavelet basis function is used as the transfer function of the hidden layer nodes. The topology of the wavelet neural network is shown in the Fig.1.

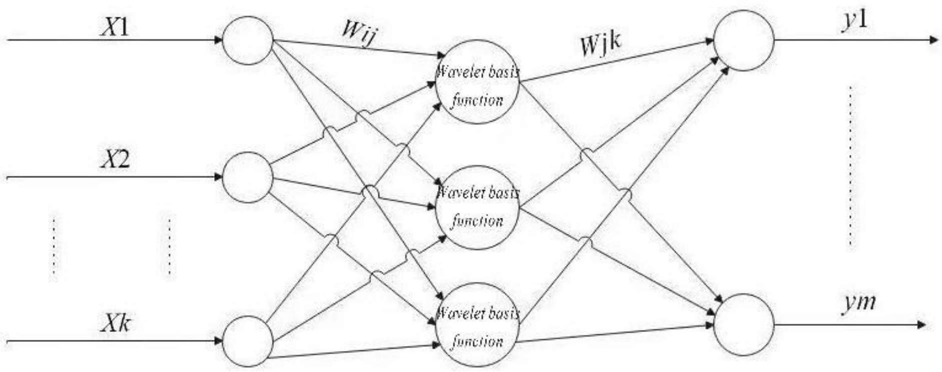

Fig. 1 Wavelet neural network topology

Among them, X_1, X_2, …, X_k are the input parameters of the wavelet neural network, Y_1, Y_2,…, Y_m are the predicted outputs of the wavelet neural network, W_{ij} and W_{jk} are the weights of the wavelet neural network.

When the input signal sequence is x_i $(i = 1,2,\ldots,k)$, the hidden layer output calculation formula is:

$$\mathrm{h(j)} = h_j\left[\frac{\sum_{i=1}^{k} W_{ij}x_i - b_j}{a_j}\right] \quad j = 1,2,\ldots,l \tag{10}$$

where h(j) is the output value of the jth node in the hidden layer, W_{ij} is the connection weight between the input layer and the hidden layer, b_j is the translation factor of the wavelet basis function h_j, a_j is the scaling factor of the wavelet basis function h_j, h_j is the wavelet basis function.

In this paper, we use the wavelet basis function as the Morlet mother wavelet basis function, and the mathematical formula is:

$$\mathrm{y} = \cos(1.75\mathrm{x})e^{-x^2/2} \tag{11}$$

Volume 8 Issue 7, 2022
DOI: 10.6919/ICJE.202207_8(7).0012

The calculation formula of the output layer of the wavelet neural network is:

$$\mathrm{y(k)} = \sum_{i=1}^{l} w_{ik} h(i) \quad k = 1,2,\dots,m \tag{12}$$

where w_{ik} is the weight from the hidden layer to the output layer, $h(i)$ is the output of the ith hidden layer node, l is the number of hidden layer nodes, m is the number of output layer nodes.

The wavelet neural network weight parameter correction algorithm is similar to the BP neural network weight correction algorithm. The gradient correction method is used to correct the network weights and wavelet basis function parameters, so that the predicted output of the wavelet neural network is constantly approaching the expected output.

2.4 Model Evaluation Indicators

In order to verify the prediction results of the precipitation by the wavelet divine general network, this paper selects three classical geostatistical indicators to evaluate, namely the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2). The formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - P_i^*)^2} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - P_i^*| \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - P_i^*)^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \tag{15}$$

where P_i is the observed value of precipitation, P_i^* is the predicted value, $\bar{P}$ is the average value of the actual precipitation.

3. Analysis of Results

3.1 Processing of Precipitation Anomalies

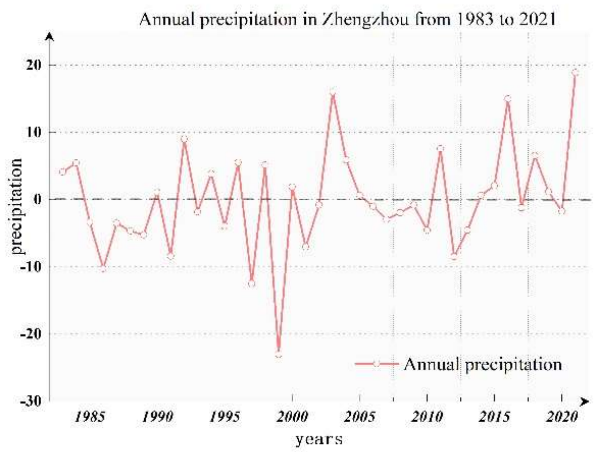

Fig. 2 Annual precipitation in Zhengzhou from 1983 to 2021

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

After analysis, it was found that there were too many missing data before 1983 in the statistics of the weather station. In order to improve the accuracy, we analyze the annual precipitation from 1983 to 2021 [5]. In order to make the data wavelet analysis reflect the precipitation cycle law more accurately, it is necessary to eliminate or reduce the "boundary effect" that may occur at both ends of the data series. The precipitation anomaly is processed first in the calculation, and the results are shown in Fig.2.

