

RESEARCH ARTICLE

Trend detection and stochastic simulation prediction of streamflow at Yingluoxia hydrological station, Heihe River Basin, China

Chenglong ZHANG, Mo LI, Ping GUO (✉)

Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China

Abstract Investigating long-term variation and prediction of streamflow are critical to regional water resource management and planning. Under the continuous influence of climate change and human activity, the trends of hydrologic time series are nonstationary, and consequently the established methods for hydrological frequency analysis are no longer applicable. Five methods, including the linear regression, nonlinear regression, change point analysis, wavelet analysis and Hilbert-Huang transformation, were first selected to detect and identify the deterministic and stochastic components of streamflow. The results indicated there was a significant long-term increasing trend. To test the applicability of these five methods, a comprehensive weighted index was then used to assess their performance. This index showed that the linear regression was the best method. Secondly, using the normality test for stochastic components separated by the linear regression method, a normal distribution requirement was satisfied. Next, the Monte Carlo stochastic simulation technique was used to simulate these stochastic components with normal distribution, and thus a new ensemble hydrological time series was obtained by combining the corresponding deterministic components. Finally, according to these outcomes, the streamflow at different frequencies in 2020 was predicted.

Keywords Monte Carlo, nonstationary, trend detection, streamflow prediction, decomposition and ensemble, Yingluoxia

1 Introduction

A streamflow time series (STS) is a comprehensive outcome under the influence of the climatological and geographical conditions and relevant human activities^[1]. Generally, these data should accord with the principles of representation, reliability and consistency in long-term hydrologic frequency analysis. However, streamflow is increasingly affected by climate change and human activity. For example, climate change is an exogenous process associated with precipitation and temperature^[2], and human activity includes hydraulic engineering, cropping and irrigation water withdrawals, which are the endogenous factors to variation in the STS. Therefore, STS cannot usually meet the requirements of being consistent/stationary, so it is defined as the nonstationary STS (NSTS). The NSTS is composed of the deterministic components and stochastic components of streamflow. Deterministic components are the trends of time series over a specific period of time, while the stochastic components are usually stable and satisfy specific random probability distributions (such as, Pearson type III distribution, normal distribution, log-normal distribution, uniform distribution and extreme value distribution). Identification and detection of these components in the NSTS are important parts for the study of the formation of and influences on hydrologic time series, and they are also the foundation of streamflow stochastic simulation, which underpins streamflow prediction.

For trend detection and analysis, statistics methods, which will help to better investigate the changes of a long-term streamflow, are widely used^[3–9]. There are more than 20 methods used for analysis of trends and abrupt changes^[10]. Parametric regression methods like linear regression and nonlinear regression are used to quantify trends^[11,12]. Nonparametric methods, such as kernel, spline, and rank correlation methods, are used for the

Received June 12, 2016; accepted September 18, 2016

Correspondence: guop@cau.edu.cn

NSTS trend analysis, wavelet analysis^[13] and the Hilbert-Huang transform^[14]. The Mann-Kendall trend test is used for shift trends^[13,15], and Wilcoxon-Mann-Whitney test^[16] and Pettitt test for abrupt changes^[7,17]. These methods have made significant contributions to information for hydrologic time series trend detection and analysis. However, these methods cannot identify and detect trends and abrupt change points simultaneously. In other words, these methods focus on either trend analysis or change point detection. Additionally, streamflow trends are also characterized by statistical parameters, such as mean value or variance. These test results may be inconsistent and unsystematic, and even lead to different understanding of variation in the NSTS because of their specific conditions and limits^[18]. Therefore, it is difficult to effectively detect trends and analyze streamflow using only one or two of these methods.

A better understanding of the NSTS can assist in forecasting the future inflow levels and support regional water resources management and planning. However, current hydrological frequency analysis methods with the assumption of same distributions are no longer applicable. After trend analysis and detection of the NSTS, deterministic and stochastic components of streamflow can be decomposed into the constituent parts. To derive the frequency distribution of streamflow in the past, present and future, stochastic simulation techniques can be adopted. The Monte Carlo stochastic simulation technique (MCST)^[19], which enables simulation of the actual situation and investigation of the intrinsic principles by generating numerous random numbers, can be applied to simulate and generate sampling points of stochastic components. When using the MCST, sufficient time is needed to repeatedly generate random numbers for simulating stochastic components^[1]. Then, an ensemble hydrological time series for hydrological frequency calculation can be derived by adding the corresponding deterministic components.

Therefore, this study was conducted in three major sections. Initially, the preliminary test of significant trend analysis was conducted to investigate change/variation in the long term NSTS using a correlation coefficient test, the Spearman correlation analysis test and the Kendall correlation test. The nonstationary nature of trends in annual streamflow were shown. Secondly, five different methods for trend identification and detection of deterministic and stochastic components of the NSTS motivated by the concept of decomposition and ensemble were selected and their corresponding results and performance were compared and evaluated to further improve their detection accuracy. Finally, the MCST was applied to simulate the stochastic components and thus the frequency distribution of the NSTS in 2020 was derived by integrating with the corresponding deterministic components. Therefore, the annual streamflow frequency was

calculated to obtain the predicted streamflow at different frequencies in 2020. These predicted results provide a scientific basis for the water resources management in the Heihe River Basin, China. This study was organized as shown in Fig. 1.

2 Study area

The Heihe River Basin is the second largest inland river basin in Northwest China^[20]. The basin covers an area of 11.6×10^4 km² and the mainstream of the river is 821 km long. The river originates from the Qilian Mountains and flows through the Hexi corridor of Gansu Province and finally enters into the western part of the Inner Mongolia Plateau. It is divided into three segments including upper, middle and lower reaches which are controlled by Yingluoxia and Zhengyixia hydrological stations. The Yingluoxia station is at the outlet of the Heihe River from the upstream mountainous area^[21] and the mid and downstream areas of the Heihe River Basin are the main areas of water consumption. Especially, the midstream region, which is one of the major irrigation and grain/cereal crop production areas of China. Agriculture is the dominant activity, with 84% of the total available water being used for agricultural irrigation. Moreover, with the population growth and eco-social development, mismatch between water supply and demand has seriously restricted the sustainability of the local economy, society and ecology. Therefore, to support sustainable development, it is crucial to investigate the changes of the long-term annual streamflow in the Heihe River Basin. The objectives of this study were to (1) examine the annual streamflow at Yingluoxia hydrological station from 1945 to 2013, (2) reveal the underlying mechanisms of change in streamflow, and (3) predict future streamflow. Figure 2 shows the study area within the Heihe River Basin. Figure 3 shows the mean annual streamflow at Yingluoxia hydrological station from 1945 to 2013.

3 Methodology

3.1 Preliminary data test

Before studying the variation in the long-term streamflow, related statistical trend analysis was conducted to detect the overall trend in streamflow. Furthermore, the trend analysis can identify whether there were increasing or decreasing changes over the historical period and whether the change was gradual. To investigate the potential overall streamflow changes, three methods for trend analysis were undertaken: the correlation coefficient test, the Spearman correlation analysis test and the Kendall correlation test.

