Annual precipitation analysis and forecasting – taking Zhengzhou as an example

Wei Tuo, Xianqi Zhang, Chao Song, Dengkui Hu and Tianyi Liang

ABSTRACT

Precipitation is greatly affected by natural conditions and human activities, and its sequence has the characteristics of periodic variation and non-stationarity. In order to study the periodic characteristics of regional precipitation, reduce the non-stationarity of the water sequence and improve the forecasting accuracy, the Morlet wavelet model and CEEMD – Elman + ARIMA model are introduced, and applied to the precipitation sequence of Zhengzhou. The results show that the annual precipitation in Zhengzhou is mainly affected by periodic fluctuations of 55a, 40a and 15a, and the mean absolute error of the CEEMD – Elman + ARIMA model in 2013–2017 is 14.1%. Based on the analysis of precipitation period and forecast of future precipitation, the characteristics and evolution of precipitation in Zhengzhou are revealed. This provides a theoretical basis for rational utilization of water resources, prevention, industrial and agricultural production in Zhengzhou.

Key words | CEEMD – Elman + ARIMA, Morlet wavelet, precipitation, Zhengzhou

Wei Tuo
Xianqi Zhang (corresponding author)
Chao Song
Dengkui Hu
Tianyi Liang
School of Water Conservancy,
North China University of Water Resources and Electric Power,
Zhengzhou 450046,
China
E-mail: *zxqi@163.com*

Xianqi Zhang
Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering,
Zhengzhou 450046,
China

INTRODUCTION

China's total water resources are less than 3 trillion m^3, while the per capita water resources are only 0.22 million m^3, only 0.25 times of the global average (Li 2014). However, Zhengzhou has a large population base, and the total amount of water per capita is only 0.2 times of the average level of China, belonging to the area of extreme shortage of water resources. Precipitation is an important source of regional water resources and precipitation anomaly is the direct driver of regional flood and drought disasters. Analysis of precipitation periodic characteristics and precipitation forecasting can provide data support for rational utilization of regional water resources, flood control and drought relief, and have important guiding significance for social and economic development.

Many eastern and western scholars have carried out annual variation analysis of precipitation series in various regions, using the wavelet analysis method (Duan *et al.* 2018; Partal 2018; Zhao *et al.* 2019). The wavelet analysis method can not only clearly show the intensity distribution and precipitation change trend of various time-scales, but also analyze the main precipitation cycle of the region. In this paper, the wavelet transform is used to analyze the annual precipitation in Zhengzhou, and thus the periodic variation characteristics of precipitation in Zhengzhou were obtained.

Wavelet analysis mainly reflects the change of precipitation period on time-scales, but it cannot reflect the specific changes in the next few years. In order to further reflect the precipitation in the next few years, it is necessary to make a precise forecast of precipitation. Precipitation forecasting at home and abroad mainly focuses on remote sensing observation and mathematical models. In terms of remote sensing observation, the studies of Ganguly & Bras (2003), Diomede *et al.* (2008) and Pombo *et al.* (2015) are the most representative. In terms of mathematical models, many scholars mainly focus on the study of precipitation

doi: 10.2166/ws.2020.067

in the aspects of artificial neural networks (Teng *et al.* 2016; Boonyuen *et al.* 2019), regression analysis (Nayagam *et al.* 2008; Nicholson 2015) and combined models (Spiridonov *et al.* 2017; Xu & Wang 2019). However, according to the characteristics of the precipitation sequence itself, it is rare to build a coupled forecasting model by reducing the non-stationarity of the sequence. Precipitation evolution is a complex multidimensional system driven by many uncertain factors (Song *et al.* 2018; Li *et al.* 2019). The development trend and direction of regional precipitation forecasting are combined with modern nonlinear theory, which can effectively improve the forecasting accuracy. Complementary Empirical Mode Decomposition (CEEMD) can decompose non-stationary time-series into Intrinsic Modal Function (IMF) components and trend, and the amplitude, fluctuation and trend of the IMF components are smaller than the original sequence. Thus, the non-stationarity of the sequence can be reduced. Combined with the inherent advantages of the Elman neural network in dealing with nonlinear problems (Liu & Zhu 2019), the Autoregressive Integrated Moving Average model (ARIMA) has a strong approximation ability to stationary sequences (Liu *et al.* 2017). The CEEMD – Elman + ARIMA model is constructed, which provides a new method for precipitation forecasting.

METHODOLOGY

Wavelet periodic analysis

Morlet in 1984 proposed wavelet analysis with time-frequency multi-resolution function, which provided the possibility of better research on time-series problems (Goupillaud *et al.* 1984). It can clearly reveal various change cycles hidden in time-series, fully reflect the change trend of the system under different time-scales, and make qualitative estimation of the future development trend of the system.

The wavelet transform

Wavelet analysis mainly uses the localization characteristics of the wavelet function in the time domain and frequency domain to reveal the multi-level variation rules of hydrological phenomena. The wavelet function is expressed as follows:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0,\ \psi(t) \in L^2(R) \tag{1}$$

where $\psi(t)$ is the wavelet basis function, and its function system is expressed as follows:

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

where $\psi_{a,b}(t)$ is the daughter wavelet; a is the scale factor, which reflects the period length of the wavelet function; b is the displacement factor, which reflects the shift of the wavelet function in time.

If $\psi_{a,b}(t)$ is the subwavelet given by (2), for a wavelet function that satisfies certain conditions ψ(t), $L^2(R)$ means it is defined on the real axis, as a measurable square integrable function on the real axis. The wavelet transform of $f(t) \in L^2(R)$ is as follows:

$$W_f(a, b) = |a|^{-1/2} \int_R f(t)\overline{\psi}\left(\frac{t-b}{a}\right)dt\, a, b \in R \text{ and } a \neq 0 \tag{3}$$

where $W_f(a, b)$ is the wavelet transform coefficients; $f(t)$ is a square integrable function; a is an extension scale; b is a translation parameter; $\overline{\psi}\left(\frac{x-b}{a}\right)$ is the complex conjugate function of $\psi\left(\frac{x-b}{a}\right)$.