It can be clearly seen from Fig. 2 that the annual precipitation in Zhengzhou has strong non-stationarity and volatility, and the years with higher annual precipitation are: 1992, 2003, 2011, 2016, and 2021.

3.2 Analysis of Wavelet Transform

We have drawn the contour map of the wavelet mode, the contour map of the wavelet mode, the real part map of the wavelet coefficient, and the wavelet variance map of the precipitation in Zhengzhou from 1983 to 2021 [7], which is shown in Fig. 3 to Fig. 6.

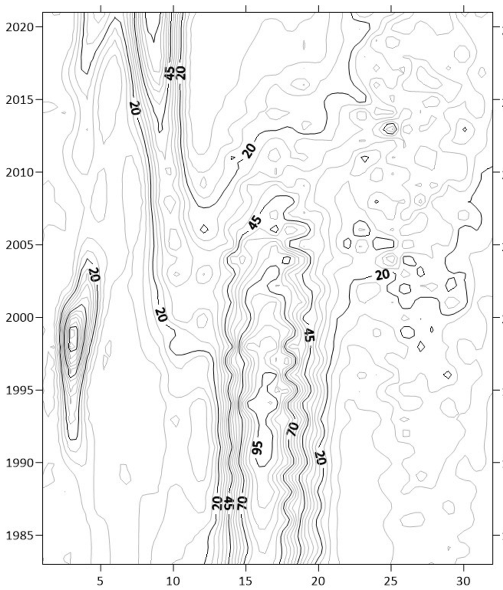

Fig. 3 Real part plot of wavelet coefficients

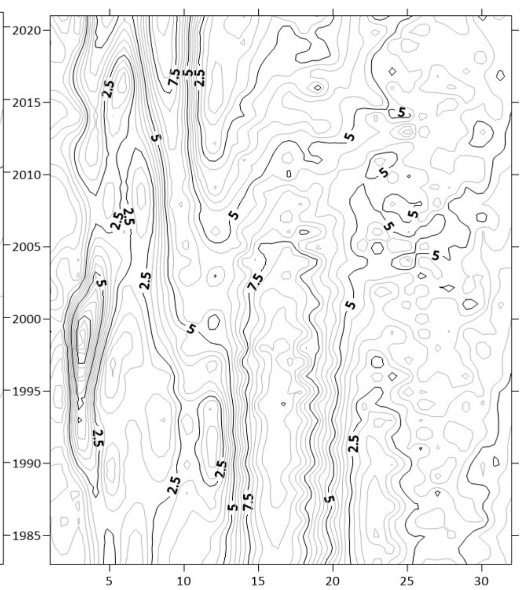

Fig. 4 Wavelet mode contour plot

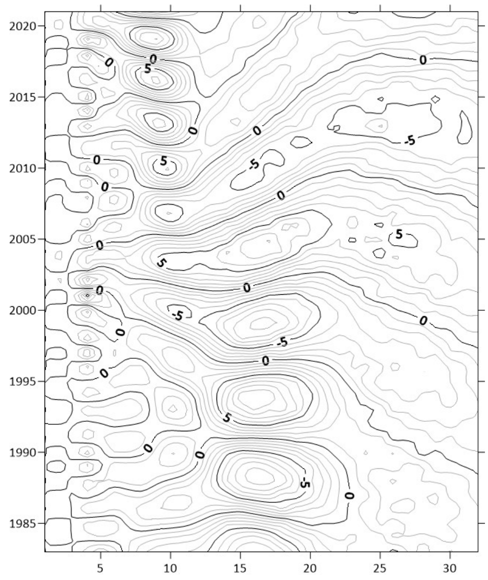

Fig. 5 Contour plot of wavelet modulus

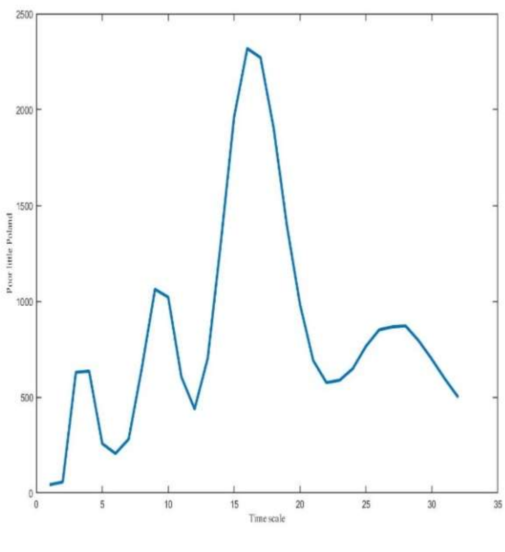

Fig. 6 Morlet wavelet variance plot

Volume 8 Issue 7, 2022
DOI: 10.6919/ICJE.202207_8(7).0012

In Fig.3, we draw it with a solid line, and when it is negative, it means a dry period, and it is drawn with a dashed line. Fig. 3 clearly shows the multi-time-scale features in the evolution of precipitation. In general, there are three types of periodic variation laws in the evolution of watershed runoff at 25-32 years, 10-25 years and 5-10 years. Among them, there are quasi-two oscillations of heavy precipitation and drought on the 25-32-year time scale, and quasi-four oscillations on the 10-25-year time scale. At the same time, it can also be seen that the periodic changes of the above two scales are very stable and global in the whole analysis period; while the periodic changes of the 5-10-year scale are relatively stable after 2000.

Fig. 4 shows the modulus contour map. The modulus value of the Morlet wavelet coefficient is the reflection of the energy density distribution in the time domain corresponding to the variation period of different time scales. It can be seen from Fig. 3 that in the evolution of annual precipitation, the 10-25-year time scale has the largest modulus value. It shows that the periodic change of this time scale is the most obvious, followed by the periodic change of the 25-32 year time scale, and the periodic changes of other time scales are smaller.

