Prediction of sediment concentration based on the MEEMD-ARIMA model in the lower Yellow River

Xianqi Zhang, Fei Liu, Chao Song and Xiaoyan Wu

ABSTRACT

There are many factors influencing the evolution of sediment concentration, and it is difficult to determine and extract, which brings great difficulties to the high-precision prediction of sediment concentration. Accurate prediction of annual sediment concentration in the lower Yellow River can provide a theoretical basis for flood control and disaster reduction and rational utilization of water and soil resources in the lower Yellow River. For the defects of pseudo-components in data decomposition of Complementary EEMD, the Modified EEMD (MEEMD) method proposed in this paper has the advantage of eliminating pseudo components of IMF and reducing non-stationarity of sediment bearing sequences. Then, combined with the Autoregressive Integrated Moving Average (ARIMA) model with strong approximation ability to the stationary sequence, the MEEMD-ARIMA model for predicting the annual sediment concentration in the lower Yellow River was constructed. Through fitting and predicting the annual sediment concentration in Gaocun Station, it is shown that the model not only considers the evolution of sediment concentration in various frequency domains, but also solves the problem that the ARIMA model requires sequence to be stable, the relative error of prediction is within ±6%, and the prediction accuracy is high, thus providing a new method for the prediction of sediment concentration.

Key words | annual sediment concentration, lower Yellow River, MEEMD-ARIMA model, prediction

Xianqi Zhang
Fei Liu (corresponding author)
Chao Song
Xiaoyan Wu
North China University of Water Resources and Electric Power,
Zhengzhou 450046,
China
E-mail: *772827492@qq.com*

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

INTRODUCTION

The fitting and prediction of sediment concentration sequences has always been the focus of sediment research, and it plays an important role in analyzing the evolution of sediment yield and sediment transport in the basin (Chen *et al.* 2017). However, the evolution of sediment concentration is influenced by many factors, and it is random and uncertain, which brings great difficulties to the prediction of sediment concentration. At present, scholars at home and abroad have carried out much research in this field, and have achieved fruitful results in the method of fitting and predicting the sediment concentration sequence. Joshi *et al.* (2015) used an artificial neural network to establish the sediment concentration prediction model of Gangotri and successfully predicted the evolution of sediment concentration in Gangotri in the Himalayas. Tilahun *et al.* (2015) predicted the evolution of sediment concentration in Debre Mawi watershed by the Parameter Efficient Distributed (PED) model. Chen & Dyke (1998) proposed a multivariate time series model suitable for predicting sediment concentration and transport and verified its accuracy. Miao *et al.* (2012) predicted the sediment concentration change in Lanzhou Station of the Yellow River by using Particle Swarm Optimization (PSO) and the Back Propagation (BP) neural network coupled model. Yao *et al.* (2015) analyzed sediment movement and predicted sediment content changes through flume experiments. Chen (2005) studied and predicted the annual average sediment concentration in Wujiang River basin through a BP artificial

doi: 10.2166/wcc.2019.077

neural network model. Based on the research results at home and abroad at the present stage, the main methods used in sediment concentration prediction include mathematical statistics (Itai & Amit 2018), hydraulic models (Li & Shang 1998), hydrological methods (Shi et al. 2008) and artificial neural networks (Barua et al. 2010). However, the traditional mathematical statistics model does not deal well with some high frequency mutation data; the hydraulic prediction model needs to consider more factors and collect more data; the hydrological method does not pay attention to the physical mechanism of sediment content change, and the prediction performance is vulnerable to the conditions of sediment content change; the neural network also has the shortcomings of over-training, which makes the difference between the model output and the actual situation too large. Therefore, it is necessary to further optimize and improve the prediction model.

There are many pre-influence factors involved in the evolution of sediment concentration, and it is characterized by randomness and non-stationarity, which puts forward higher requirements for scientific and accurate prediction of sediment concentration (Vigiak et al. 2015). The Ensemble Empirical Mode Decomposition (EEMD) method was proposed by Wu & Huang (2009), which is mainly a noise auxiliary data analysis method for the shortcomings of the Empirical Mode Decomposition (EMD) method. Yeh & Shieh (2010) further optimized and improved on the basis of EEMD and proposed the Complementary EEMD (CEEMD) method. Although CEEMD can decompose non-stationary time series and reduce the non-stationarity of the sequence, there are also defects of pseudo components. Therefore, reducing the non-stationarity of sediment concentration series by constructing a coupled prediction model became a new method to improve the prediction accuracy of sediment concentration in the lower Yellow River.