The measured time-series is usually a discrete signal, so the discrete wavelet transform of Equation (3) is:

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

In accordance with Equations (3) and (4), the basic principle of wavelet analysis is to obtain signal information at various frequencies by adjusting the extension scale a. Then the local information of the signal is analyzed so as to realize the analysis of time-scales and local spatial characteristics of different signals.

The wavelet square difference reflects the change of the fluctuation energy of the sequence with the time-scale, and can determine the main period of the evolution of the

time-series. The expression is as follows:

$$Var(a) = \int_{-\infty}^{\infty} |W_f(a, b)|^2 \, db \tag{5}$$

Wavelet variance

The square value of the wavelet coefficient is integrated in the b domain, and the wavelet variance is obtained, that is:

$$Var(a) = \int_{-\infty}^{\infty} |W_f(a, b)|^2 \, db \tag{6}$$

The change process of the wavelet squared with scale a is called the wavelet variance diagram. According to Equation (6), it can reflect the energy distribution of signal fluctuation with scale a. Therefore, the wavelet variance diagram can be used to determine the relative intensity of different scale disturbances in the signal and the main time-scale, namely the main period.

Morlet wavelet

This paper selects the Morlet wavelet for the periodic analysis, which is suitable for the evolution characteristics of precipitation. The expression of the Morlet wavelet is as follows:

$$\psi(t) = e^{ict} e^{-t^2/2} \tag{7}$$

where c is a constant, generally taken as 5.4; i is the imaginary number.

The relation between the scaling scale a of the Morlet wavelet and the period T is as follows:

$$T = \left(\frac{4\pi}{c + \sqrt{2 + c^2}} \right) \times a \tag{8}$$

The wavelet is a kind of continuous complex wavelet whose real part reflects the time distribution and phase information of the signal at different scales. In order to simplify the calculation, the precipitation time-series was processed centrally. The centralization process $f(k\Delta t)$ ($k = 1, 2 \ldots 67$; $\Delta t = 1$) and Morlet wavelet were substituted into Equation (4), and the value of a and b to calculate the wavelet transformation coefficient $W_f(a, b)$.

CEEMD

Yeh *et al.* (2010) proposed the Complementary Ensemble Empirical Mode Decomposition (CEEMD) for the poor completeness of EMD and EEMD. CEEMD mainly adds two pairs of opposite white noise into the original signal, and then EMD decomposes the original signal. The EMD decomposition results were integrated to obtain the final IMF component. The calculation steps of CEEMD are as follows.

(1) Add different white noise, decompose the original signal with EMD for N times, then IMF_1 is obtained:

$$IMF_1(t) = \frac{1}{N} \sum_{i=1}^{n} E_1(x(t) + \varepsilon_0 \omega^i(t)) \tag{9}$$

(2) Calculate the first-order residuals. The calculation formula is as follows:

$$r_1(t) = x(t) - IMF_1(t) \tag{10}$$

(3) Decompose $r_1(t) + \varepsilon_1 E_1(\omega^i(t))$, i=1, 2, 3 … n. The first intrinsic modal function is used as the IMF_2 of the CEEMD:

$$IMF_2(t) = \frac{1}{n} \sum_{i=1}^{n} E_1(r_1(t) + \varepsilon_1 E_1(\omega^i(t))) \tag{11}$$

(4) Calculate residuals $r_k(t)$ of order k. The calculation formula is as follows:

$$r_k(t) = r_{k-1}(t) - IMF_k(t) \tag{12}$$

(5) Decompose $r_k(t) + \varepsilon_k E_k(\omega^k(t))$. The first intrinsic modal function is used as the $IMF_{(k+1)}$ of the CEEMD:

$$IMF_{(k+1)}(t) = \frac{1}{n} \sum_{i=1}^{n} E_1(r_k(t) + \varepsilon_k E_k(\omega^k(t))) \tag{13}$$

(6) Repeat steps (4) to (5), until the residual cannot be decomposed. The final residual is defined as:

$$R(t) = x(t) - \sum_{k=1}^{K} IMF_k(t) \tag{14}$$

where K is the total number of $IMF(t)$. Then the original signal $x(t)$ can be expressed as follows:

$$x(t) = \sum_{k=1}^{K} IMF_k(t) + R(t) \tag{15}$$

It can be seen that the signal after CEEMD decomposition is complete, and the accuracy of the reconstructed signal is also guaranteed.

Elman neural network

The Elman neural network was proposed by Elman in the 1990s (Elman 1990), the network is a kind of feedback neural network composed of input layer, hidden layer, acceptor layer and output layer. Different from the BP neural network, the Elman neural network adds a connective layer in the hidden layer as a delay operator to realize the dynamic memory of the system. Because of the good memory and stability characteristics of the Elman neural network, it has been widely used in various fields (Jiang *et al.* 2016; Yates & Keedwell 2018; Jia *et al.* 2019; Wu & Xu 2019). Aiming at the advantages of the Elman neural network, it can also be applied to precipitation forecasting. The network structure is shown in Figure 1.

