

Application of MEEMD-ARIMA combining model for annual runoff prediction in the Lower Yellow River

Xianqi Zhang, Wei Tuo and Chao Song

ABSTRACT

The prediction of annual runoff in the Lower Yellow River can provide an important theoretical basis for effective reservoir management, flood control and disaster reduction, river and beach management, rational utilization of regional water and sediment resources. To solve this problem and improve the prediction accuracy, permutation entropy (PE) was used to extract the pseudo-components of modified ensemble empirical mode decomposition (MEEMD) to decompose time series to reduce the non-stationarity of time series. However, the pseudo-component was disordered and difficult to predict, therefore, the pseudo-component was decomposed by ensemble empirical mode decomposition (EEMD). Then, intrinsic mode functions (IMFs) and trend were predicted by autoregressive integrated moving average (ARIMA) which has strong ability of approximation to stationary series. A new coupling model based on MEEMD-ARIMA was constructed and applied to runoff prediction in the Lower Yellow River. The results showed that the model had higher accuracy and was superior to the CEEMD-ARIMA model or EEMD-ARIMA model. Therefore, it can provide a new idea and method for annual runoff prediction.

Key words | annual runoff, ARIMA, Lower Yellow River, MEEMD, prediction

Xianqi Zhang

Wei Tuo (corresponding author)

Chao Song

School of Water Conservancy,
North China University of Water Resources and
Electric Power,
Zhengzhou 450046,
China
E-mail: 1181309403@qq.com

Xianqi Zhang

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

INTRODUCTION

One of the important tasks of hydrologists and water resource engineers is to assess and predict the quantity of water available in a basin over longer periods, for example, months and years, and manage the resource for practical applications involving conservation, environmental disposal, and efficient water supply (Wang *et al.* 2013). The prediction of annual runoff in the Lower Yellow River can provide an important theoretical basis for flood control and disaster reduction, river and beach management, rational utilization of regional water and sediment resources. To the best of our knowledge, the runoff has changed greatly in the Lower Yellow River with the influence of climate change and human activities (Li *et al.* 2009; Zhao *et al.* 2015). The evolution of runoff is a complex system, with randomness, ambiguity, and uncertainty. All this makes it very difficult to predict

runoff in the Lower Yellow River. In recent years, some scholars have done a great deal of research on runoff prediction and achieved fruitful results. Since hydrological models require a great deal of data input, many parameters need to be determined, and parameter calibration is difficult. Hydrological models should be specifically analyzed in different regional conditions, which may be feasible in one region and need to be considered in another region. Relevant western research has mainly focused on mathematical models to predict river or regional runoff: for example, Sedki *et al.* (2009) used a neural network based on a real coding genetic algorithm to predict daily runoff in the semi-arid climate catchment area of Morocco; Coulibaly *et al.* (2015) used a circular neural network based on low-frequency climate change index and predicted the annual runoff in the northern

province of Quebec; Mahabir *et al.* (2010) used fuzzy logic to predict seasonal runoff in rocky and middle stream basins; Alizadeh *et al.* (2017) used wavelet neural network to predict rainfall and runoff in the Nettolt river basin for the following two months. However, research on runoff prediction in the Yellow River has mainly taken place in China. Pan *et al.* (2017) used the GM(1.1) model to predict annual runoff and precipitation in the Lower Yellow River; Zhang *et al.* (2008) used the life cycle–Markov combination model to predict the annual runoff of Longmen hydrographic station in the Yellow River; Zhao *et al.* (2001) used the Fletcher-Reeves method to improve the backpropagation algorithm and predicted the runoff of Huayuankou and Lijin hydrological stations in the Lower Yellow River; Tu *et al.* (2018) used PSO-KELM to predict the annual runoff of Lanzhou station and the Jingou River. The above research mainly focuses on the traditional statistical models such as neural networks, etc. The traditional mathematical statistics model cannot predict the high frequency mutation data well; a neural network has the defect of training transition, which makes the network deviate too far from training data. Hydrological time series prediction is one of the most important applications in modern hydrology. The establishment of coupling a prediction model by reducing the non-stationarity of runoff series is a new way to improve the runoff prediction accuracy in the Lower Yellow River. The complementary ensemble empirical mode decomposition (CEEMD) can decompose the non-stationary time series and reduce the non-stationarity of the series, but there are also defects of pseudo-components. Using the PE to improve the CEEMD method, proposed modified ensemble empirical mode decomposition (MEEMD), the pseudo-components of IMF were extracted. Because the pseudo-component is very unstable, it cannot achieve a good prediction effect. Therefore, the pseudo-component is decomposed by ensemble empirical mode decomposition (EEMD), and the series after decomposition presents good stationarity. Combined with ARIMA for stationarity the series has a strong ability of approximation, and the MEEMD-ARIMA model for annual runoff prediction in the Lower Yellow River is constructed, so as to provide new ways for runoff prediction.

METHODOLOGY

CEEMD

Yeh & Shieh (2010) proposed CEEMD on the basis of EEMD (Wu 2009). The CEEMD method mainly adds two opposite white noise signals to the signal to be analyzed and the empirical mode decomposition (EMD) decomposition, respectively. The reconstruction error caused by white noise is reduced under the condition that the effect of EEMD decomposition is equal to that of EEMD decomposition. Although CEEMD solves the problem of poor completeness of EEMD, signal decomposition relies on the selection of noise amplitude and integration times, and pseudo-components (components that do not have physical significance) will appear (Zheng *et al.* 2012). Therefore, the PE is proposed to improve it and extract the IMF pseudo-component of CEEMD decomposition.

Permutation entropy algorithm

To improve the CEEMD, the crucial step is to detect the randomness of the signal (that is, extract the pseudo-component of the IMF component). There are many existing detection methods (Borgnat *et al.* 2010; Terrien *et al.* 2011), but they all have deficiencies in extracting pseudo-components.