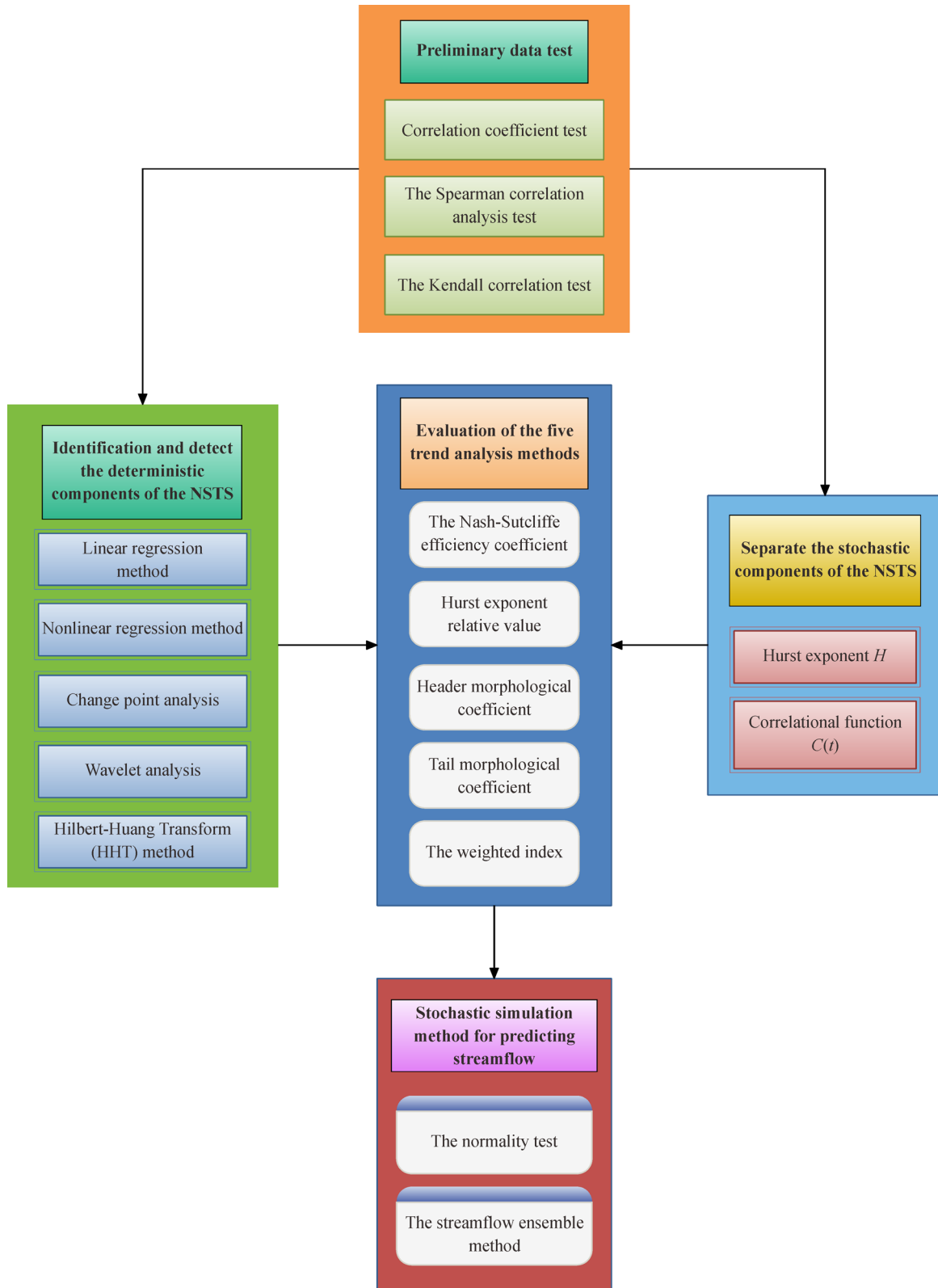


Fig. 1 The framework used for the study of streamflow at Yingluoxia hydrological station, Heihe River Basin, China

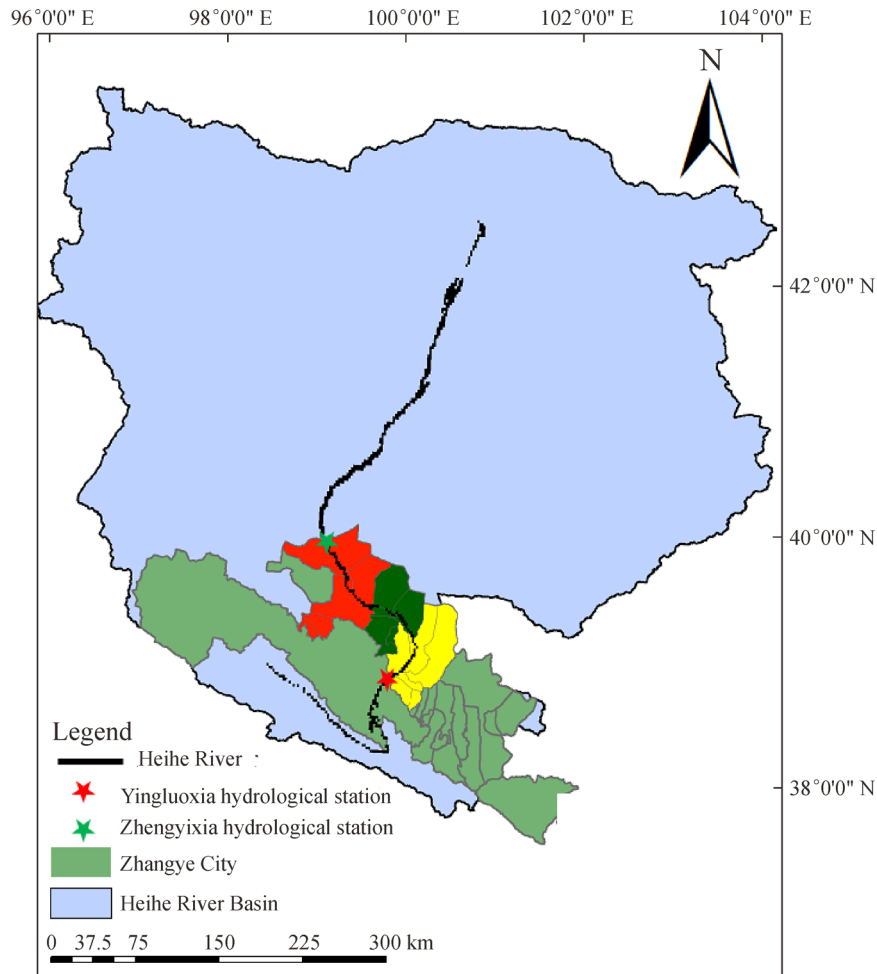


Fig. 2 The Heihe River Basin showing the location of its two hydrological stations. The red, dark green and yellow areas in this map are Gaotai County, Linze County and Ganzhou District in the midstream of Heihe River Basin, respectively.

3.1.1 Correlation coefficient test

If there is a linear monotonic trend in the NSTS, then it can be described by linear regression equation. Therefore, the test was conducted by calculating the correlation coefficient (r) from the corresponding time series. The equation was formulated as follows:

$$r = \frac{\sum_{t=1}^n (x_t - \bar{x})(t - \bar{t})}{\sqrt{\sum_{t=1}^n (x_t - \bar{x})^2 \sum_{t=1}^n (t - \bar{t})^2}} \quad (1)$$

where $\bar{t} = \frac{1}{n} \sum_{t=1}^n t = \frac{n+1}{2}$; $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$, x_t is the time series, t is time, n is the number of time series.

In this study, standard normal distribution was employed to test whether the time series data can be accepted as significant changing. When the absolute values of correla-

tion coefficient are no larger than the typical critical value (i.e., $1.645/\sqrt{n}$, $1.96/\sqrt{n}$ and $2.576/\sqrt{n}$ under the 10%, 5% and 1% significance level, respectively), the NSTS data was accepted^[1].

3.1.2 The Spearman correlation analysis test

The Spearman rank correlation coefficient (r) was formulated as follows:

$$r = 1 - \frac{6 \sum_{t=1}^n d_t^2}{n^3 - n} \quad (2)$$

where n is the number of time series, $d_t = R_t - t$, R_t is the corresponding rank when the numbers in ascending order. The t test method was used to test the statistical variable:

$T = r \left(\frac{n-4}{1-r^2} \right)^{\frac{1}{2}}$. When the typical critical value $|T| \geq t_{\alpha/2}$

($\alpha = 0.01, 0.05, 0.1$), the correlation relationship between the NSTS and time was explained, and also indicated there was long-term significant trend.

3.1.3 The Kendall correlation test

Similar to the Spearman correlation analysis test, the Kendall rank correlation coefficient (r) was formulated as follows:

$$u = \frac{\lambda}{[\text{Var}(\lambda)]^{1/2}} \quad (3)$$

where $\lambda = \frac{4m}{n(n-1)} - 1$ and $\text{Var}(\lambda) = \frac{2(2n+5)}{9n(n-1)}$, m is the number of all dual values when $x_i < x_j (x_i, x_j, j > i)$. When the number of time series was increasing, the test statistics follows a standard normal distribution. When the typical critical value $|U| \geq U_{\alpha/2} (\alpha = 0.01, 0.05, 0.1)$, this indicates there was long-term significant trend.