In Figure 1, input vector u is an r-dimensional vector; the output vector y is an m-dimensional vector; x is the n-dimensional output vector of the receiving layer; W^1 is the connection weight from the hidden layer to the receiving layer; W^2 is the connection weight from the input layer to the hidden layer; W^3 is the connection weight from the hidden layer to the output layer. The Elman network model is expressed as follows:

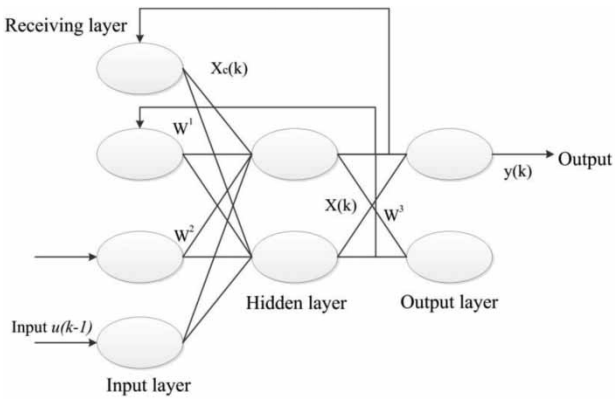

Figure 1 | The Elman neural network structure.

$$y(k) = g(w_2 x(k)) \tag{16}$$

$$x(k) = f(w_2 x_c(k) + w_2(u(k-1))) \tag{17}$$

$$x_c(k) = x(k-1) + a x_c(k-1) \tag{18}$$

where $g()$ is the activation function of the output neuron; $f()$ is the activation function of the hidden layer neurons; a is the self-connecting feedback gain factor, $0 \leq a < 1$. The network is a standard Elman network when $a = 0$; the network is a modified Elman network when $a \neq 0$.

The Elman network adopts the BP algorithm for weight correction, and the learning index function adopts the sum of squared errors. The expression is:

$$E(w) = \sum_{k=1}^{n} (y_k(w) - \hat{y}_k(w))^2 \tag{19}$$

where $\hat{y}_k(w)$ is the target output vector.

The Elman neural network learning algorithm adopts a momentum gradient descent back-propagation algorithm. The weights and thresholds are optimized by using the difference between the network output values and the output samples, to minimize the sum of squared errors of the network output layer.

ARIMA

Basic principles

Box *et al.* (1997) proposed the ARIMA model in the 1970s, which has been widely used in time-series analysis. The data can be processed smoothly and normally by studying the probability distribution of noise, thus solving the problem of sequence random disturbance.

The modeling ideas of ARIMA: the forecast object is regarded as a set of random series, which are approximately described by a certain mathematical model. Once the model is identified, the past and present values of the series can be used to forecast the future values. In the ARIMA (p, d, q)

model, AR is the autoregressive component (Zhao & Chen 2015); I is difference; MA is the moving average component; p is the order of the autoregressive component; d is the differential times; q is the order of the moving average component.

The ARIMA model is defined below:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} - \theta_1 \xi_{t-1} - \theta_2 \xi_{t-2} - \ldots - \theta_q \xi_{t-q} + \xi_t \quad (20)$$

where y is the time-series; $\phi_1, \phi_2 \ldots \phi_p$ are the autoregressive coefficients; $\theta_1, \theta_2 \ldots \theta_q$ are the moving average coefficients; ξ_t is the error series; p is the autoregressive order number ($p > 0$ and an integer); q is the autoregressive order number ($q > 0$, and an integer). If the differential order is represented by d, the model can be written as ARIMA (p, d, q).

CEEMD – ELMAN + ARIMA model

The sequence of precipitation can be decomposed into low-frequency and high-frequency components by CEEMD decomposition, thus reducing the non-stationarity of the precipitation sequence. For high-frequency components, the non-stationarity of the sequence is relatively large. If the linear model is used, the forecasting effect will be insufficient. Therefore, the Elman neural network suitable for dealing with nonlinear problems was selected for forecasting. The low-frequency components show good regularity (periodicity and linearity). In the case of maximum decomposition scale, the low-frequency component is close to the original sequence in numerical value and shows good stationarity. Therefore, the ARIMA model suitable for linear problems can be selected for forecasting. The precipitation forecasting technical route is shown in Figure 2.

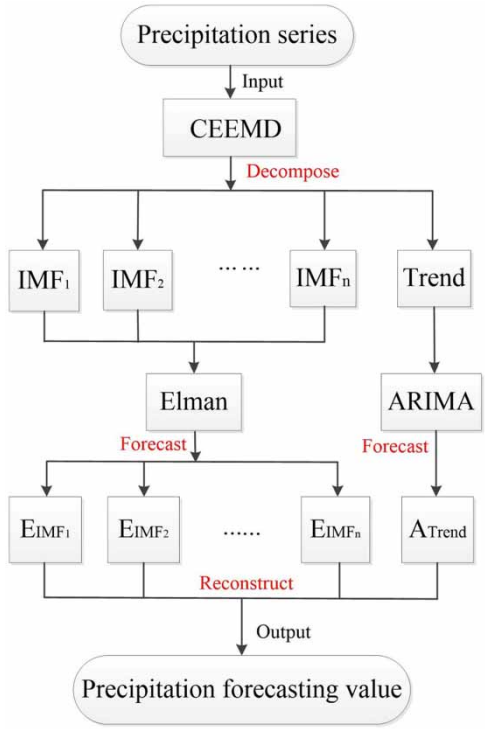

Figure 2 | The precipitation forecasting technical route.

CASE STUDY

The data source

Zhengzhou is located in the south of the north China plain, the lower Yellow River, and central Henan province in the north. Zhengzhou is affected by the temperate continental monsoon climate instability and the variability of the weather system, resulting in great variation in annual precipitation. The data comes from the national meteorological information center of China. The location as shown in Figure 3.

As can be seen from Figure 4, the annual precipitation series in Zhengzhou shows an irregular fluctuation trend with rising and falling, and the series shows a greater randomness. Precipitation variation is affected by regional topography, land and sea position, underlying surface, monsoon climate, etc, so the sequence presents non-stationarity. However, in some years, precipitation may be less affected by external factors, and the sequence shows certain linear characteristics. Therefore, CEEMD suitable for nonlinear processing is selected as the decomposition tool, the Elman network suitable for nonlinear modeling and ARIMA suitable for linear modeling as the forecasting tool.