In view of the above problems, this paper proposes a Modified EEMD (MEEMD) algorithm based on the CEEMD method, which not only solves the problem of the Intrinsic Mode Function (IMF) pseudo component, but also reduces the non-stationarity of the sequence. Then the MEEMD algorithm is combined with the Autoregressive Integrated Moving Average (ARIMA) Model. By using the ARIMA to have a strong approximation ability to the stationary sequence (Myronidis et al. 2018), the MEEMD-ARIMA coupling model for predicting the annual sediment concentration in the lower Yellow River is constructed. The MEEMD algorithm is used to decompose the sediment concentration sequence and eliminate the pseudo component, then the ARIMA model is used to predict the sediment concentration. The obtained results are reasonable and the prediction accuracy is high, which can provide a scientific basis for rationally dispatching reservoirs, mitigating reservoir siltation, and maximizing the benefits of the reservoir.

THEORIES AND METHODS

The theory of CEEMD

Huang et al. (1998) proposed Empirical Mode Decomposition (EMD). It is an adaptive decomposition algorithm for non-linear and non-stationary signals. This method has been affirmed and widely used in academia since it was put forward. At present, it has been applied in many fields such as image processing, signal denoising and financial prediction. This method not only breaks through the limitation of Fourier transform, but also does not need to pre-select the wavelet basis function like wavelet transform. The EMD method has good time-frequency resolution and adaptability, but it will produce mode mixing in signal decomposition, which will affect the decomposition effect.

Wu & Huang (2009) put forward the EEMD model on the basis of EMD. EEMD is mainly a noise auxiliary data analysis method aimed at the deficiency of mode mixing in EMD methods, which aims to eliminate mode mixing by adding original white noise into original sequence signals. Although EEMD can eliminate the phenomenon of mode mixing, it also has some shortcomings, such as poor completeness, and a large amount of computation is needed to eliminate reconstruction errors.

With the development of mathematics, as well as people's continuous efforts, some scholars have improved EEMD to improve its performance. Yeh & Shieh (2010) proposed CEEMD on the basis of EEMD. The CEEMD method mainly adds two opposite white noise signals to the signal to be analyzed and decomposes them by EMD, which reduces the reconstruction error caused by white noise while ensuring the equivalent decomposition effect as EEMD. Although CEEMD solves the problem of poor EEMD completeness, the decomposition of the signal depends on the selection of noise amplitude and integration times, and pseudo components (components without physical meaning) are prone to appear. Therefore, it is proposed to use the permutation entropy (PE) method to improve it, and eliminate the IMF pseudo component of CEEMD decomposition.

Permutation entropy algorithm

Detecting the randomness of signals (i.e. detecting the pseudo components of IMF components) is the key to improve CEEMD. There are many existing detection methods (Borgnat *et al.* 2010; Terrien *et al.* 2011), but there are more or less deficiencies in eliminating the pseudo component.

The PE algorithm is proposed by Bandt & Pompe (2002), which is a theory based on physical mechanism and reflects the inherent law of system evolution. It shows good robustness in detecting the randomness and mutagenicity of nonlinear time series (Riedl *et al.* 2013). Therefore, the PE algorithm has inherent advantages in the detection of signal randomness and the elimination of the pseudo component. The steps of the PE algorithm are shown below (Yan & Gao 2007):

- The time series $\{x(i), i = 1, 2, \ldots .N\}$ with length N is reconstructed in phase space, and the following sequence is obtained:

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

- Each row in the matrix is treated as a reconstructed component. m and λ represent embedding dimension and time delay respectively, where $K = (m-1)\lambda$.
- Each reconstructed component is arranged in order from small to large, i.e.

$$X(i) = \{x(i+(j_1-1)\lambda) \le x(i+(j_2-1)\lambda) \le \ldots x(i+(j_m-1)\lambda)\} \tag{2}$$

If there are $x(i+(j_{i1}-1)\lambda) = x(i+(j_{i2}-1)\lambda)$ in the sorting, they can be arranged in the order of j from small to large, that is, when $j_{i1} < j_{i2}, x(i+(j_{i1}-1)\lambda) = x(i+(j_{i2}-1)\lambda)$.

Therefore, for any vector *X(i)*, the following sequence of symbols must exist:

$$S(k) = [j_1, j_2, j_3 \ldots j_m] \tag{3}$$

where $k = 1, 2, 3 \ldots l$, $l \le m!$. $[j_1, j_2, j_3 \ldots j_m]$ corresponds to $m!$ kinds of arrangements, so there are $m!$ kinds of symbol sequences. Assuming that the probability of occurrence of each symbol sequence is P_i, then $\sum_{i=1}^{l} P_i = 1$.

According to Shannon's definition of information entropy, the PE of sequence $\{x(i), i = 1, 2, \ldots. N\}$ can be defined as:

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

when $H_{PE}(m)$ reaches the maximum value $In(m!)$ at $P_i = \frac{1}{m}$, the normalized form of the permutation $H_{PE}(m)$ is:

$$H_{PE} = H_{PE}(m)/In(m!) \tag{5}$$

where $0 \le H_{PE} \le 1$. H_{PE} can reflect the randomness of time series. The larger H_{PE} is, the stronger the randomness of time series is, and the higher the probability of mutation. On the contrary, the weaker the randomness of time series is, the smaller the probability of mutation.