The modulus of the wavelet coefficient is equivalent to the wavelet energy spectrum, which can analyze the oscillation energy of different periods. As shown in Fig. 5, the time scale of 10-25 years has the strongest energy and the most significant period, but its periodic variation is localized (before 2005), although the energy of the time scale of 5-10 years is relatively weak, the periodic distribution is relatively obvious, occupying almost the entire research time domain (1983-2021).

From the wavelet variance diagram in Fig.6, it can be seen that 16 years is the highest peak of the time scale, which also indicates the strongest periodic oscillation. It can be seen that the main period of precipitation in Zhengzhou is 16 years, and the secondary main period is 9 years [1].

3.3 MK Mutation Test

We draw the MK mutation curve from 1983 to 2021, where the significance level $\alpha = 0.05$, and the corresponding critical value is ± 1.96. The MK statistic curve is shown in Fig.7.

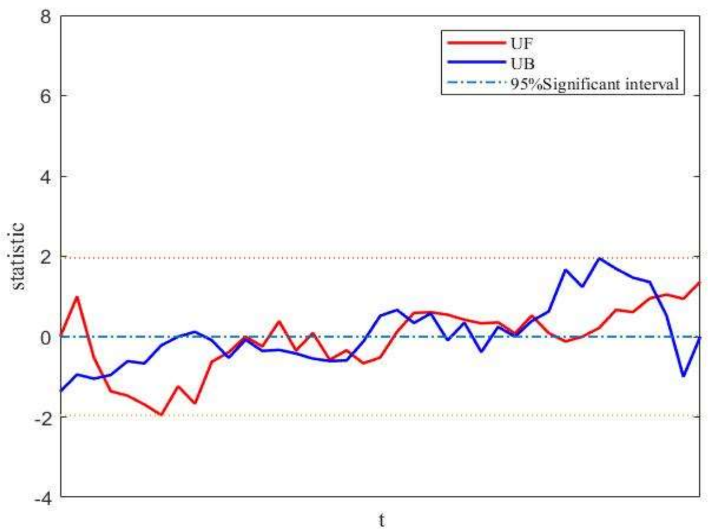

Fig. 7 MK mutation test of annual precipitation in Zhengzhou from 1983 to 2021

It can be seen from Fig. 7 that UF and UB have multiple intersections. Among them, from 1985 to 1986, and before and after the intersection of 2018 to 2019, UF and UB changed in different trends. According to the mutation principle of MK test, the intersection of UF and UB.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

It can be seen from the Fig.7 that during the period from 1983 to 2021, there were many abrupt changes in annual precipitation in Zhengzhou. The most recent change in annual precipitation was in 2019, and after that, the UF value was 0, indicating a continuous growth trend. The 0.05 significance level line indicates that it has passed the 0.05 significance test, so it will lead to heavy rainfall in Zhengzhou in 2021. And according to the cycle of the up and down trend, it is also in line with the analysis of the first main cycle in the wavelet analysis.

