

Annual runoff prediction of the upstream of Heihe River Basin, China

Y Y Xie¹, S Y Liu^{1,2}, Q Huang¹ and Y M Wang¹

¹State Key Laboratory Base of Eco-Hydraulic Engineering in Arid Area, Xi'an University of Technology, Xi'an city, 710048, China

E-mail: 719685861@qq.com

Abstract. A prediction method of annual runoff is proposed for the upstream Heihe River Basin. The Mann-Kendall test, order cluster analysis, variance analysis and R/S analysis methods were employed to investigate trends, mutations, periods and persistence of annual runoff time series, respectively. Based on various components, the linear superposition method was applied to forecast annual runoff. The Nash-Sutcliffe efficiency coefficient (*NSE*), the mean absolute percentage error (*MAPE*) and the percentage of pass (*POP*) were utilized to evaluate the simulation and prediction effects of annual runoff. Moreover, the calibration period was between 1958 and 2010, while the validation period was from 2011 to 2015. The results show that i) the annual runoff time series had a significant increasing trend ($0.65 \times 10^8 \text{ m}^3/10\text{a}$), a probable mutation at 2006 and two significant periods (6 years and 22 years) in the calibration period; ii) the randomization of annual runoff time series was improved after removing the trend, mutation and periodic components, with the Hurst exponents 0.52; and iii) the simulation effects of calibration period (*NSE*=0.74, *MAPE*=6.6% and *POP*=90.6%) and the prediction effects of validation period (*MAPE*=6.5%, *POP*=80%) jointly indicate that the annual runoff prediction method is suitable in this river basin.

1. Introduction

Heihe River, the second largest inland river of China, is located in the northwest of China with the arid or semi-arid climate. It supplies the important water resource to the socioeconomic development and ecological environmental protection of northwestern China. The upstream of the river basin is the main runoff-yield region which is fed by rainfall and ice-snow melt water [1]. In the middle of the basin, there distribute many agricultural irrigation districts which consume a large amount of water each year [2]. The downstream of the river flows through desert and Gobi area, where the evaporation is intensive and the ecosystem is fragile [3].

In Heihe River Basin, it is very important to correctly handle the relations among irrigation, ecologic environment and hydroelectric generation under the condition of limited water resources. From 1960th, the annual runoff in the downstream river basin reduced sharply because of the rapid population growth and extensive irrigation area in the middle river basin, causing a serious degradation of ecological environment of the middle and lower river basin [4-6]. Since 2000, the government of China has taken many water allocation measures to relieve the water contradiction between irrigation and ecological environment in the river basin, achieving some success [7-9]. However, the water contradiction is still serious currently, and the downstream ecological environment has not yet restored to the level of the 1980th [10, 11]. Besides, the upstream of the river has rich



waterpower resources, and many hydropower plants were built there in the first decade of 21st century. These hydropower stations provide reliable electric power for the local socioeconomic development, so the requirements of the hydropower plants should be properly taken into account when allocating the surface water of Heihe River [12].

The annual runoff at the upper stream could directly influence the water allocation between different users of Heihe River Basin, since the upstream basin is the main runoff generation area. In order to better realize the comprehensive benefits of irrigation, ecologic environment and hydroelectric generation, the decision makers always make the water distribution schemes based on the next year's runoff volume in advance. Therefore, it is of great significance to predict the annual runoff at the upstream of the river basin.

The prediction of annual runoff belongs to long-term runoff forecast, which could be solved by three kinds of common methods. The first one is the time series analysis method, which acquires the future runoff based on the internal relations (autocorrelation and relations with time) of the historical runoff series [13, 14]. The second one is the relevant factor analysis method, which forecasts the runoff based on the statistical relations between runoff and the other factors such as precipitation, atmospheric temperature and wind speed [15-17]. The last one is runoff generation and concentration model, which predicts the runoff according to the physical mechanism of runoff formation [18-20]. Among the three methods, the first one is the simplest because it only needs the historical runoff data, but it lacks physical mechanism and excessively depends on the constant statistical characteristics of runoff series. The second one could demonstrate the explicit causes and course of runoff formation, but it is the most complicated approach which needs a great mass of data (hydrologic data, meteorological data, vegetation data, soil data, land use data, etc.). As to the third one, its advantages and disadvantages are between the first and second methods. On the whole, it is difficult to prove which way is best for forecasting the annual runoff due to many reasons (e.g. assumed conditions, empiric formulas, data deficiencies and error propagation) without considering specific river basins.

The time series analysis method is often employed to forecast the annual runoff of Heihe River because of the insufficient observation data and the unrevealed ice-snow melting mechanism of the river basin. Chen *et al* [17] adopted grey model to the extract tendency part of runoff time series, employed the wavelet analysis method to extract the period part, utilized the autoregressive method to deal with the stochastic part, and then built the runoff prediction model of Heihe River. Jiang and Liu [21] combined the wavelet analysis method with radial basis function network to forecast the annual runoff of Heihe River. Chu *et al* [22] applied Mann-Kendall method to study the trend of runoff time series of Heihe River, used the coupling method of periodic wave extension and stepwise regression to determine the period part. Zhang and Ding [23] employed the grey topological method to predict runoff volume at the valley outlet of Heihe River. Hou *et al* [24] established the annual runoff forecast model of Heihe River through the periodic mean superposition method. Li *et al* [25] proposed a new runoff prediction model of Heihe River based on R/S analysis method and grey system theory. It can be seen that the applications of the time series analysis method to runoff prediction of Heihe River in the past mainly focused on three components of runoff time series (trend, periodic and stochastic components), and merely provided the deterministic forecasting results to the decision-makers. However, the annual runoff series of the upstream Heihe River probably has mutation and becomes more uncertain in the background of regional climate change [26, 27], which may bring to the decision of water allocation certain risks in future.