According to Bandt's suggestion, the embedding dimension m ranges from 3 to 7. However, if m is too small, the reconstructed vector contains less information and the algorithm is meaningless; if m is too large, the phase space reconstruction will homogenize the time series, and the subtle changes of the series will be weakened. In this paper, $m = 5$ is selected. Since the value of time delay λ has little effect on PE, $\lambda = 1$ is taken in this paper.

MEEMD modeling steps

Although the EEMD method can eliminate mode mixing to a certain extent, its computational complexity is large, the added white noise cannot be completely neutralized, and its completeness is poor. The CEEMD method adds white noise with opposite symbols to the target signal in pairs, which greatly reduces the reconstruction error. Combined with the CEEMD and PE algorithm, the Modified EEMD method (MEEMD) is proposed. After detecting abnormal components of CEEMD decomposition, the MEEMD method directly performs EMD decomposition. This method can not only suppress the mode mixing during EMD decomposition, but also reduces the computation and reconstruction error.

The MEEMD method modeling steps based on PE are as follows:

1. In the original signal $x(t)$, white noise signals $C_i(t)$ and $-C_i(t)$ with zero mean value are added, that is:

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

where $C_i(t)$ is additional white noise signal; a_i is amplitude of noise signal, $(i = 1, 2, 3 \ldots .. Ne)$; N_e is logarithm of adding white noise.

2. By EMD decomposition of $x_i^+(t)$ and $x_i^-(t)$, the first order IMF component sequences $\{I_{i1}^+(t)\}$ and $\{I_{i1}^-(t)\}$ $(i = 1, 2, 3 \ldots .. Ne)$ are obtained.
3. Average the IMF components obtained in step (2), as shown below:

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

4. The PE algorithm is used to detect whether $I_1(t)$ is an abnormal signal. If the PE value of IMF component is greater than the threshold θ_0, then the component is an abnormal component. Otherwise, the IMF component is considered to be an approximate stationary signal.
5. If $I_1(t)$ is an abnormal component, steps (1)–(4) are continued until the IMF component $I_p(t)$ is not an abnormal component.
6. The anterior $n-1$ decomposed components are extracted from the original signal, which is:

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

ARIMA modeling principles and steps

ARIMA modeling principles

The ARIMA was proposed by Box et al. (2015) in the 1970s, and the model is widely used in time series analysis. By studying the probability distribution of noise, the data can be smoothed and normalized, so that the problem of random interference of sequence can be well solved. At present, the ARIMA forecasting model has been applied in many fields, including logistics demand, disease incidence forecasting, regional finance and hydrological information forecasting.

The modeling idea of ARIMA is to treat the predicted object as a random sequence and describe the sequence approximately with a certain mathematical model. Once the model is identified, the past value and present value of the sequence can be used to predict the future value.

In the ARIMA (p, d, q) model, the autoregressive model (AR) is an autoregressive component; MA is a moving average component; I is a difference; p is the order of autoregressive component; q is the order of moving average component; d is the difference number that makes time series stable.

The ARIMA model as follows:

$$\begin{aligned} y_t = {} & \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} - \theta_1 \xi_{t-1} - \theta_2 \xi_{t-2} \\ & - \ldots - \theta_q \xi_{t-q} + \xi_t \end{aligned} \tag{9}$$

where y is a time series, $\phi_1, \phi_2 \ldots \phi_p$ is an autoregressive coefficient, $\theta_1, \theta_2 \ldots \theta_q$ is a moving average coefficient, ξ_t is an error sequence, p is an autoregressive order ($p > 0$, and is an integer), and q is an autoregressive order ($q > 0$, and is an integer). If the resolution order is represented by d, the model can be represented as ARIMA (p, d, q).

In addition, there are two important functions in the ARIMA model, namely Autocorrelation function (ACF) and Partial Autocorrelation function (PACF). In the application process of the ARIMA model, we can identify and order selection for the ARIMA model according to the different morphology and change trends of the ACF and PACF images. There are three main types of model recognition: AR, Moving average model (MR) and ARIMA. The AR(p) model is actually equivalent to the ARIMA (p, 0, 0) model, while the MA(q) model is actually the ARIMA (0, 0, q) model.

AR(p) model is a linear model of P-order Markov hypothesis, and its expression is as follows:

$$\begin{cases} Xt = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t \\ \varphi_p \neq 0 \\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma_\varepsilon^2,\ E(\varepsilon_t \varepsilon_s) = 0,\ s \neq t \\ E(X_s \varepsilon_t) = 0,\ \forall s < t \end{cases} \tag{10}$$

where ε is the residual error, φ_p is the autoregressive parameter, and when $\varphi_0 = 0$, the model is the centralized AR(p) model.