3.2 Identification and detection of the deterministic components

Hydrological frequency calculation require conditions of the hydrologic time series to be consistent/stationary. If there are increasing or decreasing changes during the period of records and the change is gradual or abrupt, it does not meet the condition of consistency. Therefore, it is necessary to identify, test and describe these components (i.e., deterministic components, period terms and stochastic components), and to separate them from the NSTS. Many methods have been applied to identify and detect the

trends. From these methods, the linear regression, non-linear regression, change point analysis, wavelet analysis and Hilbert-Huang transform (HHT) with empirical mode decomposition (EMD) were chosen to identify and detect the deterministic components of the NSTS in this study.

3.2.1 Linear regression method

Generally speaking, according to the results from the preliminary data test, if the NSTS has an overall increasing or decreasing changes in the period of records, this trend can be described as either a linear or nonlinear trend. However, from Fig. 3, it is not evident whether the streamflow trend at Yingluoxia is of a linear monotonic or nonlinear form. In this study, the NSTS was first fitted by linear regression. The mathematical model of linear regression and the estimated values of parameters \hat{a}, \hat{b} using the least square method was as follows:

$$X(t) = a + bt + \eta_t, t = 1, 2, \dots, n \quad (4)$$

$$\hat{b} = \frac{\sum_{t=1}^n (t-\bar{t})(x_t - \bar{x})}{\sum_{t=1}^n (t-\bar{t})^2} \quad (5)$$

$$\hat{a} = \bar{x} - \hat{b}\bar{t} \quad (6)$$

where $\bar{t} = \frac{1}{n} \sum_{t=1}^n t = \frac{n+1}{2}$; $\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t$. Thus, the linear regression equation $X(t)$ was obtained.

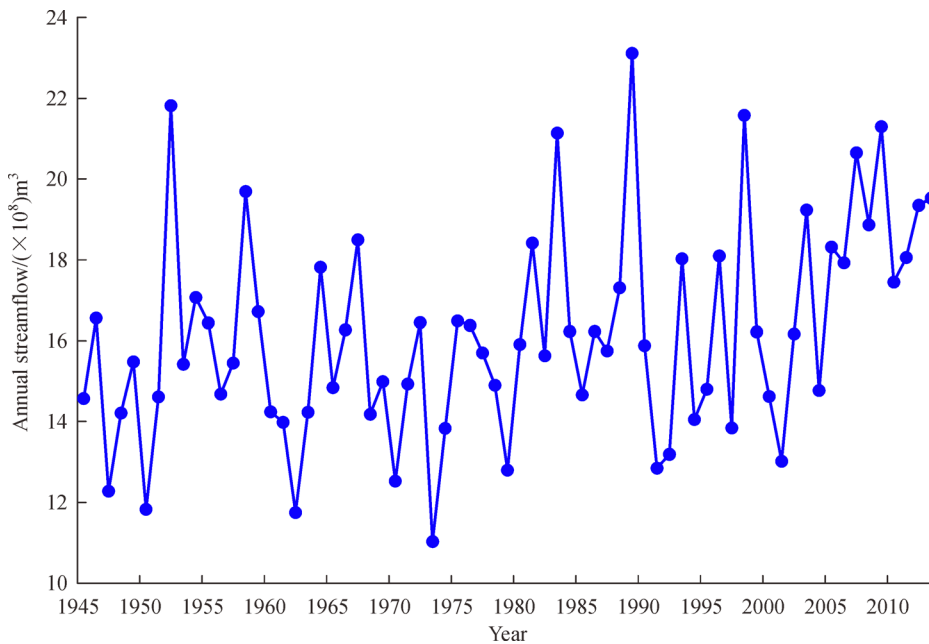


Fig. 3 Annual streamflow at Yingluoxia hydrological station from 1945 to 2013

3.2.2 Nonlinear regression method

Besides linear regression, nonlinear regression was also used in the streamflow trend analysis for Yingluoxia. It was not clear which kind of nonlinear regression should have been adopted, so the streamflow at Yingluoxia was fitted with the exponential, logarithmic and polynomial functional curves. By comparing the correlation coefficients of these nonlinear regression methods, the highest correlation coefficient was taken as the nonlinear fit (Table 1). It was found that exponential fits were the best, and thus the deterministic components trend equation was obtained.

Table 1 Fits for different functional forms to the NSTS at Yingluoxia hydrological station

Functional form	Fitting formula	$R^2/\%$
Exponential	$y = 14.443 e^{0.0028x}$	16.76
Logarithmic	$y = 0.8148 \ln(x) + 13.847$	8.08
Polynomial	$y = 0.0014x^2 - 0.0527x + 15.699$	12.91

3.2.3 Change point analysis

In this study, trend analysis of annual streamflow was undertaken to examine changes during the period for which records were available. The Mann-Kendall method is commonly used to detect abrupt changes in a trend^[10]. The assumption of normality is not required but serial correlation should be used for correcting the P -values^[22,23]. The procedure for this method is repeated for a time series that includes all possible periods of at least 10 years of records^[23]. The P -values (significance level of 5%) of the trend tests was adopted here. Abrupt change point detection was conducted to determine potential change points (joint points). Therefore, the NSTS was divided into several segments according to the potential change points. Finally, based on the mean values of annual streamflow in different periods before and after the occurrence of abrupt change points, the corresponding deterministic components were obtained.

3.2.4 Wavelet analysis

Wavelet analysis is capable of signal representation both in time and frequency domain, especially for the study of nonstationary time series like hydrological processes (i.e., precipitation and streamflow). Wavelet analysis is a time-frequency analytical method to assess the time scales of signals, and it has the powerful characteristic of multi-resolution analysis. Thus, deterministic and stochastic components at various time scales can be identified and decomposed by the wavelet analysis. When multi-resolution decomposition is performed, with the increase in the level of wavelet decomposition, the time resolution is

reduced and the frequency resolution becomes higher. Therefore, the final residual term of the low frequency component is the trend component, which can be regarded as a deterministic component^[24]. In this study, a Daubechies orthogonal wavelet was applied to decompose the annual streamflow from 1945 to 2013, and the five levels of wavelet decomposition were used to obtain d1–d5 of the high frequency components and the low frequency components of a5^[24]. Finally, the appropriate functional forms was used to fit the trend components, such as linear function, power function, logarithmic function, exponential function or polynomial function. The corresponding deterministic components were obtained.

3.2.5 Hilbert-Huang transform

Empirical mode decomposition (EMD) of the HHT was proposed by Huang et al.^[25]. Generally speaking, HHT has a similar role in wavelet analysis, but the HHT is empirical while wavelet analysis has a theoretical mathematical basis. The HHT inherits the characteristics of multi-resolution wavelet transform analysis, but it is also an adaptive method used for signal analysis, which has great advantages in the analysis of nonlinear and nonstationary time series^[26]. The HHT can obtain the physical meaning associated with the full nonlinear/nonstationary system through individual components in the linear system rather than a physically linear expansion only^[27]. Thus, the NSTS can be decomposed by EMD into a series of components including several independent intrinsic mode functions (IMF) at different time scales and a residual component, which represents the overall trend of the original time series^[26]. Finally, appropriate functional forms can be chosen to fit the deterministic trend components and the corresponding deterministic components were obtained.

3.3 Separating the stochastic components

The NSTS is generally composed of two or more components, for example, items of tendency (deterministic components), period and stochastic components. In this study, the stochastic component S_t was expressed as $S_t = Y_t - X_t$ (where Y_t is the NSTS) based on the assumption that each component of the NSTS satisfies a linear superposition principle. Therefore, after deducting the deterministic components from the NSTS, the remaining part was considered as a pure stochastic component. It is necessary to diagnose and test the consistency condition for the stochastic component of the hydrological analysis^[24]. Therefore, Hurst exponent H and the critical values of different significant levels of correlational function $C(t)$ were used to verify consistency conditions^[28]. The relationship between H and $C(t)$ was as follows:

$$C(t) = 2^{2H-1} - 1 \quad (7)$$

According to the principle of R/S analysis, the H exponent of the stochastic component was calculated first, then the $C(t)$ value was calculated. At a given significance level, the significance of the $C(t)$ value was tested. If it was less than the critical value, the long-term change in the stochastic components was considered to be not significant. It was thus shown that the decomposition was reasonable to satisfy the consistency condition. Therefore, the original NSTS was decomposed into the deterministic components (tendency) that changed significantly and the pure stochastic components.