The PE algorithm was proposed by Bandt & Pompe (2002). It is a theory based on physical mechanism and reflecting the inherent law of system evolution. Good robustness (Cao *et al.* 2004) is demonstrated in the detection of randomness and mutation of non-linear time series. This point has inherent advantages for the detection of signal randomness and the extraction of pseudo-components. The calculation of the permutation entropy is as follows (Bandt & Pompe 2002; Yan & Gao 2007):

1. Phase space reconstruction was carried out for time series of length $\{x(i), i = 1, 2, \dots, N\}$, and the following series can be obtained:

$$\begin{bmatrix} X(1) = \{x(1), x(1 + \lambda), \dots, x(1 + (m - 1)\lambda) \\ X(2) = \{x(2), x(2 + \lambda), \dots, x(2 + (m - 1)\lambda) \\ \dots\dots\dots \\ X(K) = \{x(K), x(K + \lambda), \dots, x(K + (m - 1)\lambda)\} \end{bmatrix} \quad (1)$$

2. Regard each row in the matrix as a refactoring component, m and λ represent the embedding dimension and time delay, where $K = (m - 1)\lambda$.
3. Arrange each component in order from small to large:

$$X(i) = \{x(i + (j_1 - 1)\lambda) \leq x(i + (j_2 - 1)\lambda) \leq \dots x(i + (j_m - 1)\lambda)\} \quad (2)$$

4. If $x(i + (j_{i1} - 1)\lambda) = x(i + (j_{i2} - 1)\lambda)$, it can be arranged in order of j from small to large. If $j_{i1} < j_{i2}$, take $x(i + (j_{i1} - 1)\lambda) = x(i + (j_{i2} - 1)\lambda)$, then, for any vector- $X(i)$, the following series of symbols can exist:

$$S(k) = [j_1, j_2, j_3 \dots j_m] \quad (3)$$

where, $k = 1, 2, 3 \dots l$, $l \leq m!$. Because $[j_1, j_2, j_3 \dots j_m]$ has $m!$ permutations, there are $m!$ series of symbols. Suppose the probability of each series of symbols is P_i and $\sum_{i=1}^l P_i = 1$.

According to Shannon's definition of information entropy, the permutation entropy of series $\{x(i), i = 1, 2, \dots, N\}$ can be defined below:

$$H_{PE}(m) = - \sum_{i=1}^l P_i \ln P_i \quad (4)$$

$H_{PE}(m)$ gets the maximum $\ln(m!)$ when $P_i = \frac{1}{m!}$, then, the standardized form of permutation entropy $H_{PE}(m)$ can be defined as:

$$H_{PE} = H_{PE}(m) / \ln(m!) \quad (5)$$

where $0 \leq H_{PE} \leq 1$, H_{PE} reflects the randomness of the time series; the greater the randomness of the series, the greater the catastrophe probability; on the contrary, the less random the sequence, the less chance of mutation.

According to Bandt's recommendation, the range of embedded dimension m is 3–7. If m is too small, the reconstruction vector will contain less information, and the algorithm is meaningless. If m is too large, the reconstruction of phase space will homogenize the time series, the change in the sequence is smaller, so $m = 5$. λ plays a small role in the PE algorithm, so $\lambda = 1$.

MEEMD

The steps of the MEEMD algorithm based on PE are as follows:

1. Add white noise signals $C_i(t)$ and $-C_i(t)$ in the original signal $x(t)$, where the mean values are zero.

$$\begin{aligned} x_i^+(t) &= x(t) + b_i C_i(t) \\ x_i^-(t) &= x(t) - b_i C_i(t) \end{aligned} \quad (6)$$

where $C_i(t)$ is the added white noise signal, b_i is the amplitude of noise signal ($i = 1, 2, 3 \dots Ne$, Ne is the number of adding white noise).

2. Use EMD to decompose $x_i^+(t)$ and $x_i^-(t)$, and based on this, the first-order IMF components $\{I_{i1}^+(t)\}$ and $\{I_{i1}^-(t)\}$ series can be obtained.
3. Take the mean value of step (2):

$$I_1(t) = \frac{1}{2Ne} \sum_{i=1}^{Ne} [I_{i1}^+(t) + I_{i1}^-(t)] \quad (7)$$

4. Use the permutation entropy algorithm to detect whether $I_1(t)$ is an abnormal signal. If the permutation entropy value of the IMF component is greater than the threshold value θ_0 , then this component is the abnormal component; otherwise, the IMF component is a nearly stationary signal.
5. If $I_1(t)$ is an abnormal component, repeat steps (1)–(4), until the IMF $I_p(t)$ component is not an abnormal component.
6. Extract the first $n - 1$ decomposed components out of the original signal, that is:

$$r(t) = x(t) - \sum_{i=1}^{n-1} I_i(t) \quad (8)$$

EEMD

EEMD is an improved algorithm of EMD. Compared with EMD, EEMD adds Gaussian white noise to the signal and compensates for the loss of IMF components with its uniform distribution characteristics. Through the separation of the frequency scales, the occurrence of mode mixing can

be reduced. The biggest characteristic of EEMD is that it can extract the components and changing trends of signals in the high frequency and low frequency domains, so as to reduce the non-stationarity of series.

Based on the properties of the above EMD method, the procedure of EEMD can be shown as:

1. Add a white noise series to the original data.
2. Decompose the original data with white noise into IMF components.
3. Add different white noise series with equal root mean square every time; repeat steps (1) and (2) to get a group of different IMF components and residuals.
4. Take the corresponding IMFs' mean as the final IMF group.

ARIMA

Basic principles

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

In modeling ideas of ARIMA, the predicted 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 predict 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; and q is the order of the moving average component.

The ARIMA model is defined below as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 \xi_{t-1} - \theta_2 \xi_{t-2} - \dots - \theta_q \xi_{t-q} + \xi_t \quad (9)$$

where y is time series, $\phi_1, \phi_2 \dots \phi_p$ is autoregressive coefficient, $\theta_1, \theta_2 \dots \theta_q$ is moving average coefficient, ξ_t is error series, p is autoregressive order number ($p > 0$ and as an integer), and q is autoregressive order number ($q > 0$, and

as an integer). If the differential order is represented by d , the model can be written as ARIMA (p, d, q).