Specific to the upstream Heihe River, this study still adopts the time series analysis method to build the annual runoff prediction model, but it is different from the prior studies. Firstly, the mutation of annual runoff series will be investigated before establishing the model. Secondly, the model not only provides the deterministic prediction results but shows the main variation range of runoff, which could help the decision-makers make adequate preparation against the risks of uncertainty. Moreover, the effect of the extraction of deterministic components on annual runoff series will be discussed. Therefore, this study is organized as follows: Section 2 introduces the general situation of the Heihe River; Section 3 displays the data and methods for trend analysis, mutation test, period detection,

persistence analysis and prediction of annual runoff; in Section 4, the results are presented and discussed; and the conclusions are drawn in Section 5.

2. Study area

Heihe River is a typical inland river of China, flowing through Qinghai, Gansu and Inner Mongolia provinces from south to north (figure 1). The river originates from the north foot of Qilian Mountain and feeds into the West Juyan Lake and the East Juyan Lake, with a core drainage area of 130,000 km² and a mainstream length of 821 km. The upstream Heihe River (UHR), the study area, is between the river sources and Yingluoxia canyon with the drainage area of 10,009 km², the river length of 303 km and the average channel gradient of 9.1‰. UHR is characterized by the mountainous terrains with the average elevation of 3737.7 m. The climate of UHR is cold and moist, with the annual average air temperature of 2°C and the annual average precipitation of 350 mm. Many mountain areas of UHR with high elevation (more than 4,000 m) are annually snow-covered. Moreover, a number of modern glaciers distribute in UHR, which is the natural reservoir of northwestern China. Yingluoxia canyon is the outlet of UHR with the annual average runoff volume of 1.6 billion m³, which will be mainly used for irrigating crops, recharging groundwater, improving ecological environment in the middle and lower Heihe River.

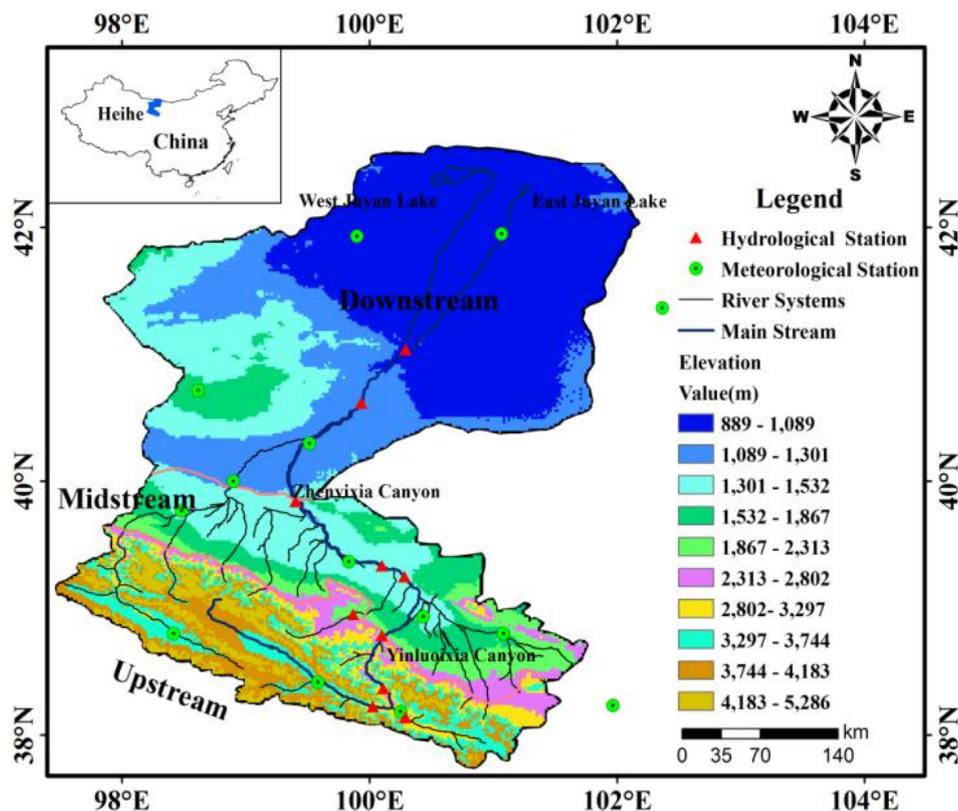


Figure 1. Map of Heihe River basin.

3. Materials and methods

3.1. Materials

The average daily flow data of Yingluoxia canyon is observed at Yingluoxia hydrologic station (figure 1), which was built in 1943. Due to the poor field monitoring condition and the backward data reorganization technology, the reliability of average daily flow data before 1958 is much lower.

Therefore, the annual runoff data of Yingluoxia station between 1958 and 2015 are employed in this study. Moreover, UHR annual runoff data between 1958 and 2010 are applied to calibrate the prediction model, and the remaining data (2011-2015) are utilized to validate the model.

3.2. Methods

3.2.1. Trend analysis method. Mann-Kendall test is widely applied to detect the trends of hydrologic time series because it is a nonparametric test and is not sensitive to the mutations of nonstationary time series. The statistic (U) is built as below [28, 29]:

$$U = \frac{12K - 3N(N-1)}{\sqrt{2N(N-1)(2N+5)}} \quad (1)$$

$$K = \sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{ij} \quad (2)$$

$$k_{ij} = \begin{cases} 1 & \text{if } X_j > X_i \\ 0 & \text{if } X_j \leq X_i \end{cases}, 1 \leq i < j \leq N \quad (3)$$

where N is the length of hydrologic time series; X_i ($i=1, 2, \dots, N$) is the i^{th} hydrologic variable (e.g. annual runoff); k_{ij} is the paired value between X_i and X_j ; and K is the summation of all the paired values.