3.4 Evaluation of the five trend analysis methods

The above five methods can identify and used to test the deterministic and stochastic components of the NSTS, but it was also useful to quantitatively distinguish and compare their applicability, advantages and disadvantages for the streamflow at Yingluoxia hydrological station. Moreover, by comparison, the most suitable method was obtained and thus the deficiencies in the method used was reduced to prevent confusion and repetitive work. The criteria for selecting a method should not just be based on goodness-of-fit testing but should also consider physical reasoning. Therefore, the evaluation indexes were determined, including (1) the Nash-Sutcliffe efficiency coefficient between the deterministic components and the original NSTS, (2) the relative value of the stochastic components of the Hurst exponent, (3) the header morphological coefficient of deterministic components for the original NSTS, and (4) the tail morphological coefficient of deterministic components for the original NSTS. The same weight coefficient for each index are then adopted in the process of evaluation, thus the comprehensive weighted indexes of five different methods were calculated, and finally the method with the maximum value of the comprehensive weighted index was regarded as the ideal method. Four evaluation indexes were selected as described below.

3.4.1 The Nash-Sutcliffe efficiency coefficient

The Nash-Sutcliffe efficiency coefficient^[29] can be used to evaluate the fit of the measured annual streamflow time series and the deterministic trend component. The formula is:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Q_{mea,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{mea,i} - \bar{Q}_{mea})^2} \quad (8)$$

where $Q_{mea,i}$ ($i = 1, 2, \dots, n$) is the measured annual streamflow time series; \bar{Q}_{mea} is the mean value of the time series; $Q_{sim,i}$ is the value of each point on the deterministic trend.

3.4.2 Hurst exponent relative value

The Hurst exponent is capable of predicting future trends according to past trends for a time series, and it has been extensively used to predict hydrological and climatological processes^[30]. Because non-stationary long-term correlated time series cannot exhibit persistent behavior^[28], the degree of consistency of time series can be expressed by the Hurst exponent. If the Hurst exponent is close to 0.5, the consistency of the stochastic component is better, and the accuracy of the corresponding method is higher. Therefore, the distance difference between the Hurst exponent and 0.5 becomes criterion. It can be expressed as Hurst exponent relative value $H_t = 1 - 2ABS(H - 0.5)$, which is in the range 0–1.

3.4.3 Header morphological coefficient

In the view of a time series, actually, the overall fitting degree between deterministic components and the original NSTS has little difference when calculated by different methods, but there is a distinct difference in the header or tail position. Therefore, the distance to the header position is selected as a rule to determine the value of the morphological coefficient. To make all the indicators are normalized to the range 0–1, therefore, the header morphological coefficient that the header position in the middle position of the five methods was 1. Then, other header morphological coefficients were determined according to the distance between them and the middle position. If the degree of closeness is higher, the value is larger. The header morphological coefficients of the corresponding five methods are: 0.2, 0.4, 0.6, 0.8 and 1.

3.4.4 Tail morphological coefficient

Similar to the header morphological coefficient, the tail morphological coefficient is determined by the distance from the tail position. The tail position in the middle position of the five methods was 1. Then, other tail morphological coefficients were determined according to the distance between them and the middle position. If the degree of closeness is higher, the value is larger. The tail morphological coefficients of the corresponding five methods are: 0.2, 0.4, 0.6, 0.8 and 1.

3.4.5 The comprehensive weighted index

The comprehensive weighted index (W) is calculated as follows:

$$W = \alpha_1 A + \alpha_2 B + \alpha_3 C + \alpha_4 D \quad (9)$$

where α_i , $i = 1, 2, 3, 4$ is the weight of each evaluation index, and $\sum_{i=1}^4 \alpha_i = 1$. In this study, $\alpha_i = 0.25$. If the comprehensive weighted index is the largest among five methods, then it indicates that the method is more applicable to the study area.

3.5 Stochastic simulation method for predicting streamflow

3.5.1 The normality test

Before the stochastic simulation of streamflow, a statistical test is employed to examine whether the stochastic component is fine-modeled by a specific distribution (i.e., normal distribution and Pearson type III distribution). Many normality tests are widely used in statistical analysis^[31]. Among these tests, the Shapiro-Wilk test^[32] was used in this study to verify the normality of stochastic components separated from the NSTS. This test generates two values: W and P . The value of W is in the range 0–1, and a higher W value means that acceptance of normality while a lower W value leads to rejection. If W is equal to 1, it means that normality is totally satisfied. If the P -value is higher than the critical significance level, normality will be accepted^[33].

3.5.2 The streamflow ensemble method

Motivated by the concept of decomposition and ensemble^[34], the frequency distribution ensemble method is adopted for the nonstationary streamflow frequency calculation. The deterministic component of a certain time (moment) is first predicted according to the trend of the deterministic component. Then the MCST^[35] is used to generate a set of samples of a pure stochastic series which satisfies the condition of randomness. Thus, the final forecasted streamflow time series can be obtained by integrating the predicted deterministic component into the stochastic component. It can deduce the ensemble hydrological distribution and its hydrological parameters. The cumulative probability density function curve can also be obtained based on the numerous simulated samples at each frequency.

4 Results and discussion

4.1 Preliminary data analysis

Table 2 shows the results of the preliminary data analysis. It was rejected by the original hypothesis at the significance level 0.1, 0.05 and 0.01 because the value of statistic index were larger than the typical critical values at three significance levels. These tests showed that the NSTS is increasing significantly. In summary, the test results showed the trend in the NSTS was significant, which also demonstrated it can be identified, detected and decomposed into deterministic and stochastic components.

4.2 Components decomposition

The results of the trend analysis using the five methods are presented in the following sections. From the linear regression method, the estimated values of parameters \hat{a}, \hat{b} can be obtained by using the least squares method. $\hat{b} = 0.0468$, $\hat{a} = 14.522$, $r^2 = 0.13$. Nonlinear regression method was fitted with exponent functional form, and the parameters were obtained. As for the change point analysis, the Mann-Kendall test (Fig. 4) showed the intersection point of Z1 and Z2 curves in the annual streamflow time series occurred during 2003 and 2004. Z1 and Z2 are the statistical variables that follow the standard normal distribution. They were computed based on the original and adverse course, respectively. This result indicates the streamflow changes in 2003, showing that a notable fluctuation in streamflow had occurred. Thus, the NSTS can be divided into two sub periods: 1945–2003 and 2004–2013. The mean value of the first sub period was $15.7 \times 10^8 \text{ m}^3$ and the second sub period was $19.1 \times 10^8 \text{ m}^3$, with a difference of $3.4 \times 10^8 \text{ m}^3$. Wavelet analysis decomposition (Fig. 5) and deterministic trend fitting (Fig. 6) were conducted. The results for the EMD method and deterministic trend fitting are presented as Fig. 7 and Fig. 8.

If the NSTS showed no trend, the hydrologic time series was a straight line that passes through the first point ($t = 1945$). Its equation was $X(t) = X(t_1)$, which can reflect the streamflow situation before the variations of annual streamflow occurred. As a result, the deterministic components of annual streamflow was the difference in the trend before and after the variations ($X(t) - X(t_0)$). The

Table 2 The results of the preliminary data tests of streamflow at Yingluoxia hydrological station

Test	Typical critical value			Statistical index	Test result
	0.1	0.05	0.01		
Correlation coefficient test	0.20	0.23	0.30	0.36	Significant
The Spearman correlation analysis test	1.668	1.996	2.651	3.111	Significant
The Kendall correlation test	1.64	1.96	2.58	2.98	Significant