Modeling steps

The steps of ARIMA modeling are as follows:

- The stationarity of the series is identified according to the scatter diagram of the series and the autocorrelation and partial autocorrelation function (PACF) graph.
- If the series is not stable, use difference or moving average to smooth the data.
- The corresponding model was established according to the autocorrelation function (ACF) and PACF. If ACF censoring and PACF trailing, the MA (q) model was adopted. If ACF trailing and PACF censoring, the AR(p) model was adopted. If ACF trailing and PACF trailing, the ARIMA (p, d, q) model was adopted.

MEEMD-ARIMA model

The non-stationarity of series can be reduced by MEEMD decomposition of runoff; however, the pseudo-component is extremely unstable and cannot achieve a good prediction effect, so the pseudo-component is decomposed by EEMD. The decomposition of MEEMD and EEMD provides a stationarity premise for the prediction of the ARIMA model. The specific steps of the MEEMD-ARIMA model are as follows:

1. Use MEEMD to decompose annual runoff to obtain IMF components, trend, and pseudo-component of annual runoff.
2. Apply EEMD decomposition to the pseudo-component and obtain the IMF components and trend.
3. Verify whether the IMF components, trend, and the sub-component of the pseudo-component are stationary series. If stationary, $d = 0$; otherwise, d is determined by the difference order, then, p and q are determined by the ACF and PACF figures
4. Use the ARIMA model to fit and predict IMF components and trend.
5. Add up the prediction value of IMF components, trend, and pseudo-component, that is, the predicted value of the annual runoff.

For more information about the implementation steps for the MEEMD-ARIMA model, please see Figure 1.

CASE STUDY

The data source

The Yellow River originates in the Bayan Kera mountains in Qinghai province, China, passes through Lanzhou,

Zhengzhou, and Jinan, and flows into the Bohai Sea near Dongying city, Shandong province. Huayuankou hydrological station is located in the Lower Yellow River. It is 770 km away from the estuary and the catchment area covers 730,000 km². The annual runoff data covering 1960–2014 were provided by the hydrographic station. According to the runoff time series from 1960 to 2014, the maximum annual runoff is 5.60 billion m³ and occurred in 1961; the minimum annual runoff is 1.42 billion m³ and occurred in 1997; and the average annual runoff is 3.53 billion m³. The location of Huayuankou station is shown in Figure 2.

It can be seen from Figure 3 that the annual runoff at Huayuankou station shows an irregular fluctuation trend of rising and falling. The series shows large randomness, and the runoff value before 1990 is slightly larger than that after 1990. This trend of change can be attributed mainly to the following aspects.

1. The impact of human activities: With the development of industry, agriculture, and the social economy, the demand for water has increased, and which has caused the Lower Yellow River to enter a continuous period of dry water. After 1990, the demands on the Yellow River water reached more than 60% of the natural runoff and increased year by year (Peng 2011). Meanwhile, the water storage of Longyangxia reservoir also effectively reduced the runoff of the downstream main stations (Ding & Pan 2008).

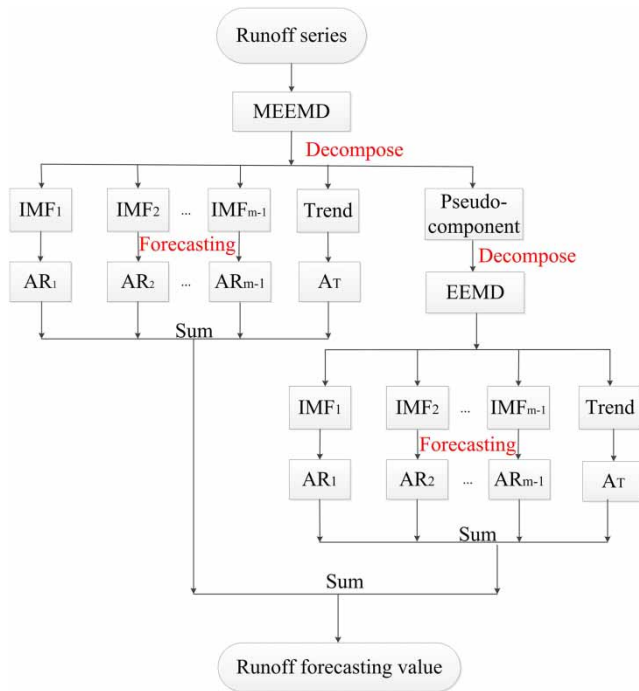


Figure 1 | Runoff prediction technical route.

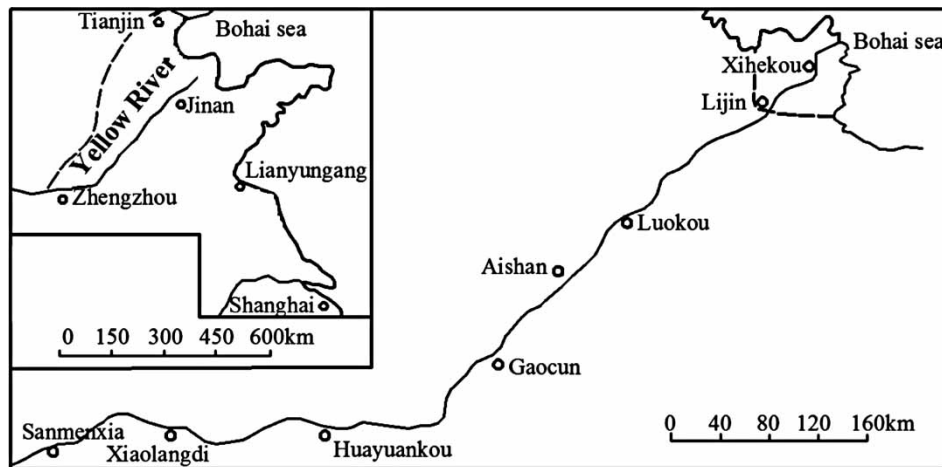


Figure 2 | Distribution map of the main hydrological stations in the Lower Yellow River.