The MA(q) model as follows:

$$\begin{cases} Xt = \varphi_0 + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \\ \qquad \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \\ \varphi_p \neq 0,\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma_\varepsilon^2,\ E(\varepsilon_t \varepsilon_s) = 0,\quad s \neq 0 \\ E(X_s \varepsilon_t) = 0,\ \forall s < t \end{cases} \tag{11}$$

where θ is the sliding average parameter, ε is the sequence residual, when $\theta_0 = 0$, the model is the centralized MA(q) model.

The basis for model recognition is shown in Table 1.

Among them, trailing refers to the monotonically decreasing or oscillating attenuation of the sequence at an exponential rate, while truncation refers to the sequence getting smaller suddenly from a certain point and keeping the state.

ARIMA modeling steps

The modeling steps of ARIMA model are as follows:

1. The stability of the sequence is identified by the scatter diagram and the autocorrelation and PACF diagram of the sequence.
2. Test the data stability. If the data sequence is non-stationary, the difference method or moving average method is used to process the data and make it become stationary data.
3. Corresponding models are established according to ACF and PACF. There are three situations to determine the parameters of ARIMA model: the first, ACF tailing, PACF truncation, using AR (p) model; the second, ACF truncation, PACF tailing, using MA (q) model; the third, ACF tailing, PACF tailing, using ARIMA (p, d, q) model.

Table 1 | ARIMA model recognition basis

Model	AR(p)	MA(q)	ARIMA (p, d, q)
ACF	Tailing	Truncation	Tailing
PACF	Truncation	Tailing	Tailing

MEEMD-ARIMA model

The MEEMD decomposition of the sediment concentration can reduce the non-stationarity of the sequence, which provides a stable premise for the prediction of the ARIMA model. The steps to use the MEEMD-ARIMA model are as follows:

1. Decomposing the sediment concentration sequence by MEEMD to obtain the IMF component, the residual term and the pseudo component sequence of the sediment concentration.
2. Judging whether IMF components, residual terms and pseudo components are stationary sequences. Stationary, d = 0; if not stationary, D is determined according to the order of difference. Then, p and q were determined according to the ACF and PACF diagrams.
3. The ARIMA model is used to fit and forecast the IMF component and the trend term. This paper uses the rolling prediction method to predict. Taking IMF_1 as an example, the IMF_1 data from 1960 to 2009 is used to predict the IMF_1 value for 2010, and then the IMF_1 value for 2011 is predicted from the IMF_1 data from 1961 to 2010 until the forecast to 2014.
4. Accumulating the predicted values of the IMF component and the trend term, plus the pseudo component, is the predicted value of the sediment concentration.

PREDICTION OF SEDIMENT CONCENTRATION

Data sources

Gaocun Hydrological Station is the first control station of the Yellow River flowing from Henan to Shandong

Province. Its cross section is a compound river bed with one shoal and one channel, and its catchment area is 734,000 km^2. Because its control section is in the lower Yellow River, and the lower Yellow River is mostly wandering, the evolution of sediment concentration is greatly affected by factors such as river potential, flow direction, erosion and siltation, and flood fluctuations. Therefore, the prediction of sediment concentration in the lower Yellow River is more difficult.

Data used in this study mainly include the measured sediment content data of Gaocun Hydrological Station from 1960 to 2014, as well as the Yellow River Water Resources Bulletin and China River Sediment Bulletin.

As can be seen from Figure 1, before 1999, the sequence showed irregular fluctuations. After 1999, the sediment concentration sequence showed a significant downward trend, which was caused by the large-scale reduction of sediment in the Yellow River caused by large-scale soil and water conservation projects in the middle and upper Yellow River (Xu 2004); In addition, when Xiaolangdi Reservoir regulates water and sediment, a large amount of sediment is transported into the sea by high sediment-laden flow (200–300 kg/m^3) (Shi & Wang 2003), which results in a significant reduction of sediment concentration during this period. From 2000 to 2014, the accumulated scouring sediment of Xiaolangdi Reservoir was 889.5 million m^3. Therefore, the evolution of sediment content in the lower Yellow River is greatly affected by human activities, and this sequence can be approximately regarded as a non-stationary sequence.

MEEMD decomposition

After repeated experiments, MEEMD has the best decomposition effect on annual sediment concentration series when the noise amplitude is 0.2, the logarithm of noise is 100, the embedding dimension is 5, the time delay is 1, the maximum decomposition number is 6, and the PE threshold θ_0 is 0.7.

The decomposition effect is shown in Figure 2.

As can be seen from Figure 2, the sediment concentration sequence is decomposed into three IMF components, one trend term and one pseudo component. Among them, the IMF_1 component has the largest fluctuation, the highest frequency and the shortest wavelength. However, the amplitude and frequency of other IMF components gradually decrease, and the wavelength gradually becomes larger. In addition, the evolution amplitude of the sub-sequence was significantly lower than that of the original sequence, indicating that the stability of the sediment concentration sequence was greatly increased after MEEMD treatment.