When N is large enough ($N \geq 10$), the statistic U approximately obeys the standard normal distribution. The original hypothesis is that the hydrologic time series has not any obvious trends. The significance level α is given before the trend test. If $|U| \leq U_{\alpha/2}$, the original hypothesis is accepted, i.e. the series has no obvious trends; otherwise, the original hypothesis is refused, which shows that there exists evident tendency in the series. Besides, the sign of U represents the trend direction ('+' represents the growth trend while '-' indicates the decreasing tendency), and the absolute value of U shows the obvious degree of trend (the degree will become more and more obvious with the increasing of $|U|$).

In addition to the overall trend analysis, the local trends are also investigated in the trend analysis of hydrologic time series. The turning points and stages of various local trends are distinguished by the trend variation index (T), which is defined as follows:

$$T_{\tau} = |t_{\tau,1} - t_{\tau,2}| \quad (4)$$

where τ ($10 < \tau \leq N-10$) is the division point of series; $t_{\tau,1}$ and $t_{\tau,2}$ represent the linear trend slopes of the former and latter subseries divided by τ , respectively.

In equation (4), when T_{τ} reaches the maximum value, the division point τ is considered as one turning point, which divides the series into two obvious different stages of local trends. Moreover, if a hydrologic time series has l ($l=1, 2, \dots$) turning points, it has $l+1$ different trend stages.

3.2.2. Mutation analysis method. The mutation represents an irreversible jump of hydrologic time series at a certain time. The order cluster analysis method could effectively find the most probable mutation point of hydrologic time series [30]. Its principle is to find out an optimal segment point dividing the series into two subsequences, where the variable difference in the same subsequence is the minimum and the variable difference between two subsequences is the maximized.

The mutation index (Z) of hydrologic time series is computed as follows [30]:

$$Z_{\tau} = \frac{(\overline{X_{\tau}} - \overline{X_{N-\tau}})^2}{S_{\tau}^2 + S_{N-\tau}^2}, 2 \leq \tau < N-1 \quad (5)$$

where τ is possible mutation point of series; $\overline{X_{\tau}}$ and S_{τ} are the mean value and the standard deviation of the subsequence before τ , respectively; $\overline{X_{N-\tau}}$ and $S_{N-\tau}$ represent the mean and the

standard deviation of the subsequence after τ , respectively.

In equation (5), the square of mean difference (the numerator part) represents the variable difference between two subsequences, and the variance sum (the denominator part) reflects the variable difference in the same subsequence. According to the principle of the order cluster analysis method, the point where Z reaches maximum is the most probable mutation point of time series.

After identifying the most probable mutation point, the next necessary step is to test its significance. Rank sum test is also a common nonparametric method in examining the obvious degree of mutation of hydrologic time series [30]. Firstly, the figures of the original series are sorted from small (large) to large (small). Secondly, the rank of each figure of the short subsequence in the sorted series is recorded. Finally, the summation of these the ranks (rank sum) is calculated and employed to establish the statistic of rank sum test.

The statistic (U) of rank sum test is built as below [30]:

$$U = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}} \quad (6)$$

where W is the rank sum of short subsequence; n_1 is the length of the short subsequence; and n_2 is the length of the long subsequence.

When n_1 is large enough ($n_1 \geq 10$), the statistic U approximately obeys the standard normal distribution. The original hypothesis is that the series has no mutation points, and the significance level α is given before rank test. If $|U| \leq U_{\alpha/2}$, the original hypothesis is accepted, i.e. the series has no mutation; otherwise, the original hypothesis is refused, i.e. there is a remarkable mutation in the series. Moreover, the difference between the means of the series before mutation and after mutation is the jump height, which could also be understood as the mutation component of series.

3.2.3. Period analysis method. Variance analysis is a very useful period identification method, and its obvious advantage is that it is not confined to the exact form of periodic wave function [31, 32]. The period detection process of variance analysis is described in the following [32].

- Divide the time series into an equal number of groups based on the length of the assumed period.

$$m_j = \begin{cases} N/P, & \text{if } r = 0 \\ (N-r)/P + 1, & \text{if } r \neq 0 \& j \leq r \\ (N-r)/P, & \text{if } r \neq 0 \& j > r \end{cases} \quad (7)$$

$$r = \text{mod}(N, P), r < P \quad (8)$$

where P ($2 \leq P \leq N/2$) is the assumed period; r is the remainder when dividing N by P ; and m_j ($j=1, 2, \dots, P$) is the size of the j^{th} group.

- Calculate the mean values of original time series and each group.

$$\bar{X} = \sum_{i=1}^N X_i \quad (9)$$

$$\bar{X}_j = \frac{1}{m_j} \sum_{k=1}^{m_j} X_{kj} \quad (10)$$

$$X_{kj} = \begin{cases} X_{(j-1)m_1+k}, & \text{if } r = 0 \\ X_{(j-1)m_1+k}, & \text{if } r \neq 0 \& j \leq r \\ X_{rm_1+(j-r-1)m_{r+1}+k}, & \text{if } r \neq 0 \& j > r \end{cases} \quad (11)$$

where X_i ($i=1, 2, \dots, N$) is the i^{th} figure of original time series; \bar{X} is the mean of original series; \bar{X}_j ($j=1, 2, \dots, P$) is the mean of the j^{th} group; and X_{kj} ($k=1, 2, \dots, m_j$) is the k^{th} figure of the j^{th} group.

- Compute the intra-group and inter-group sum of deviation square.

$$S_1^2 = \sum_{j=1}^P \sum_{k=1}^{m_j} (X_{kj} - \bar{X}_j)^2 \quad (12)$$

$$S_2^2 = \sum_{j=1}^P m_j (\bar{X}_j - \bar{X})^2 \quad (13)$$

where S_1^2 is the intra-group sum of deviation square; and S_2^2 is the inter-group sum of deviation square.