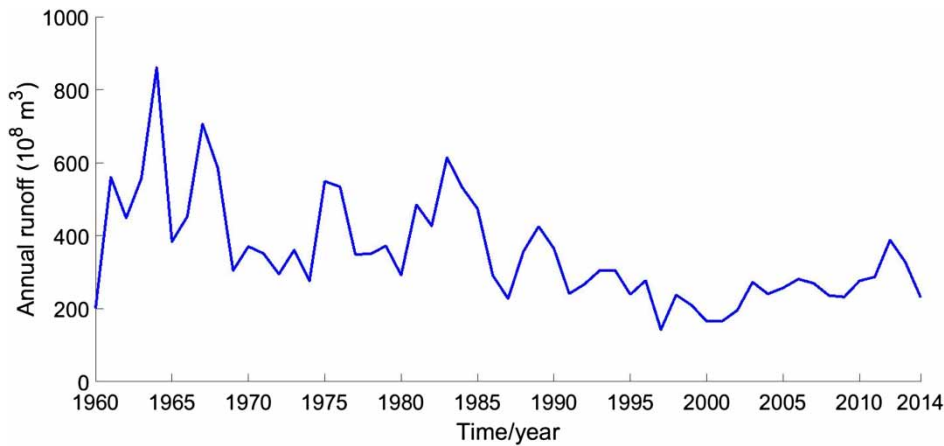


Figure 3 | Changes of annual runoff of Huayuankou from 1960 to 2014.

- Influence of precipitation: The precipitation shows a downward trend year by year, and the influence on the change of runoff is also declining year by year. Precipitation contributed 84% to the change of runoff in the Lower Yellow River from 1980 to 1992. It was 25% from 1993 to 2002, and only 10% after 2003 (Pan *et al.* 2017). The runoff will decrease without reducing water demand of the Yellow River.
- Influence of evapotranspiration: The temperature of the Lower Yellow River area increased 1.4 °C from 1960 to 2010. During this period, the evaporation capacity of the downstream was 40 mm due to the enhancement of temperature in the basin. As a result, the runoff generated by precipitation decreased and the runoff of the basin decreased.

In addition, there is the influence of water and sediment regulation of Xiaolangdi reservoir. After 2002, the runoff series showed a significant upward trend. It can be seen that the runoff evolution of Huayuankou station is greatly influenced by human activities, and the randomness of the series is relatively large, thus can be approximately regarded as a non-stationary series. Therefore, it is reasonable for this study to choose MEEMD and EEMD to decompose the runoff.

Decomposition of annual runoff

The runoff series are decomposed into sub-signals of different frequencies, which are IMF components, trend, and pseudo-component, respectively. The complex runoff

prediction is equal to the sum of the predicted values of different frequency subcomponents. By calculating the relative error of the sub-signal, the contribution rate of each sub-signal to the runoff series can be analyzed, and explain whether the relative error of a sub-signal influences the relative error of the runoff.

After repeated testing, when the noise logarithmic is 100, the noise amplitude is 0.2, the embedding dimension is 5, the maximum decomposition is 6, the PE threshold is 0.56, and the time delay is 1, MEEMD has the best decomposition effect on runoff, as shown in Figure 4.

As can be seen from Figure 4, the runoff series is decomposed into three IMF components, one trend, and one pseudo-component. Among them, the IMF₁ component has the lowest stationarity, which manifests as larger volatility, higher frequency, and shorter wavelength. The amplitude and frequency of the other IMF components gradually decreased and the wavelength gradually increased. In addition, the change amplitude of the IMF components is significantly lower than the original series. It can be seen that the volatility and tendency of the series are greatly reduced by MEEMD decomposition.

The pseudo-component shows great non-stationarity and randomness, and it is studied separately by using the EEMD method.

Pseudo-component decomposition

The MEEMD, CEEMD, and EEMD methods can be used to decompose pseudo-components. If MEEMD is used for quadratic decomposition, it will fall into the infinite loop