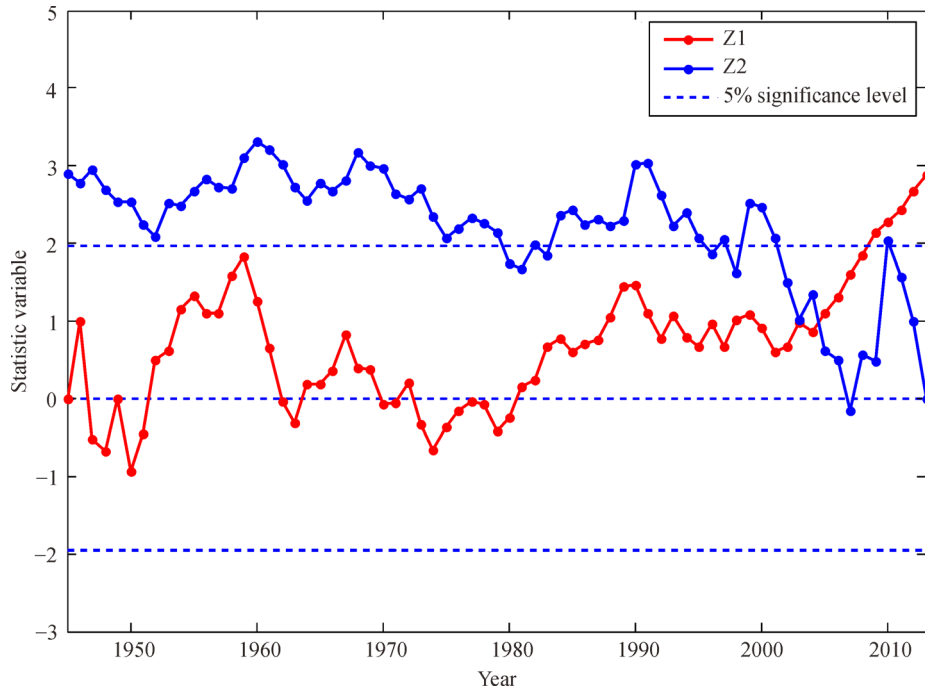


Fig. 4 Mann-Kendall trend test for the annual streamflow at Yingluoxia hydrological station from 1945 to 2013. Z1 and Z2 are the statistical variables that follow the standard normal distribution. They can be computed based on the original and adverse course.

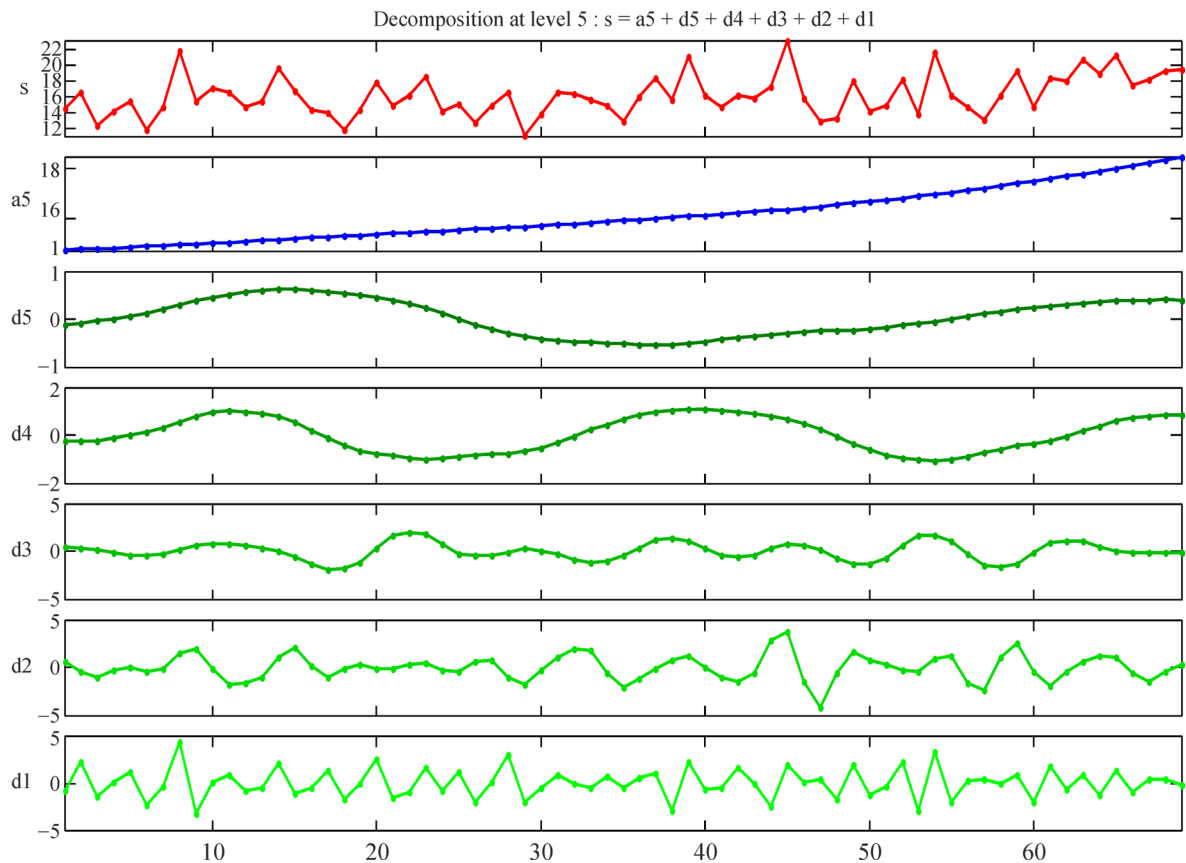


Fig. 5 Wavelet analysis decomposition for the NSTS at Yingluoxia hydrological station. d1–d5, the high frequency components at different scales; a5, low frequency deterministic components.

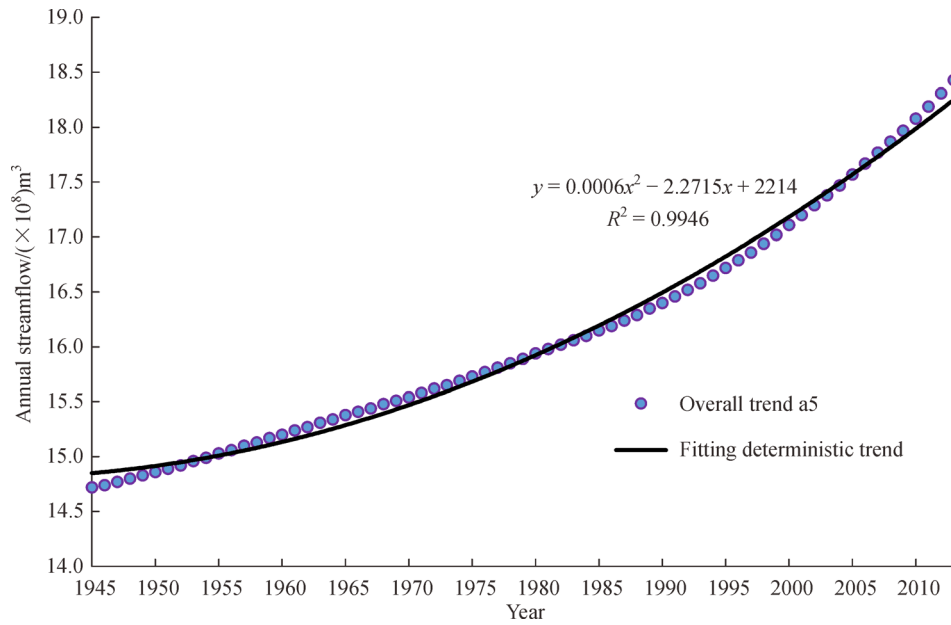


Fig. 6 Deterministic trend fitting based on wavelet analysis for streamflow at Yingluoxia hydrological station