- Calculate the value of statistic F , give the significance level α and test the significance of the supposed period.

$$F = \frac{S_2^2 / (P-1)}{S_1^2 / (N-P)} \quad (14)$$

In the definition formula (14), the statistic F follows F distribution with the degree of freedom $(P-1, N-P)$. The significance level α is given before F test. If $F > F_\alpha(P-1, N-P)$, the assumed period P is significant; otherwise, it is not obvious. After finding the first significant period, its periodic wave could be constructed according to the mean values of corresponding groups. Generally, there often exist more than one obvious period in the hydrologic time series. Hence, it is essential to continue to search an additional number of significant periods after extracting the first periodic process. Meanwhile, the number of prominent periods of series is not more than three considering the negative impact of some pseudo periods [32].

3.2.4. Persistence analysis method. The persistence analysis of runoff series has been initiated since the middle of 20th century, which was rapidly expanded to many other geophysical records including temperature and precipitation data [33]. Currently, it is proved that the time series with long range dependence property (persistence) widely exist in the natural world [34]. Hurst exponent H ($0 < H < 1$), presented by Hurst in 1951, could quantitatively describe the persistence of hydrologic time series very well [35]. Specific to Hurst exponent, there are three kinds of persistence condition in the following.

- If $H=0.5$, the hydrologic time series is an independent random sequence with a limited variance.
- If $0.5 < H < 1$, the series has long range dependence property (persistence), which indicates that the future tendency is consistent with the past trend. Moreover, the persistence of series will endlessly strengthen when H approaches 1.
- If $0 < H < 0.5$, the series still has long run memory property (anti-persistence), which shows that the future tendency is contrary to the past trend. Besides, the anti-persistence of series will continually enhance when H is close to 0.

At present, there are many methods for estimating Hurst exponent [34]. Among them, R/S analysis is the most universal method because of its robustness and nonparametric analysis feature (i.e. it has no special demands on the distribution type of time series). The calculation steps of Hurst exponent by R/S analysis method are expressed as below [33, 34].

- Create an accumulated deviation series (ADS)

$$Y_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} X_i, 4 \leq \tau \leq N \quad (15)$$

$$A(i, \tau) = \sum_{j=1}^i [X_j - Y_\tau], 1 \leq i \leq \tau \quad (16)$$

where τ is the length of ADS; Y_τ is the mean of time series $\{X_1, X_2, \dots, X_\tau\}$; and $A(i, \tau)$ is the i^{th} variable of ADS.

- Calculate the range and the standard deviation of ADS

$$R(\tau) = \max[A(i, \tau)] - \min[A(i, \tau)] \quad (17)$$

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} [X_i - Y_i]^2} \quad (18)$$

where $R(\tau)$ is the range of ADS; and $S(\tau)$ is the standard deviation of ADS.

- Estimate Hurst exponent

$$\lg \frac{R(\tau)}{S(\tau)} = H \lg \tau + C \quad (19)$$

where C is a constant.

In equation (19), a series of data points $\left(\lg \tau, \lg \frac{R(\tau)}{S(\tau)} \right)$ could be obtained when τ ($4 \leq \tau \leq N$) is assigned different values. Finally, Hurst exponent H is estimated through the least square method.

3.2.5. Prediction and evaluation method. The linear supposition method is an important time series analysis method [36]. In analyzing hydrologic time series X , it is assumed that the series could be linearly decomposed to five parts including the constant component X_C , the trend component X_T , the mutation component X_M , the period component X_P and the stochastic component X_S . Except for X_S , the other components are collectively called deterministic components. The linear supposition method is expressed as follows [36]:

$$X = X_C + X_T + X_M + X_P + X_S \quad (20)$$

In equation (20), the different components of X series should be estimated and extracted in turn. X_C is considered as the mean value X'_C of X series; X_T is estimated by the linear trend part X'_T of $X-X'_C$ series based on Mann-Kendall test result; X_M is estimated by the jump height X'_M at the mutation point of $X-X'_C-X'_T$ series according to the consequence of mutation analysis; X_P is estimated by the periodic part X'_P of $X-X'_C-X'_T-X'_M$ series through the variance analysis method; and the remaining part $X-X'_C-X'_T-X'_M-X'_P$ (X'_S) is thought as the estimation value of X_S . Moreover, X'_S could be further subdivided into a dependent random component X'_{S1} and an independent random component X'_{S2} . Hurst exponent H of X'_S calculated by R/S analysis method determines the proportions of two components in X'_S series. If $H \leq 0.4$ or $H \geq 0.6$, the proportion of X'_{S1} could not be neglected. On this occasion, X'_{S1} should be estimated through autoregressive model [17] and stripped from X'_S . Then, the remaining part of X'_S is X'_{S2} . If $0.4 < H < 0.6$, X'_S could be regarded as an independent random series, i.e. $X'_{S2} = X'_S$. Without taking X'_{S2} into account, the medium estimation value X^M of the X series could be calculated as below.

$$X^M = X'_C + X'_T + X'_M + X'_P + X'_{S1} \quad (21)$$

In order to reduce the uncertainty of prediction, it is essential to master the main variation range of X'_{S2} . According to the probability distribution type of X'_{S2} , its lower and upper bounds could be defined as

$$p = p[X'_{S2} \geq X'_{S2}(p)] \quad (22)$$

$$X^L = X^M + X'_{S2}(q) \quad (23)$$

$$X^U = X^M + X'_{S2}(1-q) \quad (24)$$

where p is the occurrence probability of event $X'_{S2} \geq X'_{S2}(p)$; X^L and X^U are the lower and upper estimation values of X , respectively; q ($0.5 < q < 1$) is the appointed probability; $X'_{S2}(q)$ and $X'_{S2}(1-q)$ represent the lower and upper bounds of X'_{S2} , respectively.