3.4 Precipitation Time Series Prediction

We use wavelet neural network [6], SVM support vector machine [3], BP neural network [2] to simulate and predict the annual precipitation in Zhengzhou from 1983 to 2021 after anomaly processing, and obtain The results are shown in Fig. 8, and the RMSE, MAE, and MAE of each model are calculated for the predicted results, and the results are analyzed. The conclusions are shown in table 1.

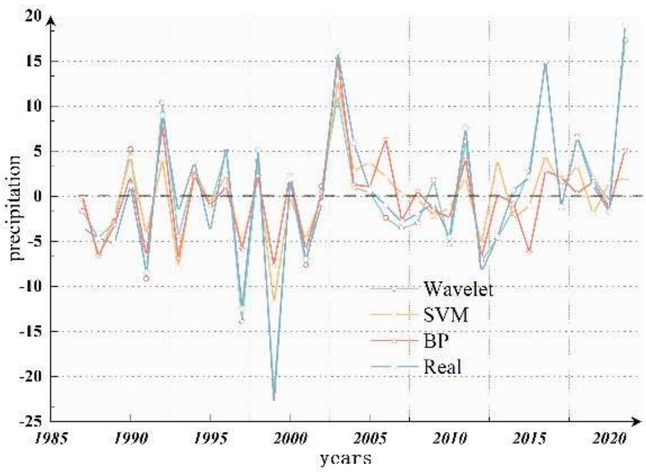

Fig. 8 Comparison of short-term forecasting methods for time series in Zhengzhou

Table 1. Model evaluation values

Model evaluation value	Wavelet neural network	BP neural network	SVM support vector machine
RMSE	1.9024	5.2328	5.2148
MAE	1.4458	3.6538	4.1308
R^2	0.9431	0.5692	0.5722

From Fig. 8, we can clearly see that the wavelet neural network time series prediction model has roughly the same trend as the actual precipitation, and the annual predicted precipitation value is close to the actual precipitation value; while the BP neural network is similar to the actual precipitation value. The change trend is roughly the same, but the error of some points is extremely high, and there is even an anti-real phenomenon that the rainy season becomes arid; the SVM support vector machine prediction and the actual precipitation prediction change trend is roughly the same, but the predicted precipitation change range is small, not very good predict extreme flooding or drought. So it can be seen intuitively that the wavelet neural network has high precision and high stability.

From table 1, we calculated the evaluation index values of the three models. It can be clearly seen that the wavelet neural network is the best in terms of RMSE, MAE, and other three evaluation indicators, reflecting its strong short-term time series prediction ability, the time correlation is high.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202207_8(7).0012

4. Conclusion

In this paper, wavelet transform is fully applied to the periodic analysis and time prediction of annual precipitation in Zhengzhou, and a wavelet neural network model is constructed to predict the time series of annual precipitation in Zhengzhou. The wavelet analysis shows that the main period of precipitation in Zhengzhou is 16 years and the secondary main period is 9 years, among the three evaluation indicators selected, the wavelet neural network time series prediction model (RMSE-2.5892, MAE-71.4605, R^2-0.8911) all showed, and the best performance is obtained, which indicates that the wavelet analysis model in this paper is a suitable tool to predict the precipitation in Zhengzhou.

In future research, we can try to input meteorological variables such as temperature, air pressure, relative humidity and wind speed into the model to further improve the precipitation prediction accuracy.

References

[1] Dai Cuixian, Jia Si, Wang Jun, Li Xiaolin. Multi-time scale analysis of annual precipitation series in Sanmenxia City based on wavelet transform [J]. Henan Science, 2022, 4(2): 270-275.

[2] Xiong Haijing, Wang Shicheng, Wang Dong. Application of BP neural network in annual precipitation forecast [J]. Proceedings of the 2012 Annual Academic Conference of Bengbu Science and Technology Association, 2021,1(1):135-141.

[3] Yu Haijiao, Wen Xiaohu, Feng Qi, He Zhibin. Simulation of daily precipitation-runoff in a typical small watershed in Qilian Mountains based on support vector machine (SVM) [J]. Chinese Journal of Water Resources and Water Engineering, 2015, 26(2): 27-31.

[4] Su Aifang, Lv Xiaona, Cui Liman, Li Zhou, Xi Le, Li Han. Basic observation and analysis of extreme rainstorm weather in Zhengzhou on July 20 [J]. Rainstorm Disaster, 2021, 40(5): 446-454.

[5] National Meteorological Science Data Center, http://data.cma.cn/.

[6] Chen Yiming, He Zijie, Benyue, Dong Qinghua. Time-frequency variation and correlation analysis of runoff and precipitation based on wavelet transform [J]. China Rural Water Resources and Hydropower, 2017, (10): 13-22.