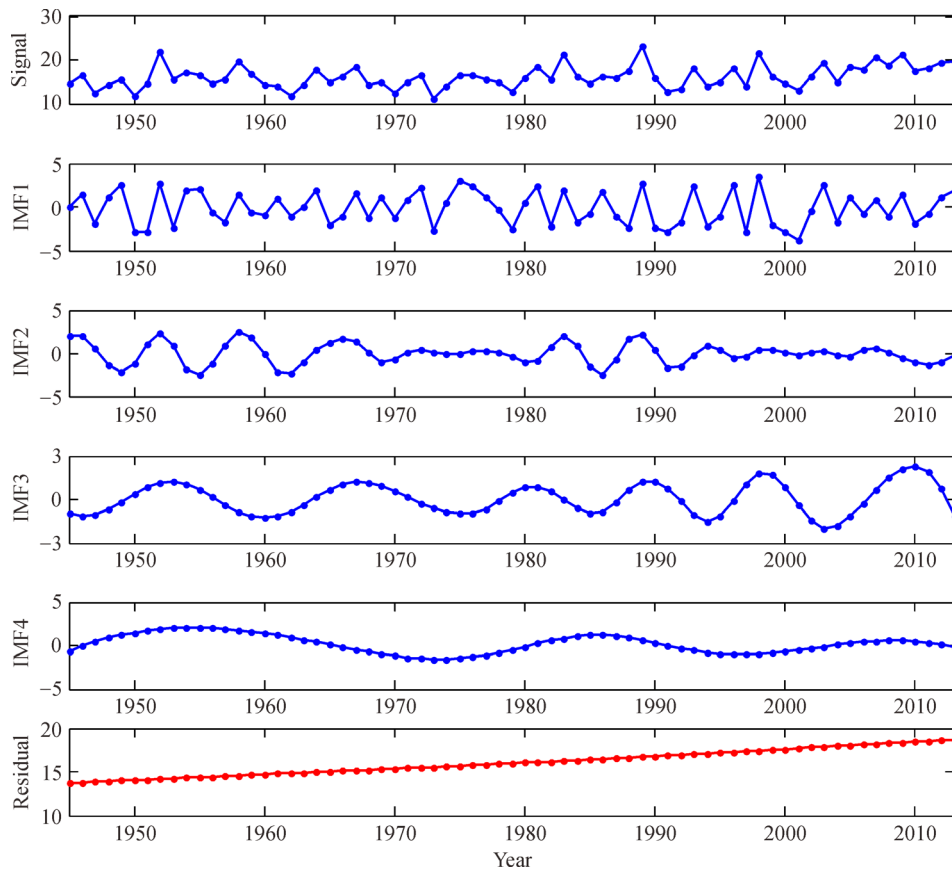


Fig. 7 Empirical mode decomposition for the NSTS at Yingluoxia hydrological station. IMF1–IMF4, the four independent intrinsic mode functions components in the order from the highest frequency to the lowest frequency, and one residual component, respectively.

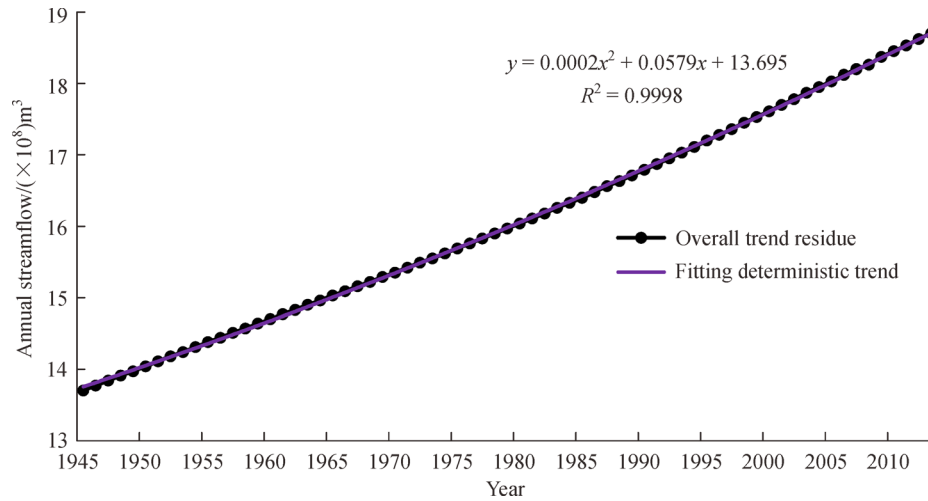


Fig. 8 Deterministic trend fitting based on Hilbert-Huang transform method for streamflow at Yingluoxia hydrological station

deterministic and stochastic components of five methods are shown in Table 3. These results provided the foundation for subsequent hydrological stochastic simulation and prediction in the future.

4.3 Evaluation of the methods

Figure 9 shows the header and tail position for the five methods. Therefore, the order of header morphological coefficients of the five methods was linear regression (1) > nonlinear regression (0.8) > wavelet analysis (0.6) > HHT (0.4) > change point analysis (0.2). The order of the tail morphological coefficients of the five methods were wavelet analysis (1) > HHT (0.8) > linear regression (0.6) > change point analysis (0.4) > nonlinear regression (0.2). Table 4 shows the results of four evaluation indexes and the final evaluation comprehensive weighted index. These results indicate the comprehensive weighted index of linear regression method was the highest, demonstrating

its applicability to the study area and thus the linear regression method was considered the appropriate ideal method for identification and detection of trend analysis and deterministic components decomposition.

4.4 Prediction of streamflow hydrological distribution and frequency analysis

According to the above results, the linear regression was the most suitable for the study area. This method can be used as a reference to investigate the trend of streamflow variation in the Heihe River Basin, and it can reduce the deficiencies in the further research. Thus, prediction of streamflow hydrological distribution and frequency analysis were performed based on the linear regression. Figure 10 shows the quantile-quantile plot of the stochastic components for the normality test. Based on the Shapiro-Wilk test, the results consist of W and P , that is $W = 0.97$, $P\text{-value} = 0.16 > 0.05$. Therefore, the stochastic compo-

Table 3 The deterministic and stochastic components of five methods of analysis of streamflow at Yingluoxia hydrological station

Components	Deterministic	Stochastic
Linear regression method	$X(t) = \begin{cases} 0, & t < 1945 \\ 0.0468t - 0.0468, & 1945 \leq t \leq 2013 \end{cases}$	$S_t = \begin{cases} Y_t, & t < 1945 \\ Y_t - X_t, & 1945 \leq t \leq 2013 \end{cases}$
Nonlinear regression method	$X(t) = \begin{cases} 0, & t < 1945 \\ 14.443e^{0.0028t} - 14.4835, & 1945 \leq t \leq 2013 \end{cases}$	
Change point analysis	$X(t) = \begin{cases} 0, & t \leq 2003 \\ 3.34, & 2003 < t \leq 2013 \end{cases}$	
Wavelet analysis	$X(t) = \begin{cases} 0, & t < 1945 \\ 0.0006t + 0.009t - 0.01, & 1945 \leq t \leq 2013 \end{cases}$	
HHT	$X(t) = \begin{cases} 0, & t < 1945 \\ 0.0002t + 0.0579t - 0.0581, & 1945 \leq t \leq 2013 \end{cases}$	

Note: Y_t , X_t and S_t are the nonstationary streamflow time series (NSTS), deterministic components and stochastic components, respectively; $t = T - 1944$; T , the time.

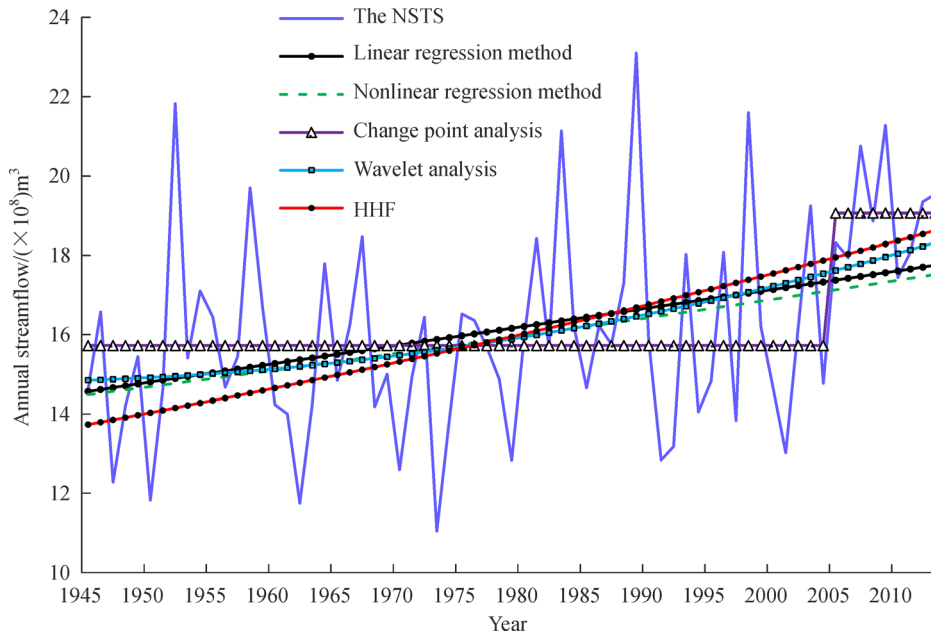


Fig. 9 Header and tail position among five methods of analysis of streamflow at Yingluoxia hydrological station