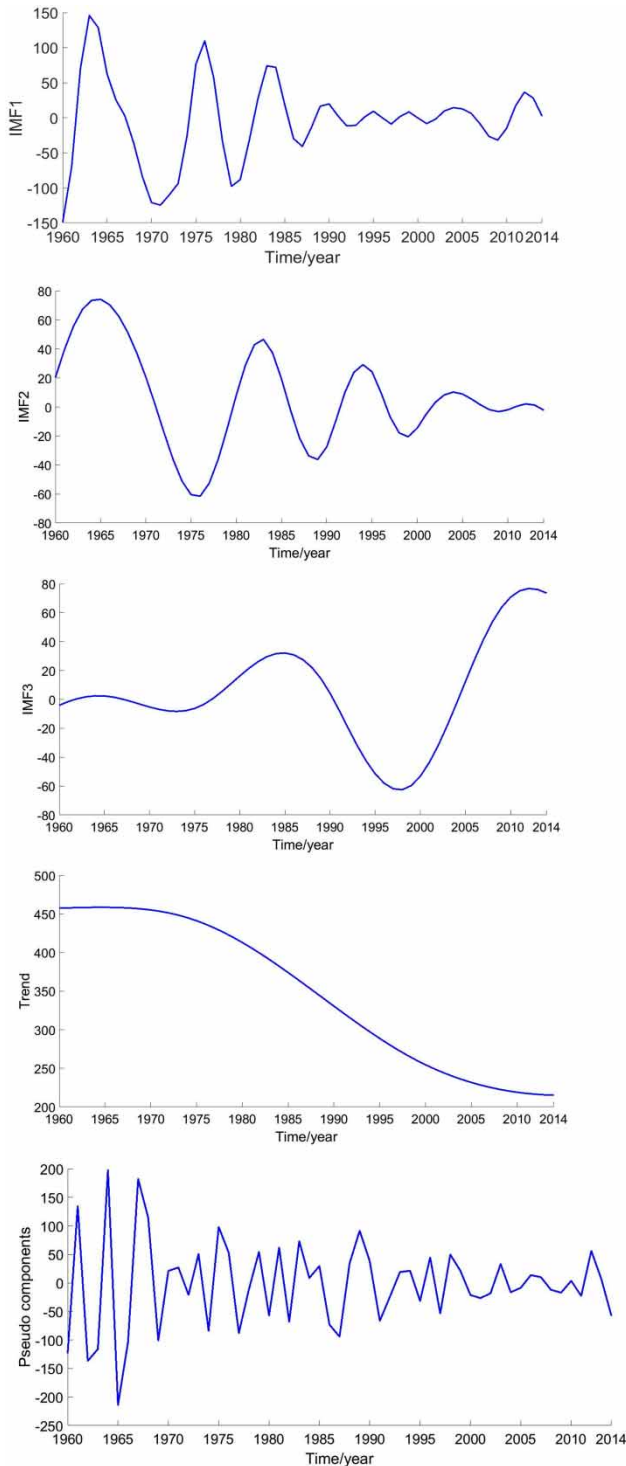


Figure 4 | MEEEMD decomposition of annual runoff at Huayankou station.

of pseudo-component extraction. If CEEMD is selected, it has four setting parameters, the determination of these parameters have no specific criteria and are difficult to

determine. Furthermore, the calculation is twice as much as EEMD (Zheng et al. 2013). Therefore, EEMD was determined to decompose the pseudo-component. The EEMD decomposition results are as shown in Figure 5.

As can be seen from Figure 5, the pseudo-component is decomposed into four IMF components and one trend. After the pseudo-component is decomposed again, the subcomponents of the pseudo-component become stable. Among them, the IMF₁ component has the lowest stationarity, which is manifested as larger volatility, higher frequency, and shorter wavelength. The amplitude and frequency of the other IMF components gradually decreased and the wavelength gradually increased. In addition, the variation amplitude of IMF components was significantly lower than the original series.

Runoff prediction

The research emphasis of this study is that the pseudo-component of IMF is extracted, decomposing the unstable pseudo-component again, and using ARIMA to predict all the obtained stationary sequences. The runoff predicted value of the MEEMD-ARIMA model is equal to the sum of each stationary component, that is IMF components, trend, and pseudo-component, respectively.

Prediction of IMF components and trend

Prediction values of IMF components and trend are shown in Table 1.

It can be seen from Table 1 that the prediction errors of IMF₁ and IMF₂ are relatively high; and the prediction error of IMF₃ and trend are relatively small. The reason can be explained by IMF₃ and trend are relatively stable, but the stability of IMF₁ and IMF₂ are relatively poor. Although the prediction error of IMF₁ and IMF₂ are relatively high, the IMF₁ and IMF₂ from 2010 to 2014 account for a small proportion in the runoff sequence, so they would not influence the runoff prediction error.

Prediction of pseudo-component

The pseudo-component was decomposed into four IMF components and one trend, and the overall prediction error of the pseudo-component is shown in Table 2.

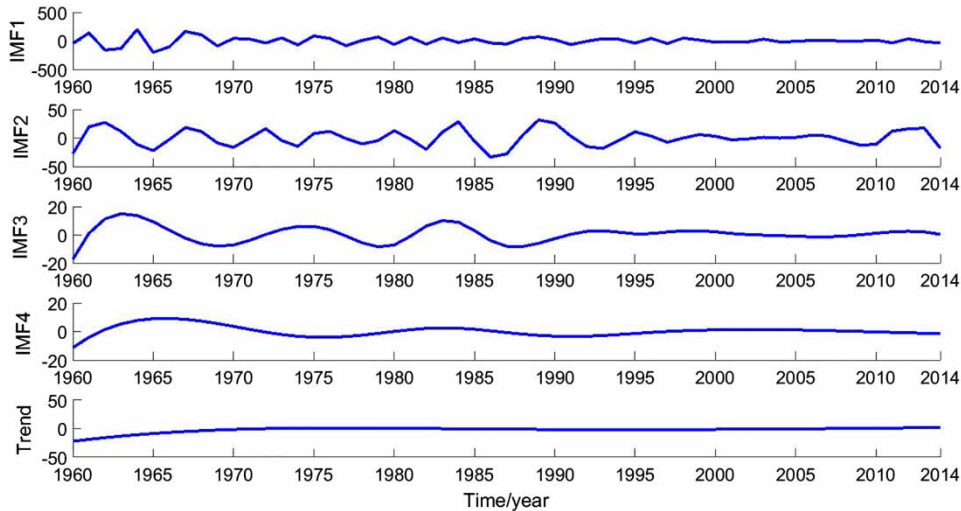


Figure 5 | Pseudo-component decomposition.

Table 1 | Prediction values of IMF components and trend

IMF component	Time (year)	True value	Prediction value	Relative error (%)	ARIMA (p, d, q)	R ²
IMF ₁	2010	-15.10	-22.65	50.04	(6,0,2)	0.84
	2011	16.67	12.63	-24.23	(6,0,2)	0.85
	2012	36.49	33.03	-9.47	(6,0,2)	0.81
	2013	28.00	29.35	4.84	(6,0,2)	0.85
	2014	2.16	-0.45	-120.81	(6,0,2)	0.87
IMF ₂	2010	-1.99	-2.55	28.26	(10,0,3)	0.96
	2011	0.48	1.96	309.81	(10,0,3)	0.96
	2012	2.13	2.80	31.66	(10,0,3)	0.95
	2013	1.24	2.07	66.46	(10,0,3)	0.92
	2014	-2.28	-2.49	9.00	(10,0,3)	0.90
IMF ₃	2010	70.79	70.58	-0.30	(16,0,2)	1.00
	2011	75.12	74.90	-0.30	(16,0,2)	1.00
	2012	76.79	76.60	-0.25	(16,0,2)	1.00
	2013	76.08	75.59	-0.64	(16,0,2)	1.00
	2014	73.61	73.03	-0.79	(16,0,2)	1.00
Trend	2010	218.94	219.14	0.09	(15,0,1)	0.97
	2011	217.50	217.83	0.15	(15,0,1)	0.98
	2012	216.40	216.62	0.10	(15,0,1)	0.97
	2013	215.64	215.66	0.01	(15,0,1)	0.97
	2014	215.20	215.21	0.01	(13,0,1)	0.97