Periodic characteristics of annual precipitation

The scale change of precipitation is mainly reflected by the wavelet real-part contour map of precipitation, where the

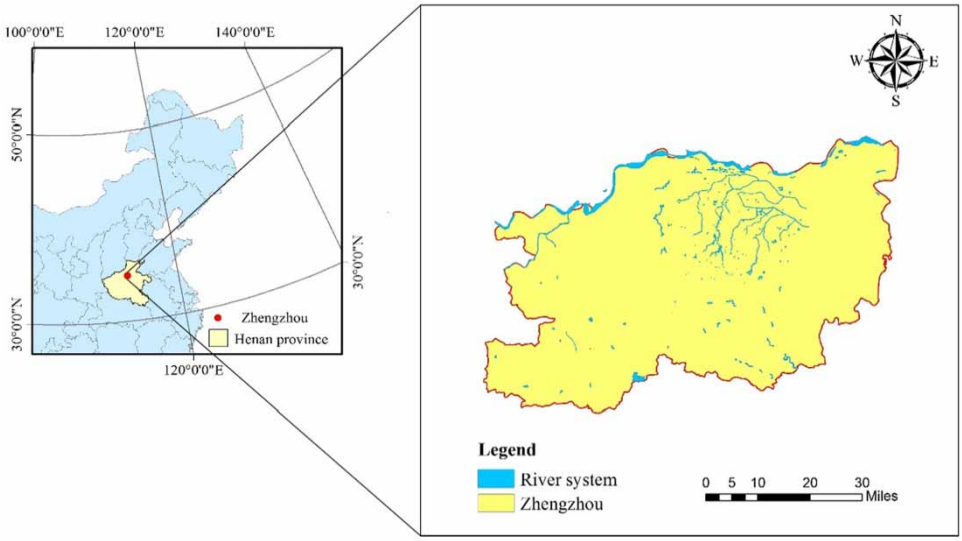

Figure 3 | The location of Zhengzhou.

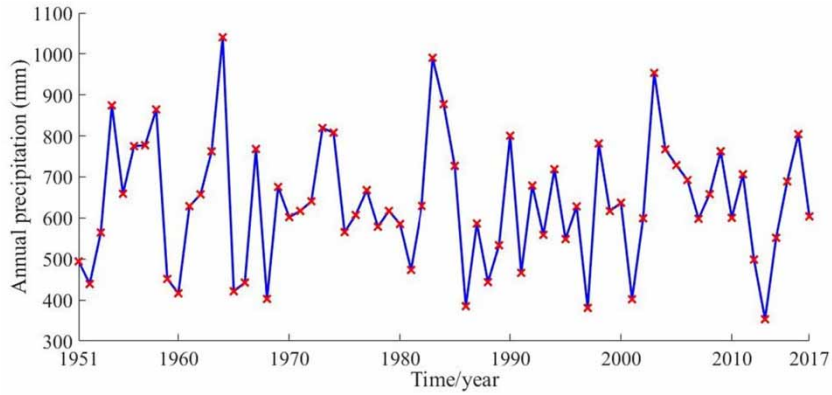

Figure 4 | Precipitation variation curve of Zhengzhou from 1951 to 2017.

contour value represents the real part of the wavelet coefficient. The wavelet coefficient is positive, indicates that the precipitation is heavy; the wavelet coefficient is negative, means the precipitation is low. The value of wavelet variance reflects the proportion of a certain time-scale in the whole time-series. The larger the variance value, the stronger the oscillation, and the more significant the periodic change of the corresponding time-scale. The smaller the variance value, the less significant the periodic change of the corresponding time-scale. The wavelet series numerical modulus reflects the energy density corresponding to different time-scales. The larger the numerical modulus of the wavelet series, the stronger the periodicity of the time-scale, and vice versa.

Wavelet coefficients are calculated by MATLAB. After obtaining the wavelet coefficient, Surfer 12.0 software was used to make the contour map. The results are shown in Figure 5.

As can be seen from Figure 5, there are various time-scale structures in the precipitation series from 1951 to 2017, which are roughly 6a–20a, 30a–47a and 48a–64a. Among them, the corresponding wavelet coefficients of 48a–64a are large, the period is obvious, and the oscillation

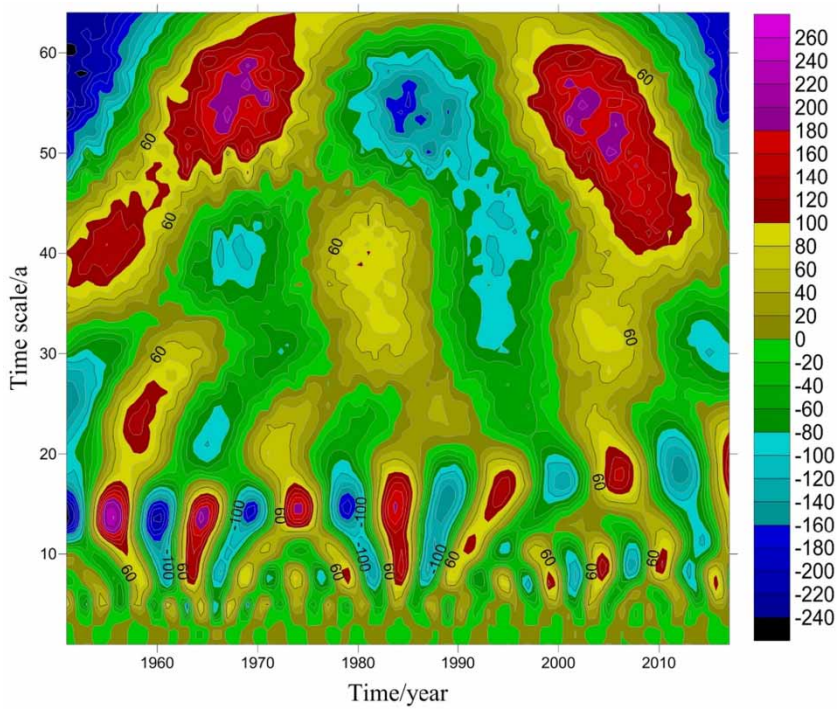

Figure 5 | Real contour map of annual precipitation wavelet coefficients in Zhengzhou.

center is 55a. Precipitation changes on the central scale of 55a show the characteristics of less-more-less-more-less. The oscillation center is 40a at the scale of 30a–47a, and the variation characteristics of precipitation are more-less-more-less-more-less. The oscillation center is 15a on the scale of 6a–20a, and the precipitation changes have undergone several alternation cycles of less and more. In addition, from the perspective of the 55a scale, the contour line of the real part of the wavelet of precipitation in 2017 is not closed. It can be seen that the precipitation after 2017 will show a decreasing trend, but it is unknown how much the value is reduced. Therefore, the CEEMD – Elman + ARIMA model is used to forecast precipitation after 2017.

