Multi-time scale co-integration forecast of annual runoff in the source area of the Yellow River

Jinping Zhang, Hongbin Li, Bin Sun and Hongyuan Fang

ABSTRACT

In order to reveal the multi-time scale of rainfall, runoff and sediment in the source area of the Yellow River and improve the accuracy of annual runoff forecast, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method is introduced to decompose the measured rainfall, runoff and sediment data series of the Tangnahai hydrological station in the source area of the Yellow River of China. With the co-integration theory, two new error correction models (ECMs) for the forecast of annual runoff in the source area of the Yellow River are constructed. The application of these two methods solves the problem of pseudo-regression caused by nonlinearity and non-stationary of hydrological time series. The results show that rainfall, runoff and sediment in the source area of the Yellow River have multi-time scales and the component sequences have co-integration relationships. For two new ECMs, the CEEMDAN component ECM has better forecast accuracy than the original sequence one. The relative error of all forecasted values is less than 15% except 2009, and the accuracy has reached level A.

Key words | CEEMDAN, co-integration, multi-time scale, runoff forecast, source area of the Yellow River

Jinping Zhang
Hongbin Li
Bin Sun (corresponding author)
Hongyuan Fang
School of Water Conservancy Engineering,
Zhengzhou University,
Zhengzhou 450001,
China
E-mail: *bin_zzu@163.com*

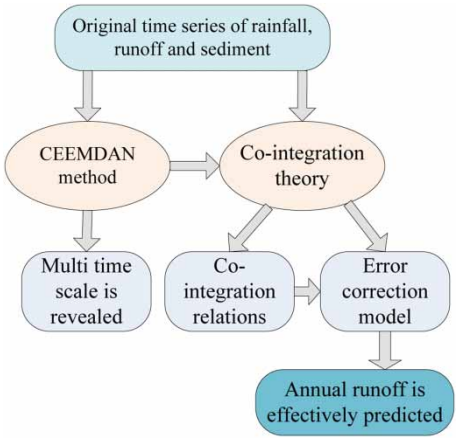
Jinping Zhang
Yellow River Institute for Ecological Protection & Regional Coordinated Development,
Zhengzhou University,
Zhengzhou 450001,
China
and
Zhengzhou Key Laboratory of Water Resource and Environment,
Zhengzhou 450001,
China

HIGHLIGHTS

- The research on the multi-time scale change law of hydrological variables reveals the multi periodic change law of hydrological variables and provides a scientific basis for the rational development of water resources.
- The non-stationary and nonlinear processing of hydrological variables can avoid spurious regression and make the result more accurate.
- Study on the co-integration relationship of rainfall, runoff and sediment.
- Study on multi-time scale dynamic relationship among rainfall, runoff and sediment.
- Multi-time scale prediction of river runoff provides a technical reference for the effective protection and scientific operation of water resources.

doi: 10.2166/wcc.2020.137

GRAPHICAL ABSTRACT



INTRODUCTION

Rainfall, runoff and sediment are important hydrological variables with complex relationships in the source area of the Yellow River. Accurately grasping the changes in these hydrological variables plays a vital role in the change of water resources throughout the Yellow River Basin (Chen & Guo 2016; Wang *et al.* 2017, 2018a). Besides, rainfall, runoff and sediment analysis are hot issues for scholars at home and abroad (Ramana *et al.* 2013; Ling *et al.* 2017; Li & Liu 2018; Wang *et al.* 2018b). At present, many scholars have conducted a lot of research on the relationship between rainfall and runoff (Nastiti *et al.* 2018; Tarasova *et al.* 2018; Chu *et al.* 2019), and many others have also conducted research on the relationship between runoff and sediment (Hou *et al.* 2013; Zhang *et al.* 2014; Wang *et al.* 2015a, 2015b) and achieved great results. However, the results of combining the three together for research are relatively few.

Runoff forecasting has always been a hot issue in the field of hydrology (Zhang *et al.* 2017a, 2017b; Zhao *et al.* 2017). The runoff forecasting with the historical data can not only realize the rational development and utilization of runoff resources but also has significance for the planning, construction and scheduling of water conservancy projects (Xie *et al.* 2019). At present, the forecast of river runoff often assumes that the time series is stationary. However, because of the influence of climate change, underlying surface and human activities, the statistical characteristics of hydrological time series always change with time. Therefore, most hydrological time series are nonlinear and non-stationary (Zhang *et al.* 2013a, 2013b). With the development of computer technology, many researchers use soft computing techniques to study hydrological variables with highly accuracy (Rezaie-Balf *et al.* 2017; Mosavi *et al.* 2018; Guru & Jha 2019). The common runoff forecast models include artificial neural network (ANN) model (Meng *et al.* 2015; Sezen & Partal 2019), support vector regression (SVR) model (Yaseen *et al.* 2018; Wu *et al.* 2019) and autoregressive moving average (ARMA) model (Wang *et al.* 2015a, 2015b, 2019). These models with non-stationary time series data will lead to pseudo-regression (Lee & Yu 2009; Jin *et al.* 2017), so their hydrological element simulation and forecasting are unbelievable.

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is an effective method for dealing with nonlinear and non-stationary time series (Torres *et al.* 2011), which is an improvement on the empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD) (Huang *et al.* 1998; Wu & Huang 2009). The EMD and EEMD methods are widely used in the fields of hydrology and water resources (Zhang *et al.* 2013a, 2013b, 2019a, 2019b; Ouyang *et al.* 2016; Adarsh & Reddy 2018). However, the EMD method presents the

mode confusion, and the EEMD method remains a noise residual problem. The CEEMDAN method solves both of these problems (Colominas *et al.* 2014) and is applied in many fields (Antico *et al.* 2016; El Bouny *et al.* 2019).

The co-integration theory was proposed by Engle & Granger (1987), which can deal with the non-stationary problem of time series and reveal the long-term equilibrium and short-term fluctuation between variables. Now, it is widely used in econometrics and the field of hydrology (Yoo 2007; Zhang *et al.* 2015). Meanwhile, this theory can also be combined with other data analysis methods so as to improve the accuracy of calculation (Zhang *et al.* 2017a, 2017b, 2019a, 2019b).

The innovation of this paper is to combine the CEEMDAN method with the co-integration theory to construct the three-variable CEEMDAN co-integration error correction model (ECM) for rainfall, runoff and sediment in the source area of the Yellow River of China to forecast the river runoff. The first is to use CEEMDAN to decompose rainfall, runoff and sediment in multi-time scales, and know the changing laws and poly-cycle and obtain the corresponding stationary time series. The second is to reveal the long-term equilibrium and short-term fluctuation relationship of the original and component time series of rainfall, runoff and sediment in the source area of the Yellow River according to the co-integration theory, and to clarify their influencing relationships. The last is to construct two new ECM models of rainfall, runoff and sediment including the CEEMDAN component ECM model and the original sequence ECM model to forecast the river runoff.