After the sediment concentration sequence is decomposed into sub-signals of different frequencies (IMF components), the prediction of complex sediment concentration sequence becomes the prediction of IMF components of different frequencies. Then, by calculating the relative error of IMF component, we can obtain the relative error of sediment concentration, and then analyze the contribution rate of the IMF component to the sediment concentration sequence.

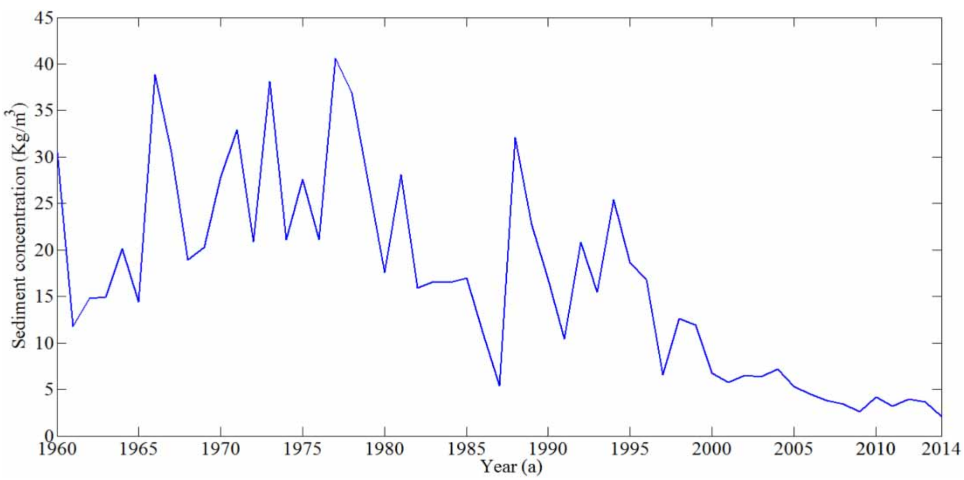

Figure 1 | Annual sediment concentration curve of Gaocun Station.