MEEMD-ARIMA prediction

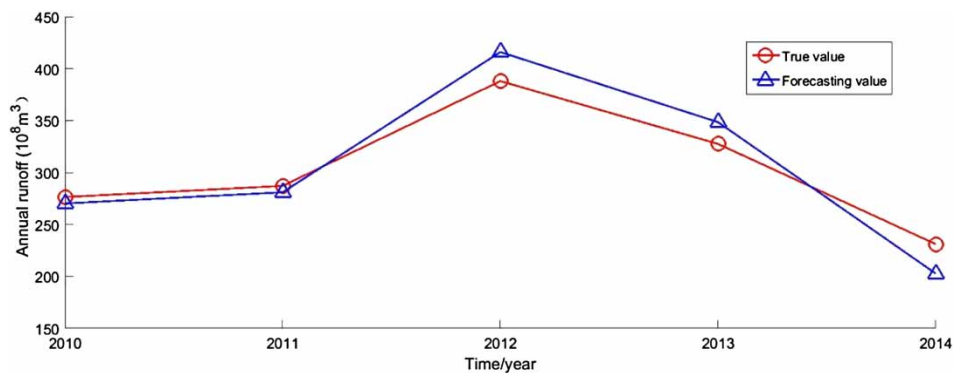
The prediction value of runoff is equal to the prediction value of three IMF components, trend, and pseudo-component. The prediction effect of runoff at Huayuankou station is shown in Figure 6.

The runoff prediction value and error of MEEMD-ARIMA model at Huayuankou station are shown in Table 3.

It can be seen from Figure 6 and Table 3, the prediction effect of MEEMD-ARIMA model was feasible, and the relative error $\leq \pm 13\%$. It can be seen that even if the Huayuankou station was influenced by human activities

Table 2 | Parameters and predicted values of pseudo-component

IMF component	Time (year)	True value	Prediction value	Relative error (%)	ARIMA (p, d, q)	R ²
IMF ₁	2010	12.83	5.02	-60.86	(3,0,2)	0.39
	2011	-13.99	-3.37	-75.91	(3,0,2)	0.37
	2012	27.53	20.61	-25.14	(3,0,2)	0.33
	2013	-11.40	-5.12	-55.09	(3,0,0)	0.38
	2014	-25.31	-6.73	-73.41	(3,0,0)	0.29
IMF ₂	2010	-6.08	0.28	-104.61	(3,0,2)	0.71
	2011	-2.88	2.06	-171.53	(2,0,2)	0.56
	2012	31.44	12.33	-60.78	(2,0,2)	0.76
	2013	20.20	26.48	31.12	(1,0,2)	0.60
	2014	-29.40	-15.75	-46.43	(3,0,4)	0.74
IMF ₃	2010	-5.57	-5.68	2.00	(3,0,2)	0.70
	2011	-5.17	-4.90	-5.19	(3,0,2)	0.92
	2012	-4.95	-4.88	-1.45	(3,0,2)	0.91
	2013	-5.08	-4.90	-3.59	(3,0,2)	0.86
	2014	-5.72	-5.39	-5.82	(3,0,2)	0.84
IMF ₄	2010	-11.16	-11.16	0.01	(3,0,2)	0.31
	2011	-11.36	-11.36	0.00	(3,0,2)	0.70
	2012	-11.55	-11.55	-0.02	(3,0,2)	0.78
	2013	-11.73	-11.73	0.03	(3,0,2)	0.90
	2014	-11.87	-11.87	-0.03	(3,0,2)	0.88
Trend	2010	13.54	13.53	-0.06	(3,0,2)	-1.02
	2011	13.86	13.85	-0.07	(3,0,2)	-0.99
	2012	14.19	14.18	-0.07	(3,0,2)	0.38
	2013	14.53	14.52	-0.06	(3,0,2)	-0.41
	2014	14.88	14.87	-0.04	(3,0,2)	-1.41

**Figure 6** | The prediction effect of annual runoff at Huayuankou station.

and climate condition, the MEEMD-ARIMA model still had good effect on short-term runoff prediction.

DISCUSSION

To the best of our knowledge, MEEMD, CEEMD, and EEMD can be used for signal decomposition. The

decomposition effect of the three methods are different under the same parameter setting. In this paper, only IMF₁-IMF₃ components with a little low stationarity and trend renderings with a large proportion are achieved, as shown in Figure 7.

MEEMD-ARIMA, CEEMD-ARIMA, and EEMA-ARIMA can be used to predict runoff. Comparison of the prediction results of the three models are shown in Table 4.