STUDY METHODS AND STEPS

CEEMDAN method

The CEEMDAN method is a time-frequency domain analysis method. It can further eliminate the mode effect by adding adaptive noise and has the strong adaptability and better convergence. Usually, the CEEMDAN method is used to deal with nonlinear and non-stationary time series.

The CEEMDAN algorithm steps are as follows:

Step 1: Add Gaussian white noise $s(t)$ to the original signal $x(t)$ and perform I test. The signal of the ith test can be expressed as:

$$x_i(t) = x(t) + s_i(t) \tag{1}$$

Step 2: Perform EMD decomposition on the signal $x_i(t)$ of the ith white noise addition, and perform an average of I test, and obtain the first intrinsic model function (IMF) component I_{IMF_1} as:

$$I_{IMF_1} = \frac{1}{I}\sum_{i=1}^{I} I_{IMF_{i1}} \tag{2}$$

In the formula, $I_{IMF_{i1}}$ is the first IMF component after EMD decomposition of signal $x_i(t)$ with white noise for the ith time.

Step 3: After decomposing to obtain the first IMF component, calculate the difference $r_1(t)$ between the original signal and I_{IMF_1} component:

$$r_1(t) = x(t) - I_{IMF_1} \tag{3}$$

Step 4: Add white noise again to the difference signal $r_1(t)$ and perform I test. Then, the difference signal $r_{i1}(t)$ of the white noise added to the ith time can be expressed as:

$$r_{i1}(t) = r_1(t) + s_i(t) \tag{4}$$

Step 5: Perform EMD decomposition on the ith white noise signal $r_{i1}(t)$, the ith is $I_{IMF_{i2}}$ component, and the second-order IMF component I_{IMF_2} obtained by the I test is:

$$I_{IMF_2} = \frac{1}{I}\sum_{i=1}^{I} I_{IMF_{i2}} \tag{5}$$

Step 6: At this time, the difference obtained by the decomposition is $r_2(t) = r_1(t) - I_{IMF_2}$, and steps 4 and 5 are repeated until $r_n(t)$ satisfies one of the following conditions: (1) cannot be further decomposed by EMD; (2) meet IMF conditions; (3) the number of local extreme points is less than 3. Finally, the original signal $x(t)$ can be decomposed

into n IMF components and a trend term $r_n(t)$:

$$x(t) = \sum_{i=1}^{n} I_{IMF_i} + r_n(t) \tag{6}$$

Co-integration theory

Co-integration concept

Co-integration describes the long-term equilibrium relationship between time series. If a time series is non-stationary but becomes stationary after d-difference, it is called d-order simple integer, which is recorded as $I(d)$. If the time series itself is stationary, it is recorded as $I(0)$. The two time series are defined as $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})^T$ and $Y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})^T$. If meeting the following conditions:

(1) $X_{it} \sim I(d)$ and $Y_{it} \sim I(d)$, $(i = 1,2,\ldots, n)$, d is an integer;
(2) there is a constant β, which makes $Y_t - \beta X_t \sim I(0)$;

Then, X_t and Y_t are co-integrated, and β is called the co-integration vector.

Stationary test

Before the co-integration test, it is necessary to conduct the stationary test of time series, and the commonly used method is the Augmented Dickey–Fuller (ADF) unit root test (Dickey & Fuller 1979). The formula is as follows:

$$\Delta y_t = \alpha + \beta_t + \delta y_{t-1} + \sum_{i=1}^{p} \xi_i \Delta y_{t-i} + \varepsilon_t \tag{7}$$

In the formula, $\Delta y_t = y_t - y_{t-1}$ is the first-order difference of variable y_t; α, β, δ, ζ_i are all parameters; t is the time; p is the lag order; ϵ_t is the white noise process.

Co-integration test

The E.G. two-step method is a common method for testing the co-integration relationship between time series, which was proposed by Engle & Granger (1987).

The first step of this method is to use the ordinary least square method (OLS) to regress multiple variables and get a residual sequence;

The second step is to test the stationarity of time series with the ADF unit root test on the residual sequence obtained in the first step. If the residual sequence is stationary, it is proved that the variables are co-integrated.

Error correction model

If the time series is co-integrated, an ECM can be constructed. This model describes the long-term equilibrium and short-term fluctuations between variables, and the modeling steps are as follows:

The first step is to perform a co-integration regression to the variables:

$$z_t = k_0 + k_1 x_t + k_2 y_t + u_t, \quad t = 1, 2, \ldots, T \tag{8}$$

to obtain k_0, k_1, k_2 and the residual sequence u_t;

$$u_t = z_t - k_0 + k_1 x_t + k_2 y_t \tag{9}$$

The second step is to make $ecm(-1) = u_{t-1}$ as an error correction term and substitute the error correction model:

$$\Delta z_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 \Delta y_t + \varphi ecm(-1) + \varepsilon_t \tag{10}$$

In the formula, β_0 is a constant term, β_1 and β_2 are the coefficients of the difference terms of each variable, which reflects the short-term dynamic changes of the model; $ecm(-1)$ is an error correction term, which reflects the degree to which the former term deviates from the long-term equilibrium in short-term fluctuations; φ is the correction coefficient, also called the adjustment speed, usually a negative value; ϵ_t is a white noise sequence.

Study steps

By using the CEEMDAN method, the time series of rainfall, runoff and sediment in the source area of the Yellow River are decomposed to obtain IMF component sequences at different time scales. Furthermore, the co-integration theory is used to construct the ECM for the original time series (ECM-OTS) and the CEEMDAN component

sequences (ECM-CEEMDAN), and then, the runoff is forecasted by ECM-OTS and ECM-CEEMDAN, respectively. Finally, the runoff forecasted value of each IMF component is reconstructed to get the runoff forecasted value of ECM-CEEMDAN, and the fitting value and forecast accuracy of these two ECM models are compared to draw a conclusion. The flow chart of study steps is shown in Figure 1.

RESULTS AND CONCLUSION

Data source

The source area of the Yellow River refers to the area above the Tangnaihai hydrological station, which is located in the northeast of the Qinghai Tibet Plateau of China. The geographic coordinates are between 95°50′–103°30′ E and 32°10′–36°05′ N (as shown in Figure 2), the basin area is 122,000 km^2, and the average annual runoff is 20.37 billion m^3. The water source is mainly supplied by rainfall, followed by glacial snow melting water and groundwater. The change of runoff in the source area of the Yellow River has a vital influence on the change of water resources in the whole Yellow River Basin.