It can be seen from Figure 5 that precipitation in Zhengzhou has a certain characteristic of periodic change. However, it is puzzling that Figure 5 reflects the range of periodic changes, and the specific periodic value of the dominant position, which we cannot know. Therefore, the water variance is calculated, as shown in Figure 6.

As shown in Figure 6, there are three large peak points of annual precipitation in Zhengzhou, which indicates that there are three main cycles of annual precipitation. Among them, the three main cycles are 55 years, 15 years and eight years respectively, and the fluctuation of these three cycles controls the variation characteristics of precipitation in the whole time domain.

Figure 7 shows the evolution of annual precipitation in Zhengzhou. The wavelet series corresponding to 48a–64a has the largest numerical modulus value, indicating that the cycle change is the most obvious. The mode value of the wavelet coefficient corresponding to 6a–20a is followed by its periodic change. The wavelet coefficients of 21a–47a are the smallest and the corresponding periodic changes are the weakest.

The above periodic analysis is to forecast the future precipitation trend from the perspective of qualitative analysis. Therefore, from the perspective of quantitative analysis, the paper also predicts the precipitation value. That is, the CEEMD – Elman +ARIMA model is used to forecast precipitation in 2018–2022.

Decomposition of annual precipitation

The sequence of precipitation is decomposed into subsignals of different frequencies, namely IMF components and trend. The complex precipitation forecasting is equal

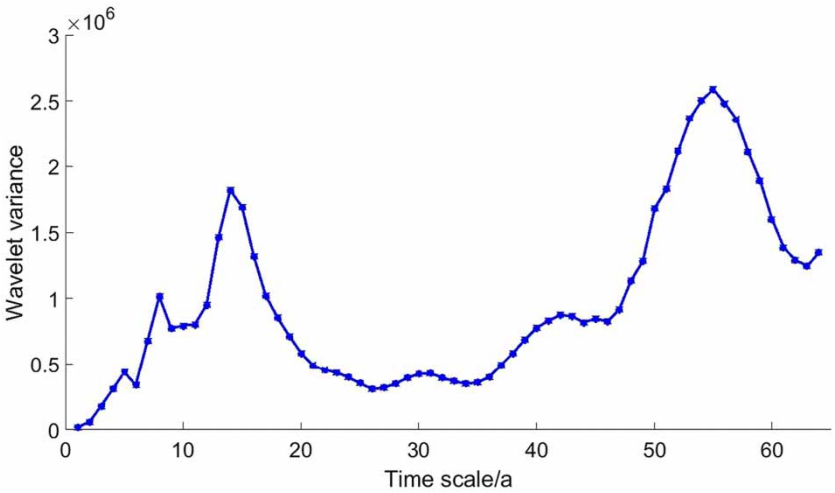

Figure 6 | Wavelet variance of annual precipitation in Zhengzhou.

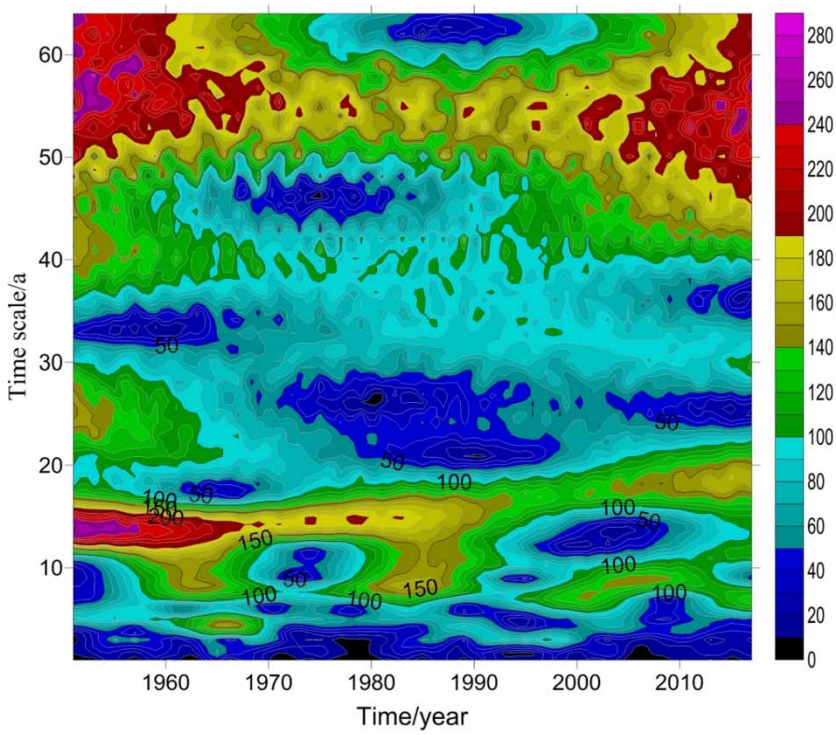

Figure 7 | Wavelet coefficient module contour map of annual precipitation in Zhengzhou.

to the sum of the forecasting values of different frequency subcomponents. By calculating the relative error of the sub-signal, the contribution rate of each sub-signal to the precipitation sequence can be analyzed, and the relative error of the sub-signal can be explained, whether it affects the relative error of precipitation.