Table 3 | Prediction error of annual runoff of Huayuankou station

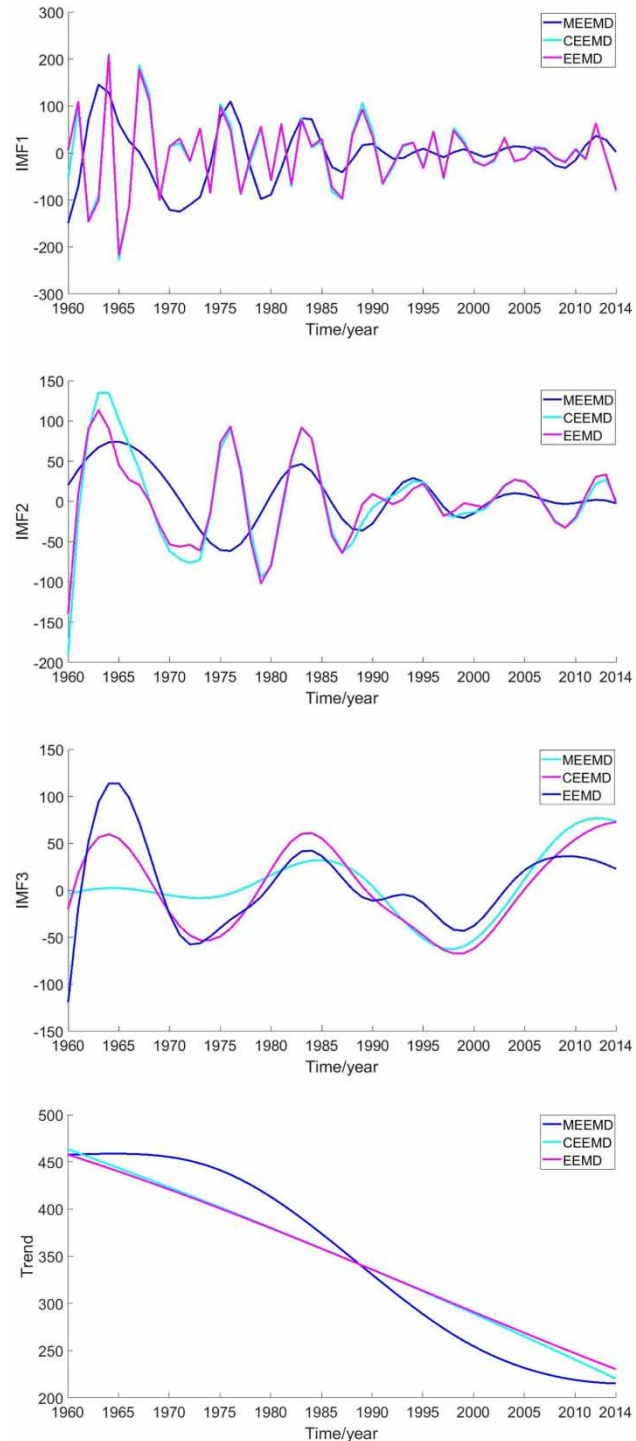
Time (year)	True value (10^9 m^3)	Prediction value (10^9 m^3)	Absolute error (10^9 m^3)	Relative error of prediction (%)
2010	276.30	270.16	6.14	-2.22
2011	287.10	280.93	6.17	-2.15
2012	388.00	415.94	-27.94	7.20
2013	327.50	348.47	-20.97	6.40
2014	231.00	202.74	28.26	-12.23

It can be seen from Figure 7, in terms of the volatility and magnitude of the IMF components, that the decomposition effect of MEEMD is slightly better than CEEMD and EEMD. In the decomposition process, the PE is used to extract the pseudo-component of IMF, which is also the reason why the decomposition effect of MEEMD is better than CEEMD and EEMD. In addition, it can be seen from Table 4 that the prediction error of the MEEMD-ARIMA model is smaller and superior to the other two models.

From the perspective of MEEMD decomposition, the stability of IMF components and trend are different, and the contribution rate of runoff is also different. The trend accounts for a large proportion in the runoff series; once the trend prediction effect is poor, the prediction effect of the runoff series is definitely poor. Therefore, the trend contributes more to runoff series. However, the IMF_1 accounts for a small proportion in the runoff series, and even if the IMF_1 component has a slightly higher prediction error, the impact on the runoff prediction accuracy is also small. Therefore, the IMF_1 contributes less to runoff series. This is also the reason why the IMF_1 prediction error does not influence the overall prediction error of the runoff.

Due to the different evolution factors of runoff in different regions, some regions are strongly influenced by human activities, such as the construction of reservoirs, and water and soil conservation. Some areas are greatly affected by natural factors, such as regional precipitation, and underlying surface conditions. Due to the lack of sufficient data in the study to verify feasibility in different regions, the feasibility of applying this model to other regions cannot be determined, but it is feasible in theory.

The factors that influence the results essentially depend on the IMF's weight itself. Generally speaking, human activities and natural factors can influence the non-stationarity of

**Figure 7** | Decomposition effect of IMF_1 - IMF_3 and trend in three models.

runoff series. Furthermore, it influences the non-stationarity of IMF components after runoff decomposition. Under the interference of different natural factors and human activities,

Table 4 | The prediction error of the three models

Time (year)	True value (10^8 m^3)	Prediction value (10^8 m^3)			Relative error (%)		
		MEEMD-ARIMA	CEEMD-ARIMA	EEMD-ARIMA	MEEMD-ARIMA	CEEMD-ARIMA	EEMD-ARIMA
2010	276.3	268.17	250.17	231.28	-2.22	-9.46	-16.29
2011	287.1	284.65	300.65	316.36	-2.15	4.72	10.19
2012	388	385.25	364.28	351.30	7.2	-6.11	-9.46
2013	327.5	329.22	314.30	315.26	6.4	-4.03	-3.74
2014	231	227.61	200.15	200.23	-12.23	-13.35	-13.32

the value of IMF components after runoff decomposition is different, which may lead to some data catastrophe points of IMF components. However, this model has a slightly lower prediction accuracy for the IMF₁ component, which has a certain impact on the model accuracy.

The Yellow River is one of the most complex rivers in the world. It is greatly influenced by human activities and natural factors, which make it very difficult to establish a model. This multiple decomposition–reconstruction model makes the application of the MEEMD-ARIMA model in the Yellow River basin feasible. However, when determining the ARIMA model for IMF components, determination one by one is necessary.

CONCLUSIONS

Annual runoff time series of Huayuankou station are characterized by high randomness and uncertainty. The short-term prediction can be accomplished by using MEEMD to reduce the non-stationarity of the series.