In this study, Nash-Sutcliffe efficiency coefficient (*NSEC*) and the mean absolute percentage error

(MAPE) are employed to evaluate hydrological forecast effects [37, 38]. NSE ($NSE \leq 1$) and $MAPE$ ($0 \leq MAPE \leq 1$) are two commonly used indexes in evaluating hydrologic prediction model. The forecast accuracy improves with the increasing of NSE and the reducing of $MAPE$. Instead of measuring the forecast accuracy, and the percentage of pass (POP) is defined to evaluate the reliability of prediction variation range.

$$NSE = 1 - \frac{\sum_{i=1}^N (X_i^R - X_i^M)^2}{\sum_{i=1}^N (X_i^R - \bar{X}^R)^2} \quad (25)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{X_i^R - X_i^M}{X_i^R} \right| \times 100\% \quad (26)$$

$$POP = \frac{M}{N} \times 100\% \quad (27)$$

where X_i^R and X_i^M are the i^{th} ($i=1, 2, \dots, N$) observed and medium estimation values of hydrologic series, respectively; \bar{X}^R is the mean value of observed hydrologic series; and M is the number of events that X^R is located between X^L and X^U .

4. Results and discussion

The average annual runoff of Yingluoxia station was $16.04 \times 10^8 \text{ m}^3$ between 1958 and 2010 (calibration phase). Thus, the constant component X^C of UHR annual runoff series X was equal to $16.04 \times 10^8 \text{ m}^3$, which would be stripped from X before extracting the other components.

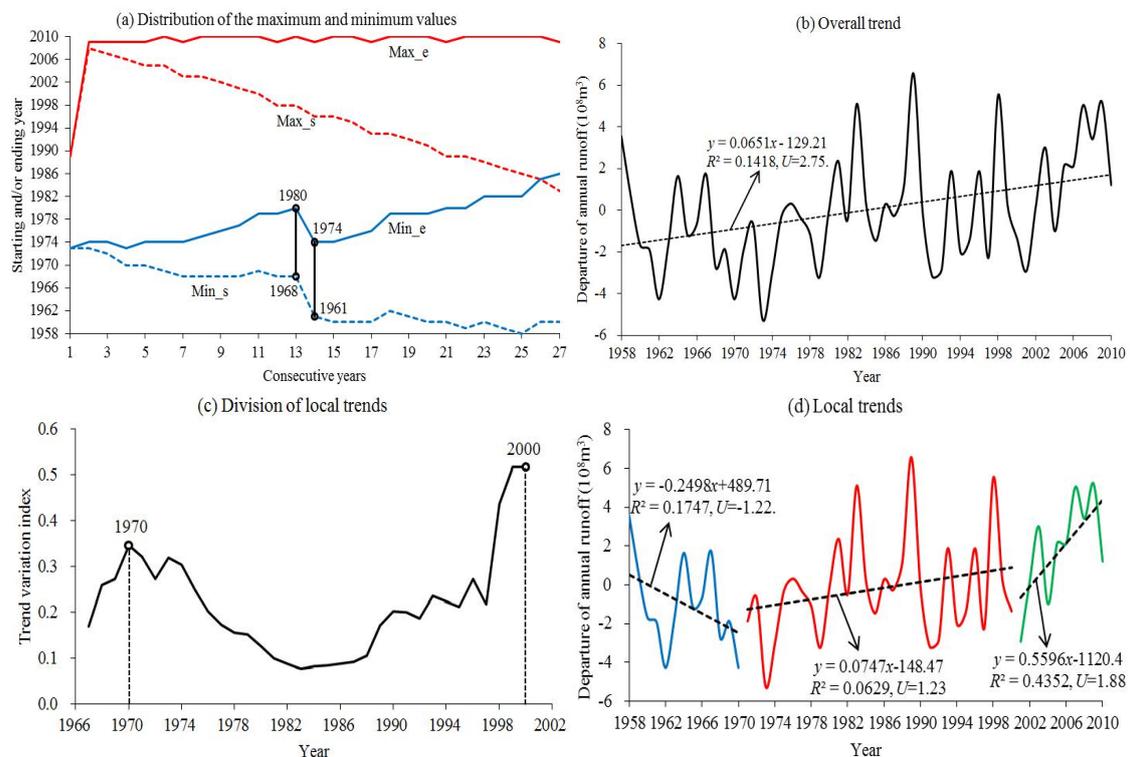


Figure 2. Trend test results.

In graph (a), Max_s and Max_e represent the start years and the end years of n ($n=1, 2, \dots, 27$) consecutive years with the maximum runoff volume, respectively; Min_s and Min_e are the start years and the end years of n consecutive years with the minimum runoff volume respectively. In graph (b) and (d), the solid curve is $X - X^C$

series, and the dotted line represents linear trend.

4.1. Trend component

Based on Mann-Kendall test with the significant level $\alpha=0.05$ ($U_{\alpha/2}=1.96$), the trend test results of $X-X_C$ series were shown in figure 2. In graph (a), Max_s curve was above Min_e curve from 1 to 26 consecutive years, which directly indicated that $X-X_C$ series had a clear increasing tendency between 1958 and 2010. In addition, Min_s and Min_e curves had obvious abrupt turns between 13 and 14 consecutive years, indicating a trough (1968-1974) of $X-X_C$ series. In graph (b), the overall trend of $X-X_C$ series was significant with a linear growth speed $0.65 \times 10^8 \text{ m}^3/10\text{a}$. The expression of X_T was the trend equation of $X-X_C$ series. In graph (c), it showed that $X-X_C$ series had two remarkable turning points (1970 and 2000). Thus, the series had three stages of local trends, which were displayed in graph (d). The first stage was between 1958 and 1970, with a decreasing speed $2.50 \times 10^8 \text{ m}^3/10\text{a}$; the second stage ranged from 1971 to 2000, with a slow growth speed $0.75 \times 10^8 \text{ m}^3/10\text{a}$; and the third stage started from 2001 and ends at 2010, with an increasing speed $5.60 \times 10^8 \text{ m}^3/10\text{a}$. However, none of the three local trends were significant through Mann-Kendall test. The overall trend and local trends of $X-X_C$ series were closely related to the variation processes of precipitation and atmospheric temperature in UHR during the past 50 years before 2010 [1, 3, 39].