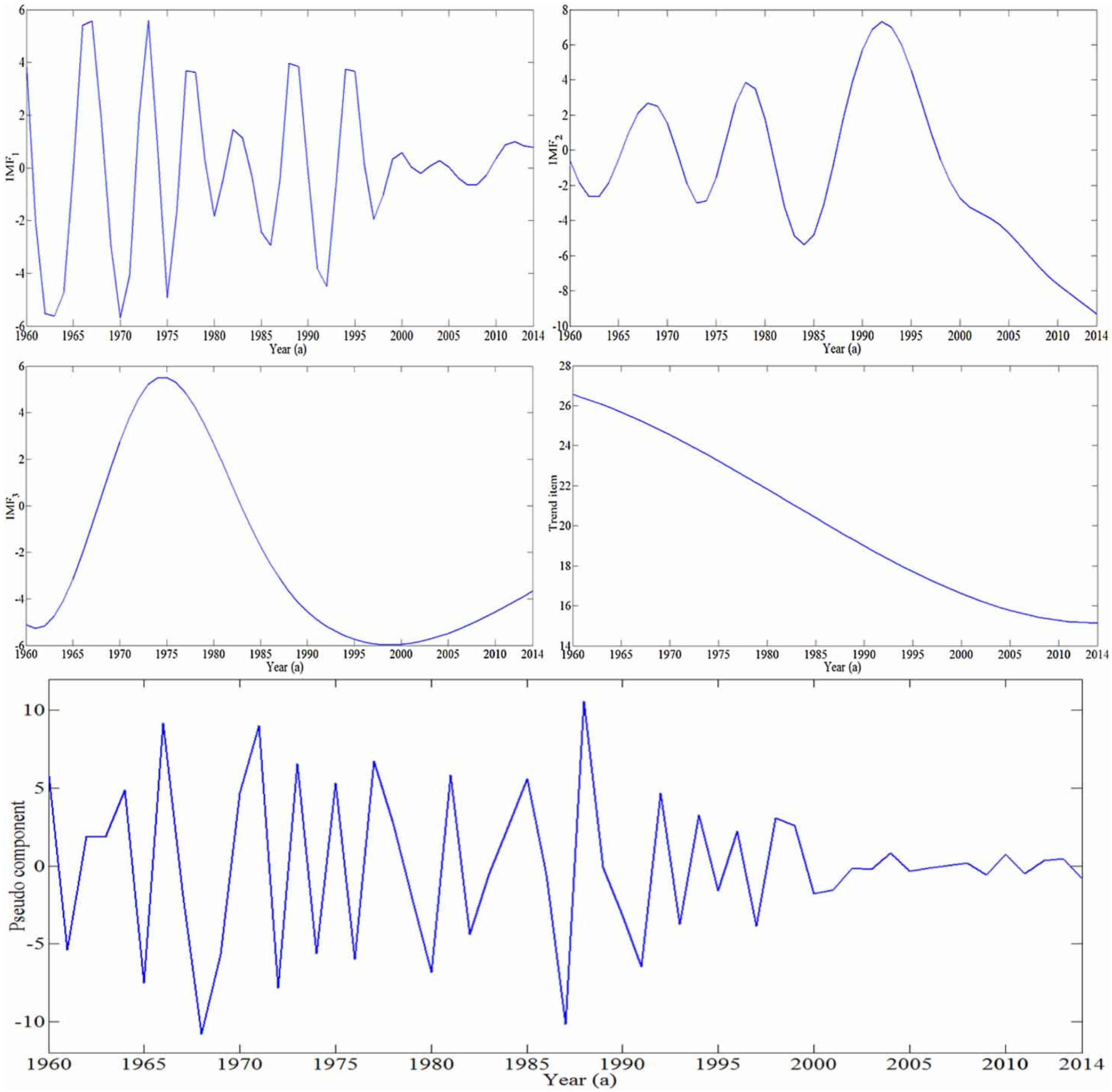

Figure 2 | MEEMD decomposition of annual sediment concentration in Gaocun Station.

ARIMA prediction

Because the IMF pseudo-component has no real physical meaning, so the MEEMD-ARIMA model only predicts three IMF components and one trend item. Then the predicted value of sediment concentration is equal to the sum of the predicted value of three IMF components, the predicted value of trend items and the pseudo-component.

The ARIMA model parameters and predicted values corresponding to IMF components are shown in Table 2.

It can be seen from Table 2 that the prediction error of the IMF_1 component is relatively high, and the prediction errors of the IMF_2, IMF_3 and trend terms are relatively small and tend to decrease gradually, and the trend term has the best prediction effect. Thus, it can be seen that the trend term has the strongest stationarity, while IMF_1 has

Table 2 | The predicted value of IMF component and corresponding ARIMA model parameters

IMF component	Time (year)	True value	Predictive value	Relative error (%)	ARIMA (p, d, q)	R^2
IMF_1	2010	0.35	0.29	−2.76	(6,0,2)	0.84
	2011	0.88	0.81	−7.80	(6,0,2)	0.85
	2012	1.00	0.83	−16.82	(6,0,2)	0.81
	2013	0.83	0.63	−23.85	(6,0,2)	0.85
	2014	0.77	0.58	−24.23	(6,0,2)	0.87
IMF_2	2010	−7.63	−7.51	−1.53	(3,0,2)	1.00
	2011	−8.06	−7.96	−1.20	(3,0,2)	1.00
	2012	−8.47	−8.38	−1.09	(3,0,2)	1.00
	2013	−8.89	−8.81	−0.93	(3,0,2)	1.00
	2014	−9.32	−9.25	−0.80	(3,0,2)	1.00
IMF_3	2010	−4.56	−4.56	0.08	(8,0,3)	0.99
	2011	−4.34	−4.34	−0.02	(8,0,3)	0.99
	2012	−4.12	−4.12	−0.01	(8,0,3)	0.98
	2013	−3.90	−3.90	0.10	(8,0,3)	0.99
	2014	−3.67	−3.67	−0.03	(8,0,3)	0.99
Trend item	2010	15.27	15.27	0.02	(13,0,1)	0.96
	2011	15.21	15.21	−0.01	(13,0,1)	0.96
	2012	15.17	15.17	−0.01	(13,0,1)	0.96
	2013	15.15	15.15	0.00	(13,0,1)	0.96
	2014	15.15	15.15	0.02	(13,0,1)	0.97

the weakest stationarity. Although the prediction error of IMF_1 is relatively high, it can be seen from Figure 2 that the IMF_1 in 2010–2014 accounts for a small proportion of sediment concentration, so it does not affect the prediction error of sediment concentration.

The prediction effect of the MEEMD-ARIMA model is shown in Figure 3.

The prediction error of sediment concentration in MEEMD-ARIMA model is shown in Table 3.

As can be seen from Figure 3 and Table 3, the MEEMD-ARIMA model has a better prediction effect and the relative error is $\leq \pm 6\%$.

Table 4 compares the relative errors of MEEMD-ARIMA model with CEEMD-ARIMA model, EEMD-ARIMA model, EMD-ARIMA model and ARIMA model in sediment concentration prediction.

According to Table 4, we can see that the MEEMD-ARIMA model has obvious accuracy advantage over the