The fitting and prediction results for annual runoff of Huayuankou station show that the decomposition of runoff series by MEEMD and EEMD not only solves the problem that the ARIMA model requires the series to be stable, but also reflects the variation characteristics of annual runoff series in various frequency domains. The established MEEMD-ARIMA coupling model was applied to the prediction of annual runoff series of Huayuankou station, and its prediction error was no more than $\leq 13\%$, which was better than CEEMD-ARIMA and EEMD-ARIMA.

This has been a new attempt to apply the MEEMD-ARIMA coupling model to prediction of the Lower Yellow River runoff. Its extension and improvement in model accuracy still need further study. Furthermore, the MEEMD-

ARIMA model can be used to predict rainfall, sediment transport, and meteorological factors, which demonstrates the prospect for broad application.

The model does not consider the physical mechanism of runoff evolution and the long-term prediction. Furthermore, how to deal with the pseudo-component properly and how to improve the prediction 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

- Alizadeh, M. J., Kavianpour, M. R., Kisi, O. & Nourani, V. 2017 [A new approach for simulating and forecasting the rainfall-runoff process within the next two months](#). *Journal of Hydrology* **548**, 588–597.
- Bandt, C. & Pompe, B. 2002 [Permutation entropy: a natural complexity measure for time series](#). *Physical Review Letters* **88** (17), 174102.
- Borgnat, P., Flandrin, P., Honeine, P., Richard, C. & Xiao, J. 2010 [Testing stationarity with surrogates: a time-frequency](#)

- approach. *IEEE Transactions on Signal Processing* **58** (7), 3459–3470.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. 1997 *Time Series Analysis – Prediction and Control*. China Statistical Press, Beijing, China, pp. 64–65.
- Cao, Y., Tung, W. & Gao, J. B. 2004 Detecting dynamical changes in time series using the permutation entropy. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* **70** (4 Pt 2), 046217.
- Coulibaly, P., Anctil, F., Rasmussen, P. & Bernard, B. 2015 A recurrent neural networks approach using indices of low-frequency climatic variability to predict regional annual runoff. *CGHU Special Issue of Hydrological Processes* **14** (15), 2755–2777.
- Ding, Y. F. & Pan, S. M. 2008 Analysis on the evolutionary characteristics of annual runoff and sediment flux distribution of the Yellow River. *Marine Science Bulletin* **3**, 49–56.
- Li, C. H., Zheng, X. K. & Yang, Z. F. 2009 Trends of annual runoff in the Yellow River Basin. *Journal of Beijing Normal University (Natural Science Edition)* **45** (1), 80–85.
- Mahabir, C., Hicks, F. E. & Fayek, A. R. 2010 Application of fuzzy logic to predict seasonal runoff. *Hydrological Processes* **17** (18), 3749–3762.
- Pan, B., Han, M. & Ni, J. 2017 Characteristics and influencing factors in the lower reaches of the Yellow River in the last 50 years. *Journal of Soil and Water Conservation Research* **24** (1), 122–127.
- Peng, J. 2011 *Variation Process of Water and Sediment in the Yellow River and Evolution of Sedimentary Environment in the Yellow River Delta*. East China Normal University, Shanghai.
- Sedki, A., Ouazar, D. & El Mazoudi, E. 2009 Evolving neural network using real coded genetic algorithm for daily rainfall-runoff prediction. *Expert Systems with Applications* **36** (3), 4523–4527.
- Terrien, J., Marque, C. & Karlsson, B. 2011 Automatic detection of mode mixing in empirical mode decomposition using non-stationarity detection: application to selecting IMFs of interest and denoising. *EURASIP Journal on Advances in Signal Processing* **2011** (1). <https://doi.org/10.1186/1687-6180-2011-37>
- Tu, Y., Wang, J. N., Zhu, Q. P., An, X. W., Mei, Y. & Chen, D. Z. 2018 Research on the prediction of hydrological time series based on PSO-KELM model. *China Rural Water and Hydropower* **429** (07), 21–24.
- Wang, W. C., Xu, D. M., Chau, K. W. & Chen, S. Y. 2013 Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *Journal of Hydroinformatics* **15**, 1377–1390.
- Wu, Z. 2009 Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis* **1**, 1–41.
- Yan, R. & Gao, R. X. 2007 Approximate entropy as a diagnostic tool for machine health monitoring. *Mechanical Systems and Signal Processing* **21**, 824–839.
- Yeh, J. R. & Shieh, J. S. 2010 Complementary ensemble empirical mode decomposition: a noise enhanced data analysis method. *Advances in Adaptive Data Analysis* **2** (2), 135–156.
- Zhang, J. X., Ma, X. Y., Zhao, W. H., Hao, J. J. & Qu, J. N. 2008 Application of the life cycle-Markov combination model for river annual runoff prediction. *Journal of Hydroelectric Engineering* **27** (6), 32–36.
- Zhao, X. H. & Chen, X. 2015 Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. *Water Resource Management* **29**, 2913–2926.
- Zhao, Q. S., Yang, T. X., Zou, J. F. & Wu, G. H. 2001 Application of improved BP algorithm to dry season runoff prediction in the lower Yellow River. *Journal of Safety and Environment* **1** (3), 30–35.
- Zhao, C. P., Xu, X. X., Chen, T. L. & Zhang, S. N. 2015 Impact of land use change on the runoff in Guanchuanhe Watershed. *Soil and Water Conservation Research* **22** (3), 83–87.
- Zheng, X., Hao, Z. Y., Lu, Z. G. & Yang, J. 2012 Separation of piston-slap and combustion shock excitations via MEEMD method. *Journal of Vibration and Shock* **31** (6), 109–113.
- Zheng, J. D., Cheng, J. S. & Yang, Y. 2013 Modified EEMD algorithm and its applications. *Journal of Vibration and Shock* **32** (21), 21–26.

First received 4 November 2018; accepted in revised form 15 March 2019. Available online 19 April 2019