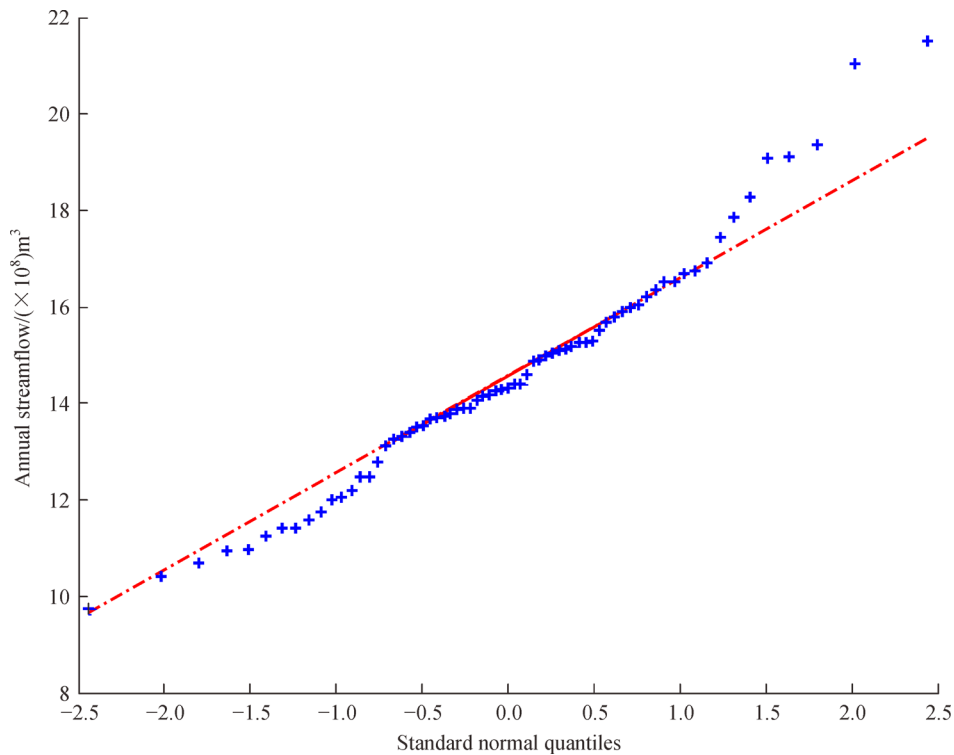


Fig. 10 Quantile-quantile plot of the stochastic components for the normality test for streamflow at Yingluoxia hydrological station

nents of the NSTS data were regarded as normally distributed. The stochastic simulation based on normal distribution was adopted and a complicated procedure of stochastic simulation for Pearson type III distribution avoided. Figure 11 is the probability distribution and cumulative distribution function according to the MCST.

Hence, the predicted streamflow at different hydrological frequencies (Table 5) was obtained from the relationship shown in Fig. 11. The predicted streamflow at hydrological frequencies 1%, 10%, 25%, 50%, 75% and 95% were 24.1×10^8 , 21.2×10^8 , 19.6×10^8 , 18.0×10^8 , 16.3×10^8 and 14.1×10^8 m³. In this study the established

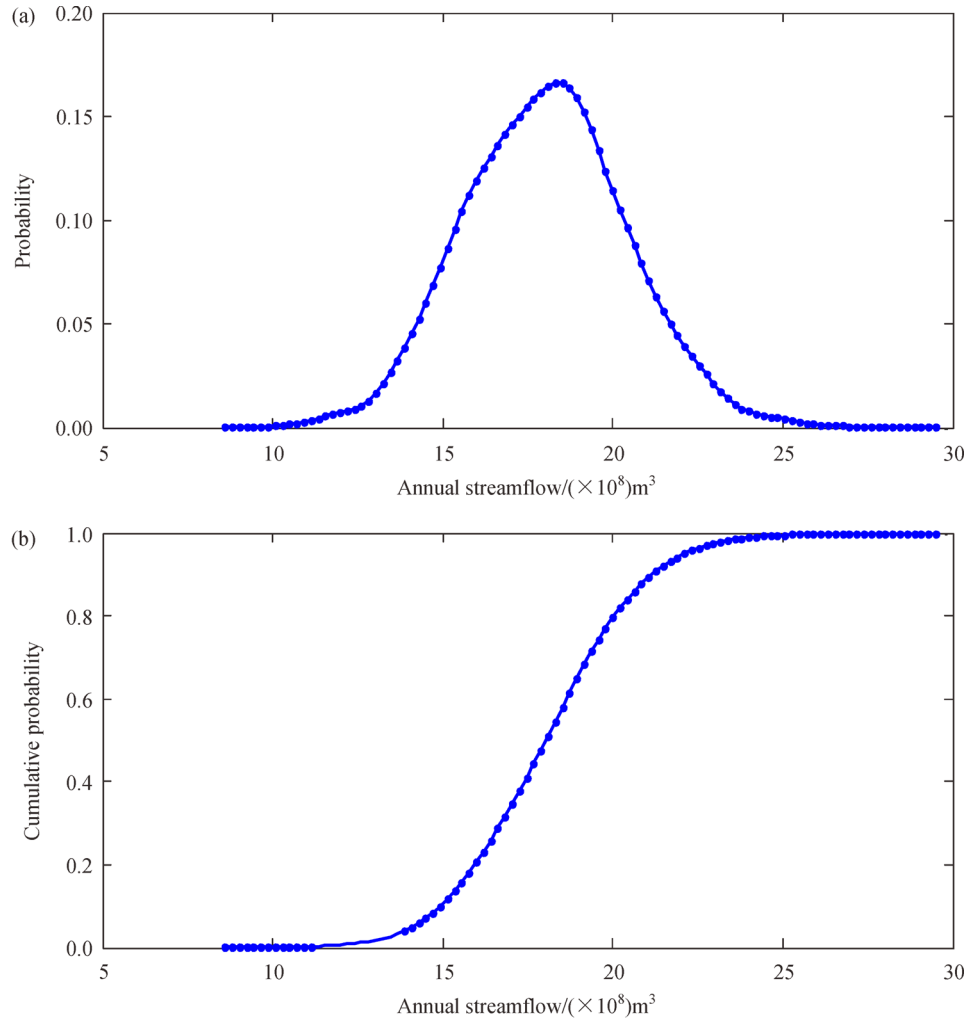


Fig. 11 The probability distribution (a) and cumulative distribution (b) function of the predicted ensemble streamflow time series in 2020 according to the Monte Carlo stochastic simulation technique for streamflow at Yingluoxia hydrological station

Table 4 Results of four evaluation index and evaluation weighted index among five methods of analysis of streamflow at Yingluoxia hydrological station

Evaluation index	Linear regression method	Nonlinear regression method	Change point analysis	Wavelet analysis	HHT
The Nash-Sutcliffe efficiency coefficient/%	36.00	35.56	43.49	39.20	31.64
The Hurst exponent relative value	0.82	0.78	0.74	0.74	0.9
Header morphological coefficient	1.0	0.8	0.2	0.6	0.4
Tail morphological coefficient	0.6	0.2	0.4	1.0	0.8
The weighted index	0.695	0.534	0.444	0.683	0.604

hydrological matching line method was not adopted, but the MCST was used to directly generate the streamflow sampling points to fit cumulative probability distribution function curve. Finally, the predicted streamflow time series at different hydrological frequencies was calculated, which effectively avoided the errors in associated with the traditional (fitting curve) method. The resultant prediction can support regional water resource system management, planning and evaluation.

4.5 Discussion

The above methods for fitting calculating and identifying deterministic components were effectively applied to an historical hydrological time series in which the various physical factors affecting the series remain unchanged. However, if there are significant changes in certain physical conditions (grassland, farmland and land use) in the future, the deterministic components in the NSTS must

Table 5 Hydrological frequency calculation for the predicted of 2020 streamflow at Yingluoxia hydrological station

Rank	Frequency/%	Cumulative probability	Predicted streamflow/($\times 10^8$)m ³	Reappearing period/Year
1	0.01	0.9999	28.44	10000
2	0.1	0.999	26.98	1000
3	1	0.99	24.08	100
4	5	0.95	22.11	20
5	10	0.9	21.15	10
6	20	0.8	20.03	5
7	25	0.75	19.62	4
8	30	0.7	19.26	3.33
9	40	0.6	18.63	2.5
10	50	0.5	18.03	2
11	60	0.4	17.40	1.67
12	70	0.3	16.71	1.43
13	75	0.25	16.34	1.33
14	80	0.2	15.93	1.25
15	90	0.1	14.92	1.11
16	95	0.05	14.09	1.05
17	99	0.01	12.33	1.01
18	99.9	0.001	10.54	1

be predicted again using the corresponding watershed hydrological model considering the changes in land use and vegetation cover that have occurred. Therefore, the prediction method should be considered to be a short-term forecast unsuitable for long-term hydrological ensemble forecasting.