The measured rainfall, runoff and sediment time series from 1966 to 2013 at Tangnaihai hydrological station are obtained by the Bureau of Meteorology and the Bureau of hydrology and water resources are shown as in Figure 3.

Table 1 shows the statistical characteristics of rainfall and runoff time series. The mean value is 556.516 mm for rainfall time series, and 203.885 billion m^3 for runoff time series, and $1{,}277.245 \times 10^4$ t for sediment time series. The

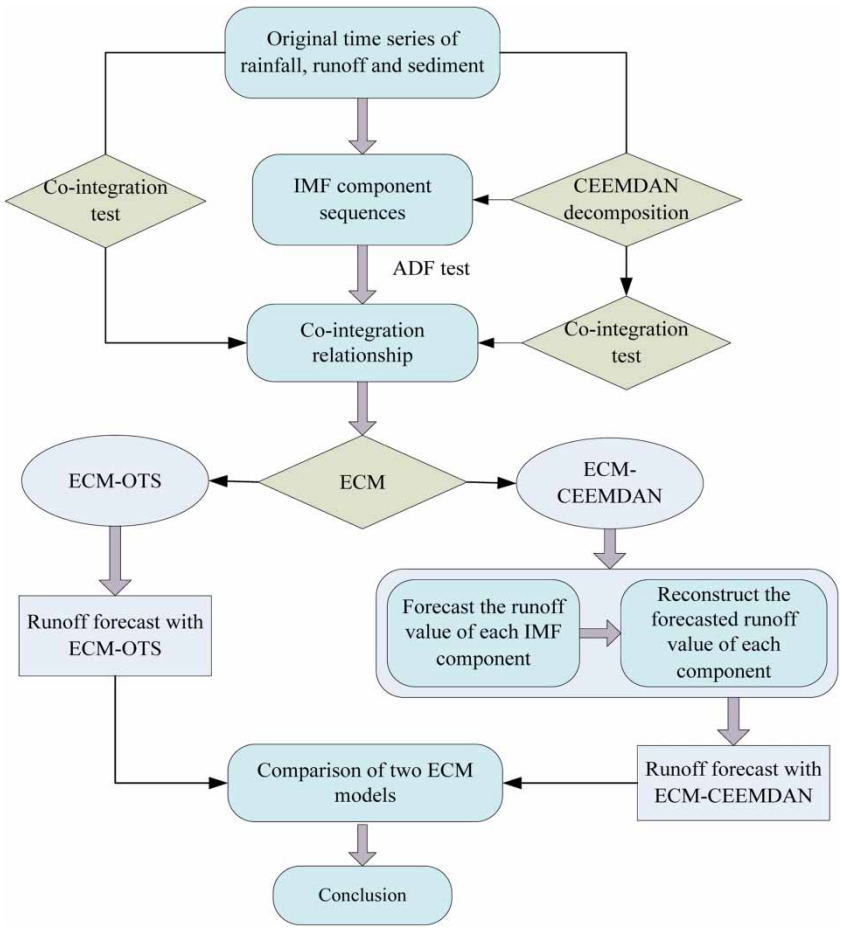


Figure 1 | Flow chart of study steps.

Figure 2 | The location map of the source area of the Yellow River.

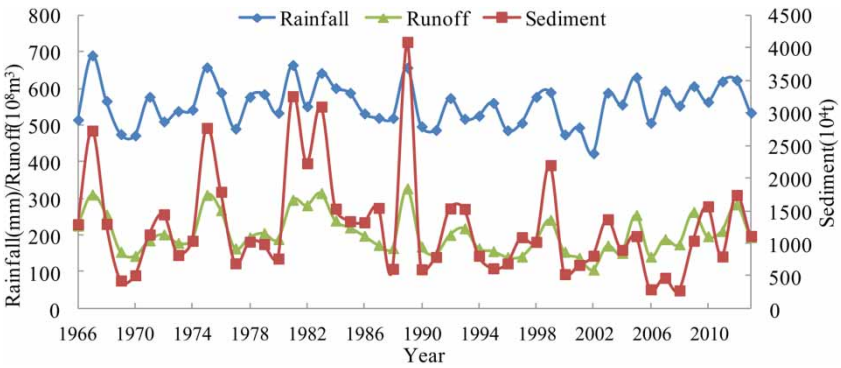


Figure 3 | Time series of rainfall, runoff and sediment at the Tangnaihai station in the source area of the Yellow River.

standard deviation of runoff is less than rainfall and sediment. For the coefficient of variation and skewness coefficient, the calculated value of sediment is larger than that of rainfall and runoff.

CEEMDAN decomposition. The CEEMDAN method is used to decompose the time series of rainfall, runoff and sediment in the source area of the Yellow River for multi-time scales. The decomposition results are shown in Figures 4–6.

With the CEEMDAN method, the annual runoff, rainfall and sediment data series at Tangnaihai hydrological station from 1966 to 2013 are decomposed into a fifth-order mode, including four IMF components and one

Table 1 | The statistical parameters of rainfall and runoff time series

Time series	Mean	Standard deviation	Coefficient of variation	Skewness coefficient
Rainfall	556.516	58.464	0.105	0.206
Runoff	203.885	55.006	0.270	0.659
Sediment	1277.245	812.960	0.636	1.534

residual. It reflects the multi-time scale evolution characteristics of rainfall, runoff and sediment in the source area of the Yellow River. The IMF1 component of each variable has the shortest period and the highest frequency, and the period of other components gradually gets longer and their frequency gradually decreases. The periodic changes of the component time series are shown in Table 2.

It can be seen from Table 2 that rainfall, runoff and sediment have four periodic changes. Specifically, rainfall, runoff and sediment all have the same short-period change, and the periodic year is 2–5 years; in the medium period, although the changing periodic years of the three are different, there is little difference, among which the rainfall is 5–8 years, runoff is 6–9 years, and sediment is 5–10 years. There are great differences between rainfall, runoff and sediment in the medium-long period, among which the span of sediment change is large with 11–30 years, 29–30 years for runoff and 9–11 years for rainfall. In terms of the long-period scale, rainfall is 28 years, runoff is 32 years, and sediment is 41 years. The residual component shows the overall nonlinear trend of rainfall, runoff and sediment. The residual component of rainfall showed a decreasing trend from 1966 to 1981, and an increasing trend from 1982 to 2013, but both residual components of runoff and sediment showed a decreasing trend. It can be seen that rainfall, runoff and sediment all have complex multi-time scale periodic change laws, but they have a good correlation in the short and the medium periods. Moreover, runoff and sediment present different periodic changes in the medium-long and long periods, while for their residual components, they show the better synchronization.