After repeated testing, when the maximum decomposition is 5, the noise logarithm is 100 and the noise amplitude is 0.2. CEEMD has the best decomposition effect on precipitation, as shown in Figure 8.

As can be seen from Figure 8, the precipitation series is decomposed into four IMF components and one trend.

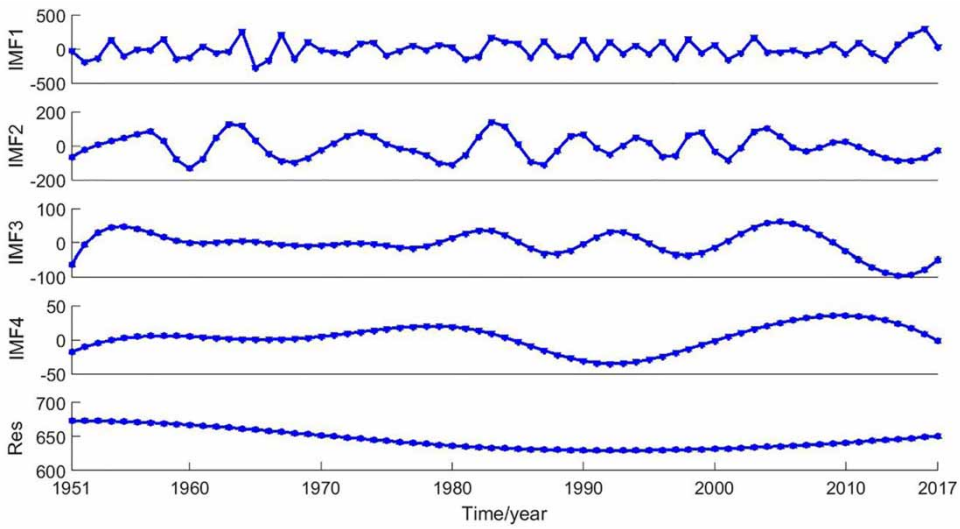

Figure 8 | CEEMD decomposition of annual precipitation in Zhengzhou.

Among them, the IMF_1 component has the lowest stationary, which is manifested as larger volatility, higher frequency and shorter wavelength. The amplitude and frequency of other the IMF components gradually decreases and the wavelength gradually increases. In addition, the change amplitude of the IMF component is significantly lower than the original series. Therefore, CEEMD decomposition greatly reduces the volatility and tendency of the sequence.

Precipitation forecasting

The sub-components obtained by CEEMD decomposition were divided into two parts for forecasting. For high-frequency components, the non-stationarity of the sequence is relatively large, so the Elman neural network suitable for dealing with nonlinear problems is selected for forecasting. For the low-frequency component (trend), it is close to the original sequence in value and has good stationarity. Therefore, the ARIMA model suitable for linear problems can be selected for prediction. In order to verify the forecasting accuracy of the model, the precipitation from 2013 to 2017 is forecast and error analysis is carried out. When the model is feasible, the future precipitation from 2018 to 2022 is forecast, so as to further study the evolution trend of precipitation.

Component forecasting

The Elman neural network was used to forecast the four nonlinear IMF components. The trend has an obvious linear characteristic, and the ARIMA was used to forecast the trend. After a lot of calculation, the ARIMA model corresponding to the trend is ARIMA(1,1,13). The forecasting effect of each component is shown in Table 1.

As described in Table 1, the forecasting error of the IMF_1 component is relatively high, while the forecasting error of the other components is relatively small, and the forecasting effect of the trend is the best. IMF_4 and the trend are the most stable, while IMF_1 and IMF_2 are relatively weak, although the forecasting error of IMF_1 and IMF_2 is relatively high. Due to the small proportion of IMF_1 and IMF_2 in the precipitation series from 1951 to 2017, this does not affect the forecasting error of precipitation.

CEEMD – Elman + ARIMA forecasting

The forecasting value of precipitation is equal to the sum of the four IMF component forecasting and the trend forecasting. The forecasting results of precipitation in Zhengzhou are shown in Table 2.

Table 2 shows that the CEEMD – Elman + ARIMA model has a good forecasting effect. The mean square error

Table 1 | Forecasting effect of IMF components and trend

Decomposition signal	Year	True value	Forecasting value	Relative error (%)	Mean relative error (%)
IMF_1	2013	−161.54	−100.00	−38.10	70.46
	2014	65.33	72.80	11.43	
	2015	207.10	100.81	−51.32	
	2016	295.76	120.00	−59.43	
	2017	31.51	−29.00	−192.02	
IMF_2	2013	−71.02	−46.76	−34.16	21.75
	2014	−87.70	−63.83	−27.22	
	2015	−88.72	−71.85	−19.02	
	2016	−69.69	−53.00	−23.95	
	2017	−26.22	−25.06	−4.41	
IMF_3	2013	−87.89	−80.98	−7.85	15.41
	2014	−95.93	−83.90	−12.55	
	2015	−93.87	−78.11	−16.79	
	2016	−79.39	−63.84	−19.58	
	2017	−50.17	−40.00	−20.27	
IMF_4	2013	28.67	28.48	−0.68	14.12
	2014	23.56	23.52	−0.19	
	2015	16.85	17.05	1.17	
	2016	8.44	9.00	6.62	
	2017	−1.79	−0.68	−61.96	
Trend	2013	644.98	644.98	0.00	0.04
	2014	646.34	646.33	0.00	
	2015	647.74	647.73	0.00	
	2016	649.18	649.15	−0.01	
	2017	650.66	650.60	−0.01	

Table 2 | Forecasting results of CEEMD – Elman + ARIMA model of annual precipitation in Zhengzhou from 2013 to 2017

Year	True value (mm)	Forecasting value (mm)	Absolute error (%)	Relative value (%)	MAPE (%)	RMSE
2013	353.20	445.71	92.51	26.19	14.1	87.86
2014	551.60	594.92	43.32	7.85		
2015	689.10	615.63	−73.47	10.66		
2016	804.30	661.31	−142.99	17.78		
2017	604.00	555.86	−48.14	7.97		

is only 87.86 and the average relative error is only 14.1%. The relative error of the forecasting in 2013 is relatively high, 26.19%, but the overall forecasting effect is feasible.