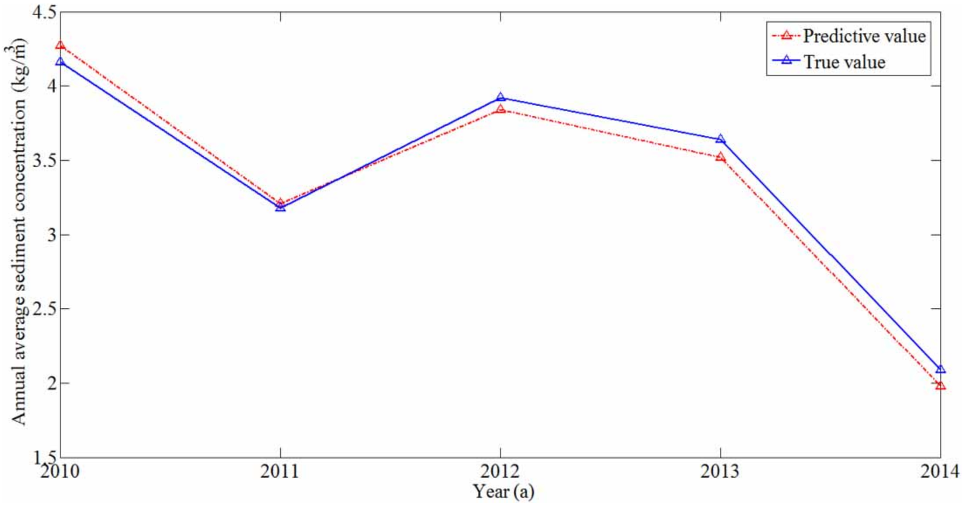

Figure 3 | Prediction effect of the annual sediment concentration of Gaocun Station from 2010 to 2014.

Table 3 | Prediction error of sediment concentration in Gaocun Station from 2010 to 2014

Time (year)	True value (kg/m^3)	Predictive value (kg/m^3)	Absolute error (kg/m^3)	Relative error (%)
2010	4.16	4.27	0.10	2.51
2011	3.18	3.21	0.03	0.95
2012	3.92	3.84	–0.07	–1.85
2013	3.64	3.52	–0.11	–3.16
2014	2.09	1.98	–0.11	–5.09

other three decomposition prediction models. The relative error of the MEEMD-ARIMA model is controlled within 6%, while the relative error of the other three prediction models is all over 10%, even up to 16.81% (ARIMA model). Thus, the MEEMD-ARIMA model still shows good prediction effect even for Gaocun Station with a complicated and wandering river section.

DISCUSSION

Sediment concentration evolution is a complex process, which is greatly affected by uncertain factors, such as water temperature, water depth, flow rate, rainfall, etc. The advantage of the MEEMD-ARIMA model is that the non-stationarity of sediment concentration sequence decreases after MEEMD decomposition, which provides a stable condition for the ARIMA model to predict. From the perspective of MEEMD decomposition, the contribution rates of IMF components and trend terms to sediment concentration are different. Trend item accounts for a large proportion, but because of its relatively high stationarity, the prediction error is relatively low, which has a greater impact on the prediction accuracy of sediment concentration, and the contribution rate is higher. However, although the IMF_1 component prediction error is slightly higher, the smaller the proportion, the smaller the impact on the prediction accuracy of the sediment concentration, and the smaller the contribution rate, so it does not affect the overall prediction error of the sediment concentration.

Table 4 | Comparison of prediction errors between MEEMD-ARIMA model and other models

	Relative error of:				
Time (year)	MEEMD-ARIMA model (%)	CEEMD-ARIMA model (%)	EEMD-ARIMA model (%)	EMD-ARIMA model (%)	ARIMA model (%)
2010	2.51	6.69	9.25	11.54	13.25
2011	0.95	5.54	8.93	10.17	13.66
2012	–1.85	2.56	–6.94	–7.87	–9.31
2013	–3.16	–7.51	–7.73	–10.44	–13.39
2014	–5.09	–10.63	–11.35	–12.61	–16.81

CONCLUSIONS

The fitting and prediction results of the annual sediment concentration of Gaocun Station by the constructed MEEMD-ARIMA model indicate the following:

1. The annual sediment concentration series at Gaocun Station are characterized by great fluctuation, and short-term prediction can be made by reducing the non-stationarity of the series by MEEMD.
2. Through the fitting and prediction of sediment concentration at Gaocun Station, it can be concluded that after MEEMD decomposition of sediment concentration sequence, the problem whereby the ARIMA model requires stable sequence can be solved, and the evolution characteristics of the sediment concentration sequence in various frequency domains can also be reflected. Therefore, it can be seen that the established MEEMD-ARIMA model has high accuracy in fitting and predicting the annual sediment concentration series at Gaocun Station.
3. Coupling the MEEMD method with the ARIMA model is a new attempt, and its representativeness and model accuracy are good through the research results. In addition, the MEEMD-ARIMA model can be applied to the prediction of runoff, groundwater, and hydrometeorological factors, and has broad application prospects.
4. This paper does not consider the long-term prediction of sediment concentration and the physical mechanism of sediment concentration evolution. In addition, how to improve the accuracy of the model under the strong effect of human activities will be the next research direction and focus.

ACKNOWLEDGEMENTS

This work was financially supported by the Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Henan Province, Water Environment Governance and Ecological Restoration Academician Workstation of Henan Province, National Natural Science Foundation of China (NSFC) (No. U1304511), Henan International Science and Technology Cooperation Project (No. 152102410052). Our gratitude is also extended to the reviewers for their meticulous work and thoughtful suggestions during the review process, which were of great help in improving the manuscript.