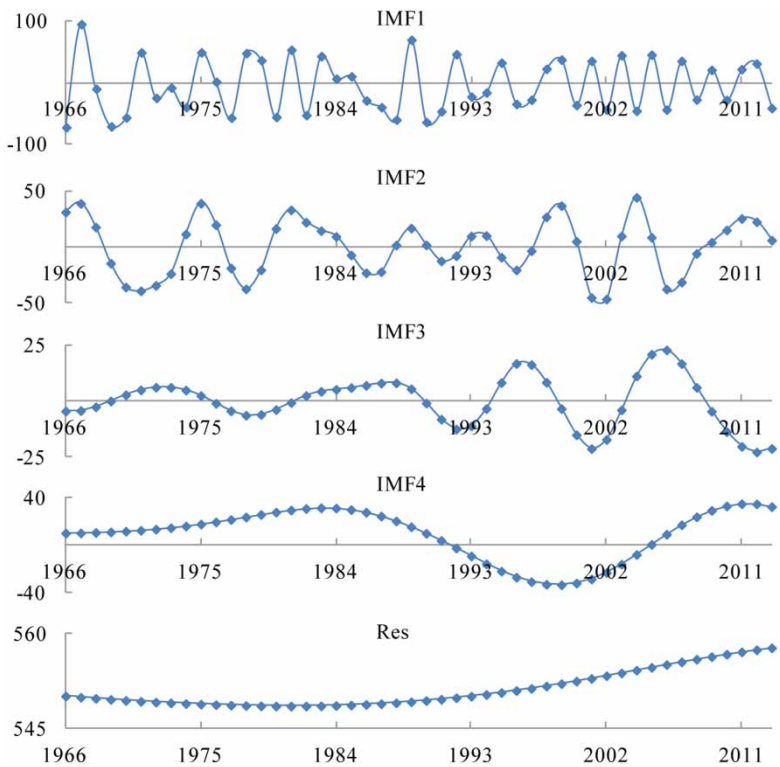


Figure 4 | The decomposed components of rainfall.

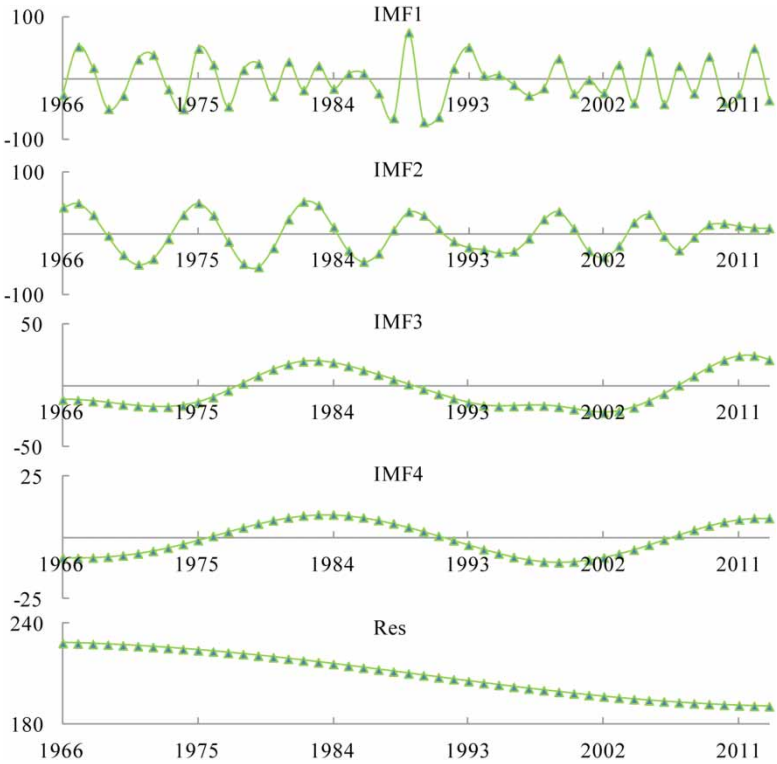


Figure 5 | The decomposed components of runoff.

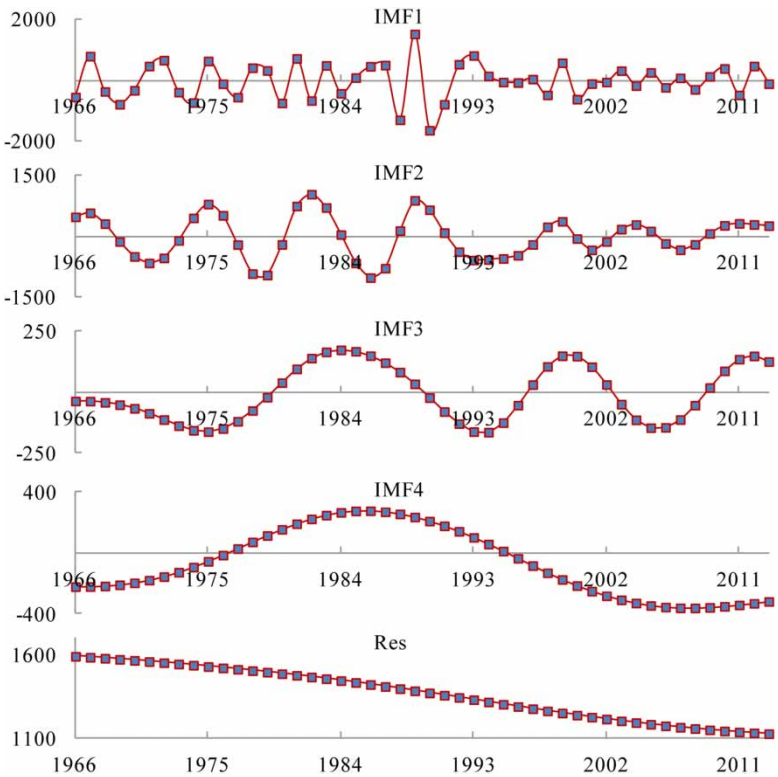


Figure 6 | The decomposed components of sediment.