When the CEEMD – Elman + ARIMA model was verified to be feasible, the precipitation in 2018–2022 was forecast, and the forecasting results are shown in Table 3.

As can be seen from Table 3, the precipitation in Zhengzhou showed a downward trend from 2018 to 2020. Precipitation has increased slightly since 2021, but annual precipitation forecasts for 2018–2022 are still generally lower than in 2017. Due to the lack of real precipitation values after 2017 as a reference, the forecast values of 2018–2020 are highly reliable, while the forecast value error of 2021 and 2022 may be large. Therefore, it is reasonable to conclude from the wavelet cycle periodic analysis that the precipitation will decline in the next few years starting from 2017.

Table 3 | Forecasting results of CEEMD – Elman + ARIMA model of annual precipitation in Zhengzhou from 2018 to 2022

Year	IMF_1	IMF_2	IMF_3	IMF_4	Trend	Forecasting value (mm)
2018	−93.46	36.35	−11.84	−11.89	650.29	569.45
2019	−152.77	16.43	46.90	−8.29	651.31	553.57
2020	−124.27	9.46	1.25	−1.40	650.15	535.20
2021	−123.66	21.45	−5.68	1.44	652.40	545.94
2022	−79.94	−4.39	0.72	0.84	653.07	570.30

DISCUSSION

In the introduction, a variety of forecasting methods are mentioned. Among them, Empirical Mode Decomposition (EMD) and the Elman combined model (EMD–Elman), Markov and ARIMA combined model (Markov–ARIMA), Back Propagation neural network (BPNN) and Weather Research and Forecasting (WRF) and Cloud theory (WRF-Cloud) are representative. To further assess the capabilities of the proposed model, the simulation results of several models in precipitation forecasting are compared. Due to the limited grasp of various forecasting methods, only the simulation results of CEEM – Elman + ARIMA, EMD–Elman, Elman and Markov–ARIMA are compared. If there is time, the next step will be a systematic and comprehensive comparison of the advantages of various forecasting methods. The forecasting errors of the four models were made, as shown in Table 4.

As can be seen from Table 4, the relative error and average error of CEEMD – Elman + ARIMA are lower than that of Markov–ARIMA, EMD–Elman and Elman, indicating that CEEMD – Elman + ARIMA is more advantageous in precipitation forecasting. Compared with the other forecasting methods, the decomposition–reconstruction forecasting model can reduce the forecasting error. The reasons are as follows: the original signal is decomposed into different frequency components by the decomposition tools (CEEMD, EEMD, EMD), and the forecasting value of precipitation is equal to the sum of the forecasting values of the high-frequency component and low-frequency component. That is, the overall forecasting error of the model is determined by the forecasting error

Table 4 | The forecasting error of the three models

Time (year)	True value (mm)	CEEMD – Elman + ARIMA	EMD–Elman	Elman	Markov–ARIMA
Forecasting value (mm)					
2013	353.2	445.71	448.13	460.83	450.7
2014	551.6	594.92	505.67	490.26	501.23
2015	689.1	615.63	588.46	570.65	580.56
2016	804.3	661.31	657.25	650.85	647.85
2017	604	555.86	698.87	705.68	710.68
Relative error (%)					
2013	/	26.19	26.88	30.47	27.60
2014	/	7.85	−8.33	−11.12	−9.13
2015	/	−10.66	−14.60	−17.19	−15.75
2016	/	−17.78	−18.28	−19.08	−19.45
2017	/	−7.97	17.69	16.83	17.66
Mean relative error (%)					
2013–2017	/	14.09	16.76	18.94	17.92

of these high- and low-frequency components. In this way, even if the forecasting effect of some components is not good, the overall forecasting effect will not be affected.

CEEMD is proposed on the basis of EMD, and the model constructed adopts different forecasting methods for high- and low-frequency components. Therefore, the forecasting effect of CEEMD – Elman + ARIMA is more advantageous than the single decomposition–reconstruction forecasting model, such as EMD–Elman. Compared with the forecasting model without decomposition, the forecasting effect of Elman is determined by the original data itself. If the forecasting effect of Elman is not good, it will directly affect the whole error. Markov–ARIMA averages the forecasting results of the ARIMA and Markov model, which reduces the forecasting error to some extent, but the forecasting accuracy is lower than that of the decomposition–reconstruction forecasting model. In conclusion, CEEMD – Elman + ARIMA is more suitable for precipitation forecasting, and has more advantages and stability.

CONCLUSIONS

Wavelet analysis was applied to the annual precipitation cycle analysis of Zhengzhou from 1951 to 2017, and it was concluded that the annual precipitation of Zhengzhou was mainly affected by scale fluctuations of 55a, 40a and 15a. Furthermore, the forecasting of precipitation in the future years of 2018–2022 was carried out, and from the qualitative and quantitative perspective, the precipitation of Zhengzhou will show a downward trend from 2017.

The non-stationarity of precipitation time-series is reduced by CEEMD decomposition, and Elman and ARIMA are used to forecast the data after decomposition. The CEEMD – Elman + ARIMA forecasting model was constructed and applied to Zhengzhou precipitation forecasting, and the average relative error of the forecasting in 2013–2017 was 14.1%.

This is a new attempt to apply the CEEMD – Elman + ARIMA coupling model to the forecasting of precipitation in Zhengzhou. Its extension and the improvement of model accuracy still need further study. The model does not consider the physical mechanism of precipitation evolution and the long-term forecasting. Furthermore, how to deal with the subcomponent properly and improve the forecasting accuracy of the model are the next research direction and emphasis.