4.2. Mutation component

The significant overall trend component of hydrologic time series easily misled the mutation analysis, taking the middle year of time series as its mean value mutation year. Hence, it is necessary to strip the trend component from $X-X_C$ series at first. Through the order cluster analysis method, the mutation analysis result was displayed in figure 3. It could be seen that the most probable mutation point of $X-X_C-X_T$ series occurred at 2006, and the jump height at 2006 was $2.0 \times 10^8 \text{ m}^3$ occupying 12.5 percent of the mean annual runoff of UHR. However, as far as the sample size of calibration phase was concerned, the length of annual runoff series after 2005 was so short (only 5 years) that the current mutation test methods (including the rank sum test method) were not suitable to examine the significance of the most probable mutation point. Thus, it could not be confirmed whether the mean value of $X-X_C-X_T$ series had a significant mutation at 2006. When estimating the annual runoff series in the validation phase, it was conservatively thought that $X-X_C-X_T$ series had no mutation or mutation of mean value at 2006, i.e. $X_M=0 \text{ m}^3$.

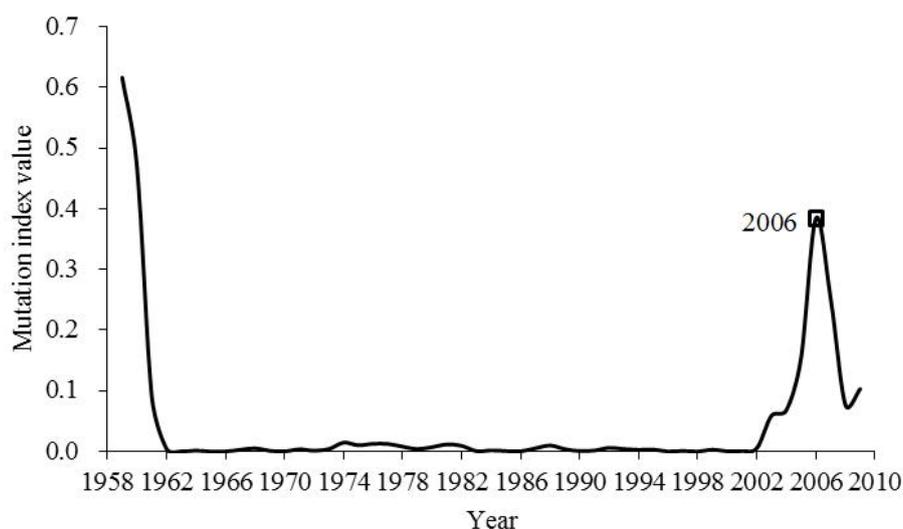


Figure 3. Mutation analysis result.

4.3. Periodic components

The period analysis results were shown in figure 4. The left graph displayed two significant simple periods ($P_1=22$ years and $P_2=6$ years) of $X-X_C-X_T-X_M$ series, and the right one presents their respective variation processes. Through the variance analysis method, the 22 years' period was found in the first F test when given the significance level $\alpha=0.05$. After extracting the 22 years' period component, there was no obvious period component in the remaining part in the second F test when $\alpha=0.05$. Generally, the potential periods in the second test were less significant than the periods detected in the first test. Based on this, it was imperative to lower the significant level to search the potential periods in the second test. When given the significance level $\alpha=0.1$, the 6 year' period was found in the second test, but no remarkable periods were found in the third test. So far, the period detection work of $X-X_C-X_T-X_M$ series by variance analysis method was finished. The two significant simple period components of UHR annual runoff series mainly result from the sun activity [17]. At last, the synthesized period component X_P was acquired through linear superposition of two simple period components.

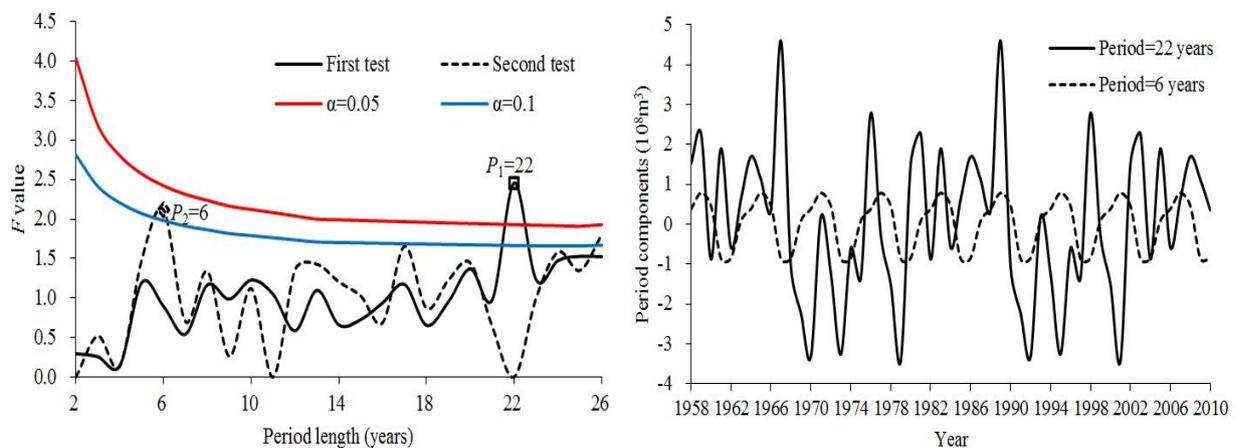


Figure 4. Period detection result (left) and period components (right).