Table 2 | Periodic changes of rainfall, runoff and sediment component time series

Component time series	Periodic changes (year)/Res changes		
	Rainfall	Runoff	Sediment
IMF1	2–5	2–5	2–5
IMF2	5–8	6–9	5–10
IMF3	9–11	29–30	11–30
IMF4	28	32	41
Res	First reduce and then increase	Reduce	Reduce

Co-integration analysis

Stationary test

The OTS and components of rainfall, runoff and sediment in the source area of the Yellow River are tested by the unit root test. It is assumed that x_i, z_i and y_i ($i = 0, 1, 2, 3, 4, 5$) are used to represent the CEEMDAN component of rainfall, runoff and sediment, and x_0, z_0 and y_0 are their original sequences, respectively. The optimal lag order is determined by the Akaike information criterion (AIC), and the unit root test results are given in Table 3.

The ADF test values of the OTS of rainfall, runoff and sediment in the source area of the Yellow River are all larger than the critical value of *t*-test, so they belong to non-stationary time series, but their first-order difference time series are stationary. Meanwhile, their CEEMDAN components are stationary.

Co-integration test

The co-integration test is conducted on their OTS and CEEMDAN decomposition sequence of rainfall, runoff and sediment with the E.G. two-step method. The first step is to perform OLS regression on the original sequence and the same time scale component sequence of rainfall, runoff and sediment to establish a co-integration equation. The second step is to perform a unit root test on the residuals of each co-integration equation. If the residual

Table 3 | Unit root test results of the OTS and components

Time series	Variables	ADF value	Test type (c, t, k)	Test critical values			Stationary or not
				1%	5%	10%	
The original	x_0	−0.5138	(0, 0, 3)	−2.6186	−1.9485	−1.6121	No
	y_0	−0.8657	(0, 0, 3)	−2.6186	−1.9485	−1.6121	No
	z_0	−0.3647	(0, 0, 3)	−2.6186	−1.9485	−1.6121	No
	Δx_0	−7.4505	(0, 0, 2)	−2.6186	−1.9485	−1.6121	Yes
	Δy_0	−6.2861	(0, 0, 2)	−2.6186	−1.9485	−1.6121	Yes
	Δz_0	−6.5489	(0, 0, 2)	−2.6186	−1.9485	−1.6121	Yes
The IMF1	x_1	−7.8693	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	y_1	−7.6006	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	z_1	−8.3002	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
The IMF2	x_2	−10.8127	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	y_2	−14.1319	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	z_2	−14.2495	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
The IMF3	x_3	−14.6845	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	y_3	−10.0076	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
	z_3	−8.6862	(c, 0, 1)	−3.5812	−2.9266	−2.6014	Yes
The IMF4	x_4	−26.8800	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes
	y_4	−23.9409	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes
	z_4	−26.7954	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes
The residual	x_5	−20.3586	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes
	y_5	−13.4521	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes
	z_5	−25.1841	(c, t, 1)	−4.1706	−3.5107	−3.1855	Yes

Note: In the test type (c, t, k), c is the intercept item, t is the time trend term (t = 0 means no trend) and k is the optimal lag length.

sequence is stationary, the co-integration relationship exists; otherwise, this relationship does not exist. It can be seen from Table 4 that the ADF test values of the residual sequences of all co-integration equations are less than the critical values of the significance levels of 1, 5 and 10%, so the co-integration relationship exists.

$$z_0 = 0.259236\,x_0 + 0.044719\,y_0 + u_0 \quad (11)$$
$$(0.014418)\quad (0.005056)$$

$$z_1 = 0.265928\,x_1 + 0.036010\,y_1 + u_1 \quad (12)$$
$$(0.071956)\quad (0.005192)$$

$$z_2 = 0.377995\,x_2 + 0.042713\,y_2 + u_2 \quad (13)$$
$$(0.094843)\quad (0.004324)$$

$$z_3 = 0.184838\,x_3 + 0.073507\,y_3 + u_3 \quad (14)$$
$$(0.220190)\quad (0.017521)$$

$$z_4 = 0.109124\,x_4 + 0.022741\,y_4 + u_4 \quad (15)$$
$$(0.025683)\quad (0.002785)$$

$$z_5 = 0.159144\,x_5 + 0.088938\,y_5 + u_5 \quad (16)$$
$$(0.001172)\quad (0.000456)$$

In the formula, u_t represents the residual sequence of the equation, and the data in brackets are the standard deviation of the corresponding coefficient of the equation.

Establishing ECM

According to the ECM method, the ECM-OTS for x_0, z_0 and y_0 and the ECM-CEEMDAN model for x_i, z_i and y_i ($i = 1, 2, 3, 4, 5$) is as follows:

$$\Delta z_0 = 0.236650\Delta x_0 + 0.037671\Delta y_0 - 0.700574 ecm_0(-1) - 0.813728 \quad (17)$$
$$(0.061223)\quad (0.004417)\quad (0.149456)\quad (3.221133)$$

$$\Delta z_1 = 0.221667\Delta x_1 + 0.035227\Delta y_1 - 1.154164 ecm_1(-1) - 0.262368 \quad (18)$$
$$(0.045599)\quad (0.003267)\quad (0.164667)\quad (2.288407)$$

$$\Delta z_2 = 0.204791\Delta x_2 + 0.040942\Delta y_2 - 0.636489 ecm_2(-1) - 0.213539 \quad (19)$$
$$(0.075293)\quad (0.004075)\quad (0.143114)\quad (1.268256)$$

$$\Delta z_3 = 0.153765\Delta x_3 + 0.035988\Delta y_3 - 0.004766 ecm_3(-1) - 0.010384 \quad (20)$$
$$(0.073105)\quad (0.010246)\quad (0.040804)\quad (0.482288)$$

$$\Delta z_4 = 0.276363\Delta x_4 + 0.015486\Delta y_4 - 0.090239 ecm_4(-1) - 0.076207 \quad (21)$$
$$(0.013844)\quad (0.001382)\quad (0.025256)\quad (0.061591)$$

$$\Delta z_5 = 0.718452\Delta x_5 + 0.152342\Delta y_5 - 0.207623 ecm_5(-1) - 0.631328 \quad (22)$$
$$(0.061745)\quad (0.005481)\quad (0.023369)\quad (0.051960)$$

In the formula, $ecm_t(-1)$ represents the error correction term, and the coefficient before $ecm_t(-1)$ is the short-period adjustment coefficient, and the coefficient before the difference terms of each variable represents the short-period dynamic change of the model.