For the analysis of the stochastic components generally, a Pearson type III distribution is widely used in China for hydrological frequency analysis. However, when the stochastic simulation of Pearson type III distribution is performed, there are still shortcomings, such as low precision, a limited range of applicable conditions and computational complexity. Therefore, the MCST based on the normal distribution is proposed on the premise that the stochastic components obey the normal distribution. It is important to establish whether the data are fine-modeled by a normal distribution, because normal distribution is common and easy to calculate provided the mean and standard deviation are known. This can effectively avoid the errors in the traditional (curve fitting) method.

5 Conclusions

In this paper, three methods were first employed to test long-term trend of annual streamflow time series from 1945 to 2013 at Yingluoxia hydrological station. The results showed that the trend changed significantly, which was detected and identified by deterministic and stochastic

components. Then five methods were employed to analyze the deterministic and stochastic components, which are the foundation of long-term hydrologic frequency calculation and variation analysis. To identify the most suitable method and reduce the deficiencies in the methods used for the investigating area, a weighted index was introduced to evaluate the methods applied. The linear regression was effective in identification and detection components of the NSTS, which can be used as a reference for further study. Finally, annual streamflow ensemble sample points in 2020 were obtained by the MCST. Thus the streamflow at different frequencies in 2020 was predicted. The predicted streamflow at hydrological frequencies 1%, 10%, 25%, 50%, 75% and 95% were 24.1×10^8 , 21.2×10^8 , 19.6×10^8 , 18.0×10^8 , 16.3×10^8 and 14.1×10^8 m³, respectively. Using the cumulative probability distribution function, the streamflow at different frequencies can be predicted, which can effectively avoid the error of the stochastic simulation for Pearson type III distribution.

These results indicated that long-term significantly increasing change in annual streamflow may provide more choices for managers. In case of sufficient water, the managers have a positive attitude to agricultural water management for guaranteeing the desired crop irrigation target and crop area planning. Moreover, it can also increase the water availability in the midstream of the Heihe River Basin and thus alleviate the water shortages. However, it will potentially result in the increasing probability of flood events occurring in future. These

predictions can support the regional water resources system management, planning and evaluation.

Acknowledgements The research was supported by the National Natural Science Foundation of China (51439006, 91425302) and the Governmental Public Research Funds for Projects of Ministry of Water Resources (201501017).

Compliance with ethics guidelines Chenglong Zhang, Mo Li, and Ping Guo declare that they have no conflict of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Seyam M, Othman F. Long-term variation analysis of a tropical river's annual streamflow regime over a 50-year period. *Theoretical and Applied Climatology*, 2014, **121**(1–2): 1–15
- Ge Y C, Li X, Huang C L, Nan Z T. A decision support system for irrigation water allocation along the middle reaches of the Heihe River Basin, Northwest China. *Environmental Modelling & Software*, 2013, **47**(47): 182–192
- McCabe G J Jr, Wolock D M. Climate change and the detection of trends in annual runoff. *Climate Research*, 1997, **8**(2): 129–134
- Déry S J, Stahl K, Moore R D, Whitfield P H, Menounos B, Burford J E. Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. *Water Resources Research*, 2009, **45**(4): 546–550
- Douglas E M, Vogel R M, Kroll C N. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology*, 2000, **240**(1–2): 90–105
- Burn D H, Hag Elnur M A. Detection of hydrologic trends and variability. *Journal of Hydrology*, 2002, **255**(1–4): 107–122
- Xu Z X, Takeuchi K, Ishidaira H. Monotonic trend and step changes in Japanese precipitation. *Journal of Hydrology*, 2003, **279**(1–4): 144–150
- Hamed K H. Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology*, 2008, **349**(3–4): 350–363
- Brillinger D R. Trend analysis: Time series and point process problems. *Environmetrics*, 1994, **5**(1): 1–19
- Zang C F, Liu J G. Trend analysis for the flows of green and blue water in the Heihe River basin, northwestern China. *Journal of Hydrology*, 2013, **502**(8): 27–36
- Henderson B. Exploring between site differences in water quality trends: a functional data analysis approach. *Environmetrics*, 2006, **17**(17): 65–80
- Zhang Q, Xu C Y, Becker S, Jiang T. Sediment and runoff changes in the Yangtze River basin during past 50 years. *Journal of Hydrology*, 2006, **331**(3–4): 511–523
- Xu Z X, Gong T L, Li J Y. Decadal trend of climate in the Tibetan Plateau—regional temperature and precipitation. *Hydrological Processes*, 2008, **22**(16): 3056–3065
- Wang W C, Chau K W, Xu D M, Chen X Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management*, 2015, **29**(8): 2655–2675
- Mann H B. Nonparametric test against trend. *Econometrica*, 1945, **13**(3): 245–259
- Mann H B, Whitney D R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 1947, **18**(1): 50–60
- Shao Q X, Li Z L, Xu Z X. Trend detection in hydrological time series by segment regression with application to Shiyang River Basin. *Stochastic Environmental Research and Risk Assessment*, 2010, **24**(2): 221–233
- Li J, Tan S, Wei Z, Chen F, Feng P. A new method of change point detection using variable fuzzy sets under environmental change. *Water Resources Management*, 2014, **28**(14): 5125–5138
- Rahman A, Weinmann P E, Hoang T M T, Laurenson E M. Monte Carlo simulation of flood frequency curves from rainfall. *Journal of Hydrology*, 2002, **256**(3–4): 196–210
- Chen Y, Zhang D, Sun Y, Liu X, Wang N, Savenije H H G. Water demand management: a case study of the Heihe River Basin in China. *Physics and Chemistry of the Earth Parts A/B/C*, 2005, **30**(6–7): 408–419
- Cheng G D, Li X, Zhao W Z, Xu Z M, Feng Q, Xiao S C, Xiao H L. Integrated study of the water–ecosystem–economy in the Heihe River Basin. *National Science Review*, 2014, **1**(3): 413–428
- Helsel D R, Hirsch R M. Statistical methods in water resources. *Studies in Environmental Science*, 2010, **36**(3): 323–324
- Zhang Z X, Dehoff A D, Pody R D, Balay J W. Detection of streamflow change in the Susquehanna River Basin. *Water Resources Management*, 2010, **24**(10): 1947–1964
- Xie P. Study on variation of regional water resources under changing environment. Beijing: *Science Press*, 2012 (in Chinese)
- Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu H H. 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*, 1971, **1998**(454): 903–995
- Huang S, Chang J, Huang Q, Chen Y. Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, 2014, **511**: 764–775
- Huang N E, Wu Z H. A review on Hilbert - Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 2008, **46**(2): 2008
- Koscielny-Bunde E, Kantelhardt J W, Braun P, Bunde A, Havlin S. Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies. *Journal of Hydrology*, 2006, **322**(1–4): 120–137
- Nash J E, Sutcliffe J V. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 1970, **10**(3): 282–290
- Li Z L, Xu Z X, Li J Y, Li Z J. Shift trend and step changes for runoff time series in the Shiyang River basin, northwest China. *Hydrological Processes*, 2008, **22**(23): 4639–4646
- Zhang H, Wang B, Lan T, Shi J, Lu S. Change-point detection and variation assessment of the hydrologic regime of the Wenyu River. *Toxicological and Environmental Chemistry*, 2015, **98**(3): 1–20
- Royston P. Approximating the Shapiro-Wilk W-test for non-

- normality. *Statistics and Computing*, 1992, 2(3): 117–119
33. Razali N M, Wah Y B. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2011, 2(1): 21–33
34. Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renewable Energy*, 2012, 37(1): 241–249
35. Gui Z Y, Zhang C L, Li M, Guo P. Risk analysis methods of the water resources system under uncertainty. *Frontiers of Agricultural Science and Engineering*, 2015, 2(3): 205–215