ACKNOWLEDGEMENTS

This work is financially supported by Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Henan Province, Water Environment Governance and Ecological Restoration Academician Workstation of Henan Province, Program for Science & Technology Innovation Talents in Universities of Henan Province (No. 15HASTIT049). Our gratitude is also extended to reviewers for their efforts in reviewing the manuscript and their very encouraging, insightful and constructive comments.

REFERENCES

Boonyuen, K., Kaewprapha, P., Weesakul, U. & Srivihok, P. 2019 Convolutional neural network inception-v3: a machine learning approach for leveling short-range rainfall forecast model from satellite image. In: *Advances in Swarm Intelligence* (Y. Tan, Y. H. Shi & B. Niu, eds), Springer, Cham, Switzerland, pp. 105–115.

Box, G. E. P., Jenkins, G. M. & Reinsel, C. 1997 *Time Series Analysis – Forecasting and Control*. China Statistical Press, Beijing, China, pp. 64–65.

Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi, A. & Malguzzi, P. 2008 Discharge prediction based on multi-model precipitation forecasts. *Meteorology and Atmospheric Physics* **101** (3–4), 245–265.

Duan, N., Yang, G. Y. & You, J. J. 2018 Temporal and spatial evolution of precipitation in Wuzhishan City in recent 60 years. *South-to-North Water Transfers and Water Science & Technology* **16** (6), 60–67.

Elman, J. L. 1990 Finding structure in time. *Cognitive Science* **14**, 179–211.

Ganguly, A. R. & Bras, R. L. 2003 Distributed quantitative precipitation forecasting using information from radar and numerical weather prediction models. *Journal of Hydrometeorology* **4** (6), 1168–1180.

Goupillaud, P., Grossmann, A. & Morlet, J. 1984 Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* **23** (1), 85–102.

Jia, Z., Guo, Q. J. & Hao, Q. W. 2019 Deformation prediction of deep foundation pit based on Elman–Markov model. *Yangtze River* **50** (1), 202–206 + 219.

Jiang, S. J., Fu, G. F., Huang, L. Y., Tang, Y. C., Cai, P. & Peng, H. Y. 2016 Forecasting on ash fusion temperatures of bituminous coal and biomass co-firing based on Elman neural network. *Journal of Central South University (Science and Technology)* **47** (12), 4240–4247.

Li, H. T. 2014 *Research on the Prediction of Rainfall in Irrigation Field Based on Combined Model*. Master's thesis, Northeast Agricultural University, Harbin, China.

Li, Y. K., Ma, X., Pan, X. Y., Bai, T., Di, S. C. & Huang, Q. 2019 Application of BP neural model based on grain in ear to medium and long term wet-season rainfall forecasting. *Soil and Water Conservation China* **17** (3), 1–6.

Liu, Z. Y. & Zhu, C. L. 2019 IGBT life prediction based on Elman neural network model. *Semiconductor Technology* **44** (5), 395–400.

Liu, Y., Yang, Y., Nie, L. & Song, Q. Y. 2017 The EEMD–ARIMA prediction of runoff at mountain pass of Manas river. *Research of Soil and Water Conservation* **24** (6), 273–280 + 285.

Nayagam, L. R., Janardanan, R. & Mohan, H. S. R. 2008 An empirical model for the seasonal prediction of southwest monsoon rainfall over Kerala, a meteorological subdivision of India. *International Journal of Climatology* **28** (6), 823–831.

Nicholson, S. E. 2015 The predictability of rainfall over the Greater Horn of Africa. Part II: prediction of monthly rainfall during the long rains. *Journal of Hydrometeorology* **16** (5), 2001–2012.

Partal, T. 2018 Wavelet based periodical analysis of the precipitation data of the Mediterranean region and its relation to atmospheric indices. *Modeling Earth Systems and Environment* **4** (4), 1309–1318.

Pombo, S., de Oliveira, R. P. & Mendes, A. 2015 Validation of remote-sensing precipitation products for Angola. *Meteorological Applications* **22** (3), 395–409.

Song, F., Yang, X. H., Wu, F. F. & Liu, T. 2018 Precipitation prediction using clustering-fuzzy-Markov chain model. *Water Saving Irrigation* **278** (10), 33–36.

Spiridonov, V., Baez, J. & Telenta, B. 2017 Heavy convective precipitation forecast over Paraguay using coupled WRF-cloud model. In: *Perspectives on Atmospheric Sciences* (T. Karacostas, A. Bais & P. T. Nastos, eds), Springer International Publishing, Cham, Switzerland, pp. 183–189.

Teng, S. H., Tang, H. T., Zhang, W., Liu, D. N. & Liang, L. 2016 Identifying local rainfall type and forecasting rainfall quantity based on mixed multiple PNN and RBF neural network models. *Journal of Chinese Computer Systems* **37** (11), 2571–2576.

Wu, M. M. & Xu, J. X. 2019 A short-term stock prediction model using an improved Elman neural network based on EMD. *Computer Engineering & Science* **41** (6), 1119–1127.

Xu, M. R. & Wang, X. M. 2019 Prediction of regional precipitation based on Markov and ARIMA combined model. *Computer Application and Software* **36** (3), 34–37.

Yates, W. B. & Keedwell, E. C. 2018 Offline learning for selection hyper-heuristics with Elman networks. In: *Artificial Evolution* (E. Lutton, P. Legrand, P. Parrend, N. Monmarché & M. Schoenauer, eds), Springer, Cham, Switzerland, pp. 217–230.

Yeh, J. R., Shieh, J. S. & Huang, N. E. 2010 Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis* **2** (2), 135–156.

Zhao, X. H. & Chen, X. 2015 Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. *Water Resources Management* **29**, 2913–2926.

Zhao, H., Zhang, M. S. & Pan, R. 2019 Prediction and analysis of future rainfall change in pilot project sponge city based on wavelet analysis. *Water & Wastewater Engineering* **55** (9), 29–35.

First received 4 July 2019; accepted in revised form 1 April 2020. Available online 20 April 2020