It can be seen that the rainfall, runoff and sediment in the source area of the Yellow River show a long-term equilibrium relationship. The component time series also has a long-term equilibrium relationship at different time scales,

Table 4 | Unit root test results of residual sequences of co-integration equations

Residual sequences	ADF value	Test type (c, t, k)	Test critical values			Stationary or not
			1%	5%	10%	
u_0	−4.6011	(c, 0, 1)	−3.6105	−2.9390	−2.6079	Yes
u_1	−6.8268	(c, 0, 1)	−3.6105	−2.9390	−2.6079	Yes
u_2	−7.7600	(c, 0, 1)	−3.6156	−2.9411	−2.6091	Yes
u_3	−6.1780	(c, 0, 1)	−3.6156	−2.9411	−2.6091	Yes
u_4	−8.7148	(c, 0, 1)	−3.6156	−2.9411	−2.6091	Yes
u_5	−5.5610	(c, 0, 1)	−3.6156	−2.9411	−2.6091	Yes

and the error correction term coefficients of all equations are all negative, which is consistent with the reverse correction mechanism. It can be seen from Equation (17) that runoff is not only affected by rainfall and sediment but also by the deviation of runoff from the equilibrium level in the previous year. The coefficients of Δx_0 and Δy_0 are 0.23665 and 0.03767, respectively, which indicates that the short-term influence of rainfall and sediment on runoff in the source area of the Yellow River is different, and the influence of rainfall is stronger than that of sediment. The coefficient before $ecm_t(-1)$ is -0.70057, which indicates that the deviation of runoff from equilibrium in this year will be adjusted by 70.06% in the next year.

Annual runoff forecast

The ECM-OTS and the ECM-CEEMDAN models of the annual runoff are established by using the measured data series of rainfall, runoff and sediment from 1966 to 2005, and the forecast test is conducted with the measured data series from 2006 to 2013. Figure 6 shows the fitting between the measured value and the fitted value of the two models. Figure 7 shows the relative error between the fitting value and the measured value of the two models. Table 6 shows the forecasted values and relative errors of the two models during the forecast period.

It can be seen from Figure 7 that both models can well describe the dynamic equilibrium relationship between rainfall, runoff and sediment in the source area of the Yellow River. Moreover, the accuracy of runoff fitting value of the ECM-CEEMDAN model is better than that of ECM-OTS.

It can be seen from Figure 8 that in the year that the relative error is greater than 20% from 1967 to 2005, the ECM-CEEMDAN model has only one 28.11% in 2002, but the ECM-OTS model has two years, 20.83% in 1997 and 32.17% in 2002. The average relative error of the ECM-CEEMDAN model is 6.21%, which is 1.42% lower than the 7.63% of the ECM-OTS model. It can be seen that the ECM-CEEMDAN model has better fitting accuracy.

According to the Standard for Hydrological Information and Hydrological Forecasting (GB/T 22482-2008) of China, 20% of the measured value is taken as the allowable error for runoff forecasting. When the error is less than the allowable error, it is available. The percent of qualified forecast times and total forecast times is the qualified rate of forecast. Meanwhile, the degree of agreement between the runoff forecasting process and the measured process can be evaluated by the deterministic coefficient, which is calculated as follows:

$$DC = 1 - \frac{\sum_{i=1}^{n} [y_c(i) - y_0(i)]^2}{\sum_{i=1}^{n} [y_0(i) - \overline{y_0}]^2} \tag{23}$$

In the formula, DC is the deterministic coefficient, $y_0(i)$ is the measured value, $y_c(i)$ is the forecasted value, $\bar{y}_0$ is the mean of the measured values, and n is the length of the sequence.

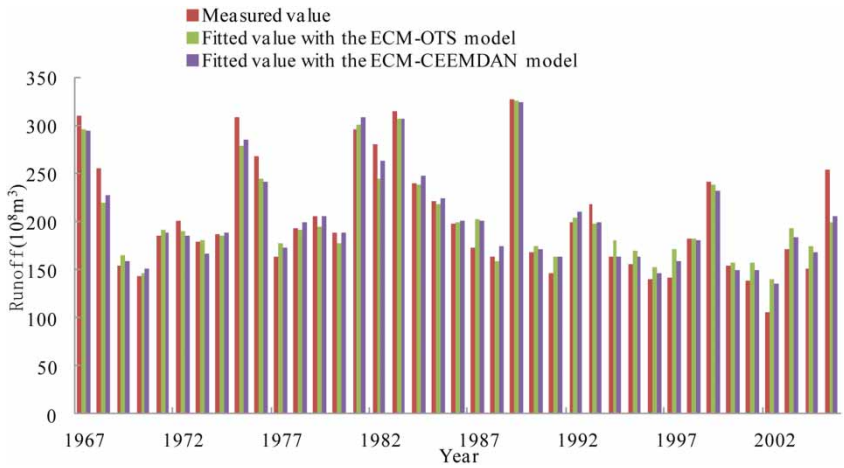


Figure 7 | Fitting between the fitted value and measured value of two models.

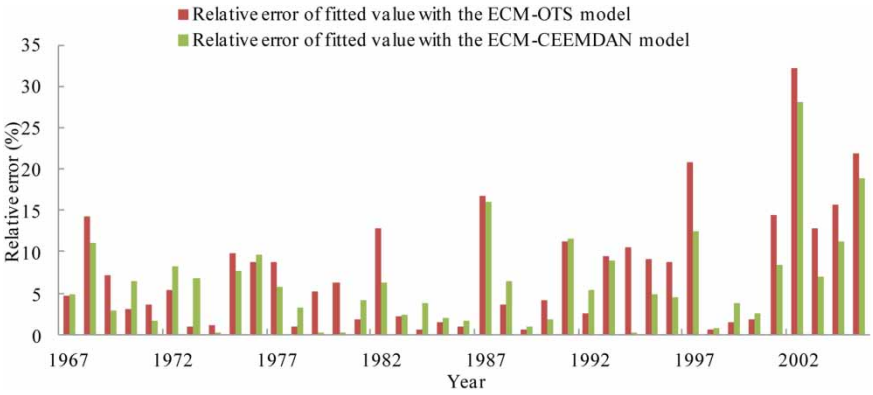


Figure 8 | Relative error between the fitted value and the measured value of the two models.

The accuracy of runoff forecast is divided into three grades according to the qualification rate or the deterministic coefficient, as shown in Table 5.

It can be seen from Table 6 that for the ECM-OTS model, in the forecasted 8 years of 2006–2013, only the relative error of runoff in 2009 exceeded 20%, and its forecasted qualified rate was 87.5%, reaching the level A. Meanwhile, for all predicted years, the relative error of runoff forecast that is less than 10% is 5 years, accounting for 62.5%. While for the ECM-CEEMDAN model, its forecasted relative error in 2009 only is 18.81% which is close to 20%. The whole forecasted qualified rate was 100%. Although the ECM-CEEMDAN model has the same as the ECM-OTS model, with the runoff forecast relative error of 10% in 5 years, its relative error value tends to be smaller on the whole, which indicates that the overall forecast accuracy of the ECM-CEEMDAN model is better. Moreover, the average relative error of the ECM-CEEMDAN model is 8.59%, which is 2.7% lower than 11.29% of the ECM-OTS model. This shows that the ECM-CEEMDAN model has a higher forecast accuracy than the ECM-OTS model.