REFERENCES

Bandt, C. & Pompe, B. 2002 Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **88** (17), 174102.

Barua, S., Perera, B. J. C. & Ng, A. W. M. 2010 Drought forecasting using an aggregated drought index and artificial neural network. *J. Water Clim. Change* **1** (3), 193–206.

Borgnat, P., Flandrin, P., Honeine, P., Richard, C. & Xiao, J. 2010 Testing stationarity with surrogates: a time-frequency approach. *IEEE Trans. Signal Process* **58** (7), 3459–3470.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. 2015 Time series analysis: forecasting and control, 5th edition. *J. Oper. Res. Soc.* **22** (2), 199–201.

Chen, J. Z. 2005 Application of artificial neural network BP model to predict annual average sediment concentration in Wujiang River Basin. *J. China Hydrol.* **25** (4), 6–9 (in Chinese).

Chen, H. & Dyke, P. P. G. 1998 Multivariate time series model for suspended sediment concentration. *Cont. Shelf Res.* **18** (2), 123–150.

Chen, H., Zhu, L., Wang, J. & Fan, H. X. 2017 Modal analysis of annual runoff volume and sediment load in the Yangtze river-lake system for the period 1956–2013. *Water Sci. Technol.* **76** (1), 1–14.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, E. H., Zheng, Q., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Lond.* **454** (1971), 903–995.

Itai, D. & Amit, H. 2018 Modern statistical tools for inference and prediction of infectious diseases using mathematical models. *Stat. Methods Med. Res.* **27** (7), 1927–1929.

Joshi, R., Kumar, K. & Adhikari, V. P. S. 2015 Modelling suspended sediment concentration using artificial neural networks for Gangotri glacier. *Hydrol. Process* **30** (9), 1354–1366.

Li, Y. T. & Shang, Q. M. 1998 Study on mathematical model of one-dimensional unsteady flow sediment. *J. Sediment Res.* **1**, 83–89 (in Chinese).

Miao, Y., Wang, W. C. & Wu, H. B. 2012 Prediction of sediment concentration in Lanzhou Station of the Yellow River based on PSO-BP model. *Yellow River* **34** (09), 35–40 (in Chinese).

Myronidis, D., Ioannou, K., Fotakis, D. & Gerald, D. 2018 Streamflow and hydrological drought trend analysis and forecasting in Cyprus. *Water Resour. Manage.* **2018** (1), 1–18.

Riedl, M., Müller, A. & Wessel, N. 2013 Practical considerations of permutation entropy. *Eur. Phys. J. Spec. Top.* **222** (2), 249–262.

Shi, W. & Wang, G. Q. 2003 A review of research on sediment transport in the lower Yellow River. *Adv. Water Sci.* **14** (1), 118–123 (in Chinese).

Shi, B., Qing, Y. & Ling, Y. 2008 The application of response function model in sediment prediction. *Bulletin of Soil and Water Conservation* **28** (2), 104–105 (in Chinese). doi: https://doi.org/10.13961/j.cnki.stbctb.2008.02.008.

Terrien, J., Marque, C. & Karlsson, B. 2011 Automatic detection of mode mixing in empirical mode decomposition using nonstationarity detection: application to selecting IMFs of interest and denoising. *EURASIP J. Adv. Signal Process.* **2011** (1), 37.

Tilahun, S. A., Guzman, C. D. & Zegeye, A. D. 2015 Distributed discharge and sediment concentration predictions in the sub-humid Ethiopian highlands: the Debre Ma-wi watershed. *Hydrol. Process* **29** (7), 1817–1828.

Vigiak, O., Malagó, A., Bouraoui, F., Vanmaercke, M. & Poesen, J. 2015 Adapting SW-AT hillslope erosion model to predict sediment concentrations and yields in large basins. *Sci. Total Environ.* **538**, 855–875.

Wu, Z. H. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assis-ted data analysis method. *Adv. Adapt. Data Anal.* **1** (1), 1–41.

Xu, J. X. 2004 Study on the macroscopical trend of sediment deposition in the lower reaches of the Yellow River under the influence of human activities. *J. Hydraul. Eng.* **35** (2), 8–16 (in Chinese).

Yan, R. & Gao, R. X. 2007 Approximate entropy as a diagnostic tool for machine health monitoring. *Mech. Syst. Signal Process.* **21**, 824–839.

Yao, P., Su, M., Wang, Z. B., van Rijn, L. C., Zhang, C. K., Chen, Y. P. & Stive, M. J. F. 2015 Experiment inspired numerical modeling of sediment concentration over sand-silt mixtures. *Coast. Eng.* **105**, 75–89.

Yeh, J. R. & Shieh, J. S. 2010 Complementary ensemble empirical mode decomposition: a noise enhanced data analysis method. *Adv. Adapt. Data Anal.* **2** (2), 135–156.

First received 17 April 2019; accepted in revised form 5 September 2019. Available online 1 November 2019