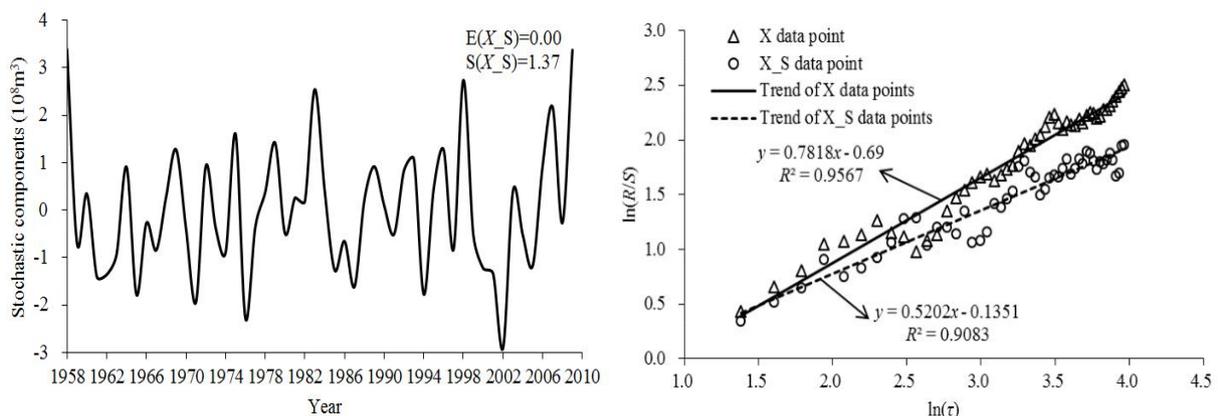


Figure 5. Stochastic component (left) and Hurst exponent variation (right).

4.4. Stochastic component

After extracting all the deterministic components from the UHR annual runoff series X , the remaining part was the stochastic component X_S which was displayed in the left graph of figure 5. Further, Hurst exponents of X and X_S were also calculated through R/S analysis method, and the results were presented in the right graph of figure 5. Seen from the right graph, Hurst exponents of X and X_S were 0.78 and 0.52. The two Hurst exponents showed that X had a remarkable long run dependence

property and the randomness of X'_s was very strong. Meanwhile, the great change of Hurst exponent also proved that the persistence of UHR annual runoff series was mainly caused by its deterministic components. Since Hurst exponent of X'_s was very close to 0.5, it was believed that X'_s did not contain any dependent random components. Hence, the following work was to investigate the main distribution range of X'_s as an independent random series.

Numerous studies had shown that the hydrologic time series of various rivers in China could be well fitted by Pearson III (P-III) type distribution [40]. Considering many negative values in X'_s series, P-III distribution type was employed to fit $X'_c+X'_s$ series. The theoretical curve-fitting result was shown in the left graph of figure 6, and it can be seen that the curve fitting ($R^2=0.9888$) was very successful. Further, the distribution range of X'_s was also obtained based on theoretical frequency curve. In the right graph, according to equations (22)-(24), the accuracy rate of prediction was equal to $q-(1-q)$ ($0.5 < q < 1$), and the forecast range is equal to $X'_s(1-q)-X'_s(q)$. The forecast range rose faster and faster with the growth of accuracy rate, which exactly brought a dilemma to the prediction of annual runoff. Even so, the accuracy rate of prediction still attracted more attention. In this study, the appointed probability q was assigned 0.95, and the forecast range was $4.66 \times 10^8 \text{ m}^3$ with the accuracy rate 90%.

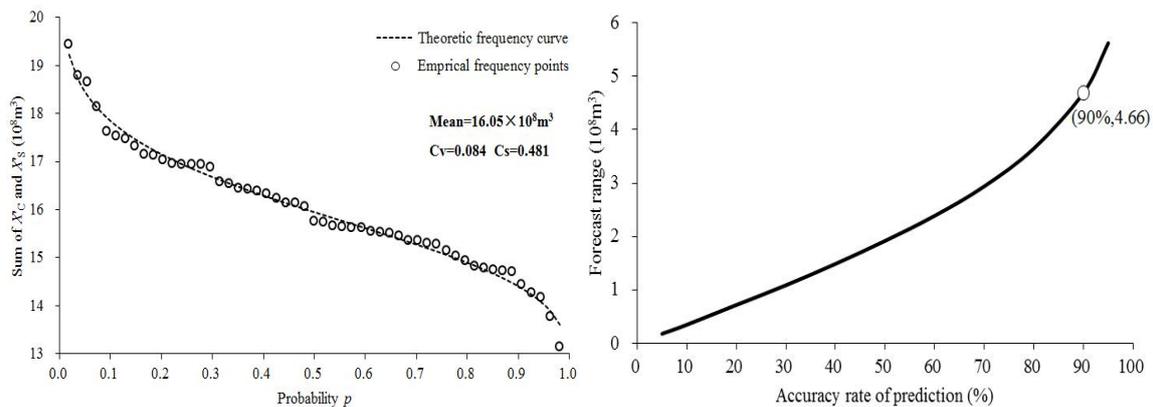


Figure 6. Theoretical curve-fitting result of probability distribution of $X'_c+X'_s$ series (left) and relation between forecast range and accuracy rate of prediction (right).