Furthermore, from the deterministic coefficient of runoff forecast, the DC value of the ECM-OTS model is 0.842, which is the level B, while the DC value of the ECM-CEEMDAN model is 0.901, reaching the level A. This shows that the ECM-CEEMDAN model has the higher degree of agreement between the runoff forecasting process and the measured process.

Table 5 | Runoff forecast accuracy class table

Accuracy class	A	B	C
Pass rate/%	$QR \geq 85$	$85 > QR \geq 70$	$70 > QR \geq 60$
Deterministic coefficient	$DC > 0.9$	$0.9 \geq DC > 0.7$	$0.7 > DC \geq 0.5$

Table 6 | The forecasted values and relative errors of the two models during the forecast period

		ECM-OTS model		ECM-CEEMDAN model	
Year	Measured value (10^8 m^3)	Predicted value (10^8 m^3)	Relative error (%)	Predicted value (10^8 m^3)	Relative error (%)
2006	141.26	164.72	16.60	157.08	11.19
2007	189.04	177.12	6.31	181.34	4.07
2008	174.60	157.93	9.55	165.21	5.37
2009	263.48	197.06	25.21	213.92	18.81
2010	197.08	210.72	6.92	209.32	6.21
2011	211.21	198.11	6.20	193.63	8.32
2012	284.04	232.62	18.10	249.97	11.99
2013	194.64	191.92	1.40	189.24	2.77

CONCLUSION

(1) The CEEMDAN method can reveal the periodic characteristics of rainfall, runoff and sediment on the multi-time scales in the source area of the Yellow River. These three variables have a good correlation in the short and the medium periods. In addition, runoff and sediment show a better synchronization in the trend

item, which reveals the law of periodic fluctuations of rainfall, runoff and sediment.

(2) With the co-integration theory and ECM, the ECM-OTS model and the ECM-CEEMDAN model are established. They can reveal the long-term equilibrium and short-term fluctuations of the original sequence and component sequence of rainfall, runoff and sediment in the source area of the Yellow River, and can also effectively forecast the runoff in this source area.

(3) Both the ECM-OTS model and the ECM-CEEMDAN model can well describe the dynamic equilibrium relationship between rainfall, runoff and sediment in the source area of the Yellow River. However, the forecast period error of the ECM-CEEMDAN model is less than 20%, and its forecast qualified rate can reach 100%, and the accuracy reaches the level A. Compared with the ECM-OTS model, it has the better forecasting accuracy, which provides a new and more accurate runoff forecasting method.

Although the ECM-CEEMDAN of rainfall, runoff and sediment has the higher prediction accuracy, rainfall, runoff and sediment in practice they are often showing the nonlinear relations affected by other many factors, such as underlying surface. More efforts are needed to reveal the nonlinear relations between rainfall, runoff and sediment. Anyway, the combination of the CEEMDAN method and co-integration theory in this paper provides a better analysis method to reveal the internal periodic changes of the hydrological variable, so it can better reflect the actual characteristic of the hydrological variable, which is available for the hydrological forecasting and water resources management.

ACKNOWLEDGEMENTS

This research is supported by the National Key R&D Program of China (Grant No. 2018YFC0406501), Program for Innovative Talents (in Science and Technology) at University of Henan Province (Grant No. 18HASTIT014) and Foundation for University Youth Key Teacher of Henan Province (Grant No. 2017GGJS006).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Adarsh, S. & Reddy, M. J. 2018 Evaluation of trends and predictability of short-term droughts in three meteorological subdivisions of India using multivariate EMD-based hybrid modeling. *Hydrological Processes* **33** (1), 130–143.

Antico, A., Torres, M. E. & Diaz, H. F. 2016 Contributions of different time scales to extreme Parana floods. *Climate Dynamics* **46** (11–12), 3785–3792.

Chen, X. & Guo, W. 2016 Variation analysis of hydrological situation in the middle reaches of Yellow River after reservoirs construction based on range of variability approach. *Water Resources and Power* **34** (11), 5–8.

Chu, H. B., Wei, J. H., Qiu, J. & Wang, G. Q. 2019 Identification of the impact of climate change and human activities on rainfall-runoff relationship variation in the Three-River Headwaters region. *Ecological Indicators* **106**, 105516.

Colominas, M. A., Schlotthauer, G. & Torres, M. E. 2014 Improved complete ensemble EMD: a suitable for tool biomedical signal processing. *Biomedical Signal Processing and Control* **14** (11), 19–29.

Dickey, F. & Fuller, W. A. 1979 Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

El Bouny, L., Khalil, M. & Adib, A. 2019 ECG signal filtering based on CEEMDAN with hybrid interval thresholding and higher order statistics to select relevant modes. *Multimedia Tools and Applications* **78** (10), 13067–13089.

Engle, R. F. & Granger, C. W. J. 1987 Co-integration and error correction: representation, estimation and testing. *Econometrica* **55** (2), 251–276.

Guru, N. & Jha, R. 2019 Application of soft computing techniques for river flow prediction in the downstream catchment of Mahanadi River Basin using partial duration series, India. *Iranian Journal of Science and Technology. Transactions of Civil Engineering* **44**, 279–297.

Hou, S. Z., Wang, P., Guo, Y. & Chu, W. B. 2013 Analysis of water and sediment regulation of Longyangxia reservoir. *Journal of Hydroelectric Engineering* **32** (6), 151–156.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, N. Y., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **454**, 903–995.

Jin, H., Zhang, S. & Zhang, J. S. 2017 Spurious regression due to neglected of non-stationary volatility. *Computation Statistics* **32** (3), 1065–1081.

Lee, L. F. & Yu, J. H. 2009 Spatial nonstationarity and spurious regression: the case with a row-normalized spatial weights matrix. *Spatial Economic Analysis* **4** (3), 301–327.

Li, J. Z. & Liu, L. B. 2018 Analysis on the sediment retaining amount by warping dams above Tongguan section of the Yellow River in recent years. *Yellow River* **40** (01), 1–6.

Ling, H. B., Deng, X. Y., Long, A. H. & Gao, H. F. 2017 The multi-time-scale correlations for drought-flood index to runoff and North Atlantic Oscillation in the headstreams of Tarim River, Xinjiang, China. *Hydrology Research* **48** (1), 253–264.

Meng, X. M., Yin, M. S., Ning, L. B., Liu, D. F. & Xue, X. W. 2015 A threshold artificial neural network model for improving runoff prediction in a karst watershed. *Enviornmental Earth Sciences* **74** (6), 5039–5048.

Mosavi, A., Ozturk, P. & Chau, K. W. 2018 Flood prediction using machine learning models: literature review. *Water* **10** (11), 1536.

Nastiti, K. D., An, H., Kim, Y. & Jung, K. 2018 Large-scale rainfall–runoff–inundation modeling for upper Citarum River watershed, Indonesia. *Environmental Earth Sciences* **77** (18), 640.

Ouyang, Q., Lu, W. X., Xin, X., Zhang, Y., Cheng, W. G. & Yu, T. 2016 Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. *Water Resource Management* **30** (7), 2311–2325.

Ramana, R. V., Krishna, B., Kumar, S. R. & Pandey, N. G. 2013 Monthly rainfall prediction using wavelet neural network analysis. *Water Resource Management* **27** (10), 3697–3711.

Rezaie-Balf, M., Zahmatkesh, Z. & Kim, S. 2017 Soft computing techniques for rainfall-runoff simulation: local non–parametric paradigm vs. model classification methods. *Water Resources Management* **31**, 3843–3865.

Sezen, C. & Partal, T. 2019 The utilization of a GR4 J model and wavelet-based artificial neural network for rainfall–runoff modeling. *Water Science and Technology – Water Supply* **19** (5), 1295–1304.

Tarasova, L., Basso, S., Zink, M. & Merz, R. 2018 Exploring controls on rainfall-runoff events: 1. Time series-based event separation and temporal dynamics of event runoff response in Germany. *Water Resources Research* **54** (10), 7711–7732.

Torres, M. E., Colominas, M. A., Schlotthauer, G. & Flandrin, P. 2011 A complete ensemble empirical mode decomposition with adaptive noise. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Prague, Czech, pp. 4144–4147.

Wang, F., Hessel, R., Mu, X. M., Maroulis, J., Zhao, G. J., Geissen, V. & Ritsema, C. 2015a Distinguishing the impacts of human activities and climate variability on runoff and sediment load change based on paired periods with similar weather conditions: a case in the Yan River, China. *Journal of Hydrology* **527**, 884–893.

Wang, W. C., Chau, K. W., Xu, D. M. & Chen, X. Y. 2015b Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resource Management* **29** (8), 2655–2675.

Wang, S., Fu, B. J., Liang, W., Liu, Y. & Wang, Y. F. 2017 Driving forces of changes in the water and sediment relationship in the Yellow River. *Science of the Total Environment* **576**, 453–461.

Wang, X. Y., Yang, T., Yong, B., Krysanova, V., Shi, P. F., Li, Z. Y. & Zhou, X. D. 2018a Impacts of climate change on flow regime and sequential threats to riverine ecosystem in the source region of the Yellow River. *Environmental Earth Sciences* **77**, 465.

Wang, Y. K., Rhoads, B. L., Wang, D., Wu, J. C. & Zhang, X. 2018b Impacts of large dams on the complexity of suspended sediment dynamics in the Yangtze River. *Journal of Hydrology* **558**, 184–195.

Wang, Y. W., Shen, Z. Z. & Jiang, Y. 2019 Comparison of autoregressive integrated moving average model and generalised regression neural network model for prediction of haemorrhagic fever with renal syndrome in China: a time-series study. *BMJ Open* **9** (6), e025773.

Wu, Z. H. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis, Theory and Applications* **1** (1), 1–41.

Wu, J., Liu, H., Wei, G. Z., Song, T. Y., Zhang, C. & Zhou, H. C. 2019 Flash flood forecasting using support vector regression model in a small mountainous catchment. *Water* **11** (7), 1327.

Xie, T., Zhang, G., Hou, J. W., Xie, J. C., Meng, L. & Liu, F. C. 2019 Hybrid forecasting model for non-stationary daily runoff series: a case study in the Han River Basin, China. *Journal of Hydrology* **577**, 123915.

Yaseen, Z. M., Allawi, M. F., Yousif, A. A., Jaafar, O., Hamzah, F. M. & El-Shafie, A. 2018 Non-tuned machine learning approach for hydrological time series forecasting. *Neural Computing and Applications* **30** (5), 1479–1491.

Yoo, S. H. 2007 Urban water consumption and regional economic growth: the case of Taejeon, Korea. *Water Resource Management* **21** (8), 1353–1361.

Zhang, J. P., Ding, Z. H., Yuan, W. L. & Zuo, Q. T. 2013a Research on the relationship between rainfall and reference crop evapotranspiration with multi-time scales. *Paddy and Water Environment* **11** (1–4), 473–482.

Zhang, J. P., Yuan, W. L. & Guo, B. T. 2013b Study on prediction of stream flow based on cointegration theory. *Water Resources and Power* **31** (05), 18–20, 99.

Zhang, J. P., Ding, Z. H. & You, J. J. 2014 The joint probability distribution of runoff and sediment and its change characteristics with multi-time scales. *Journal of Hydrology and Hydromechanics* **62** (3), 218–225.

Zhang, J. P., Zhao, Y. & Xiao, W. H. 2015 Multi-resolution cointegraton prediction for runoff and sediment load. *Water Resource Management* **29**, 3601–3613.

Zhang, J. P., Li, Y. Y., Zhao, Y. & Hong, Y. 2017a Wavelet-cointegration prediction of irrigation water in the irrigation district. *Journal of Hydrology* **544**, 343–351.

Zhang, J. P., Zhao, Y. & Lin, X. M. 2017b Uncertainty analysis and prediction of river runoff with multi-time scales. *Water Science and Technology – Water Supply* **17** (3), 897–906.

Zhang, J. P., Li, H. B., Shi, X. X. & Hong, Y. 2019a Wavelet-nonlinear cointegration prediction of irrigation water in the irrigation district. *Water Resource Management* **33** (8), 2941–2954.

Zhang, J. P., Xiao, H. L., Zhang, X. & Li, F. W. 2019b Impact of reservoir operation on runoff and sediment load at multi-time scales based on entropy theory. *Journal of Hydrology* **569**, 809–815.

Zhao, X. H., Chen, X., Xu, Y. X., Xi, D. J., Zhang, Y. B. & Zheng, X. Q. 2017 An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* **9** (3), 153.

First received 24 May 2020; accepted in revised form 24 July 2020. Available online 14 September 2020