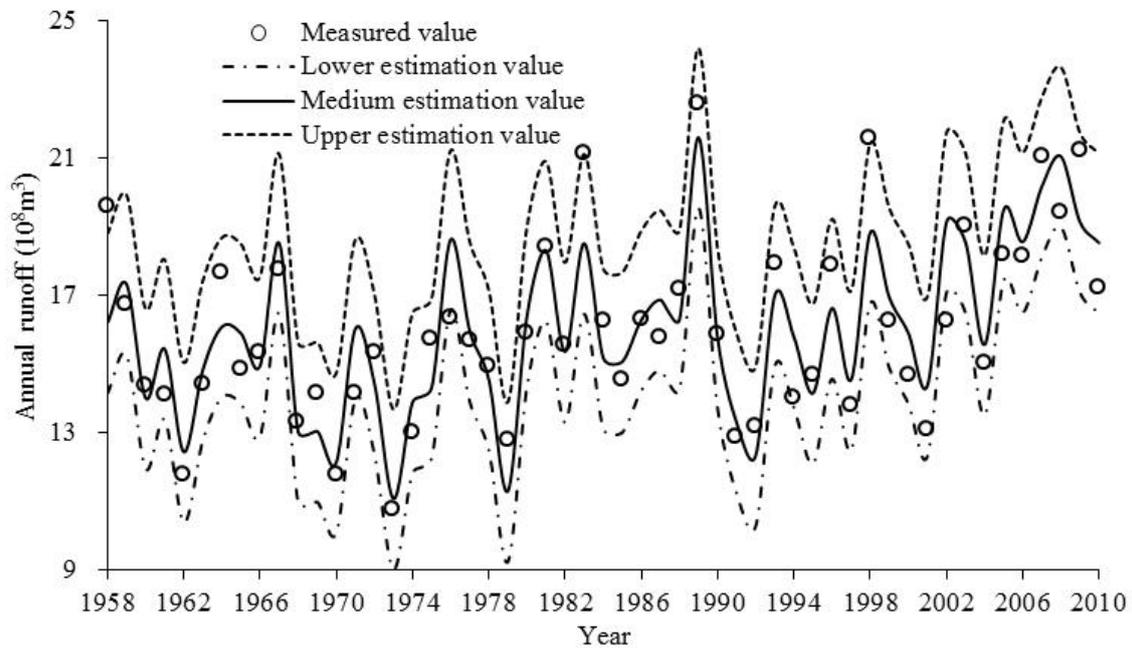


Figure 7. Forecast results of annual runoff of UHR in calibration phase.

4.5. Prediction and evaluation

Based on the deterministic and stochastic components, the estimation results of UHR annual runoff series in the calibration phase (1958-2010) were exhibited in figure 7. The evaluation indexes ($NSE=0.74$, $MAPE=6.6\%$, $POP=90.6\%$) showed that the simulation effects were satisfactory in calibration phase. Moreover, table 1 displayed the medium, lower and upper estimation values of UHR annual runoff series in the validation phase (2011-2015). Seen from this table, the estimation range failed to cover the measured value at 2011, and the medium estimation values were always smaller than the measured values in four consecutive years (2012-2015). Nonetheless, the estimation effects in validation phase were still acceptable on the basis of the evaluation indexes ($MAPE=6.5\%$, $POP=80\%$). Therefore, the model based on the time series analysis method was feasible in the prediction of UHR annual runoff.

Table 1. Estimation results of UHR annual runoff in validation phase. X is the measured value of annual runoff; X^M , X^L , and X^U are the medium, lower, and upper estimation values of X , respectively.

Year	$X(10^8m^3)$	$X^M(10^8m^3)$	$RE(\%)$	$X^L(10^8m^3)$	$X^U(10^8m^3)$
2011	18.95	21.62	14.1	19.56	24.22
2012	19.34	19.15	-1.0	17.09	21.75
2013	19.52	18.40	-5.7	16.34	21.00
2014	18.37	17.12	-6.8	15.06	19.72
2015	20.24	19.28	-4.7	17.22	21.88

Note: RE is the relative error of X^M , i.e. $RE=(X^M-X)/X\times 100\%$.

In order to provide the basis for the water allocation of Heihe river basin, the annual runoff series in future five years (2016-2020) were calculated and shown in table 2. In addition, the future UHR annual runoff series probably had some unexpected variations under the double impact of regional climate change and human activities. Hence, it was of great significance to update the current UHR annual runoff prediction model utilizing the latest observation data regularly.

Table 2. Prediction results of UHR annual runoff between 2016 and 2020.

Year	X^M (10^8 m ³)	X^L (10^8 m ³)	X^U (10^8 m ³)
2016	18.02	15.96	20.62
2017	17.17	15.11	19.77
2018	20.02	17.96	22.62
2019	19.66	17.60	22.26
2020	23.61	21.55	26.21

5. Conclusions

An annual runoff prediction method was adopted to forecast annual runoff of UHR based on the linear superposition method. In other words, various components were extracted and investigated by decomposing the historical annual runoff series, hence constructing the forecast model through the linear superposition of these components. The main conclusions are drawn as below:

- During 1958 and 2010, the annual runoff of the upstream Heihe River had an obvious increasing trend with a speed 0.65×10^8 m³/10a, a probable mutation at 2006, two significant periods (22 years and 6 years) and a persistence property.
- The adopted method based on the linear superposition method is suitable to forecast the annual runoff in the Heihe River.

Acknowledgments

This study is funded by the Natural Science Foundation of China (91325201) and China Scholarship Council.

References

- [1] Wang J, Li H, and Hao X 2010 Responses of snowmelt runoff to climatic change in an inland river basin, Northwestern China, over the past 50 years *Hydro. Earth Sys. Sci. Discuss.* **14** 1979-87
- [2] Ge Y, Li X, Huang C and Nan Z 2013 A Decision Support System for irrigation water allocation along the middle reaches of the Heihe River Basin, Northwest China *Environ. Modell. Softw.* **47** 182-92
- [3] Zhang A, Zheng C, Wang S and Yao Y 2015 Analysis of streamflow variations in the Heihe River Basin, northwest China: Trends, abrupt changes, driving factors and ecological influences *J. Hydrol-Regional Studies* **3** 106-24
- [4] Lu L, Li X and Cheng G 2003 Landscape evolution in the middle Heihe River Basin of north-west China during the last decade *J. Arid Environ.* **53** 395-408
- [5] Qi S and Luo F 2005 Water environmental degradation of the Heihe River Basin in arid northwestern China *Environ. Monit. Assess.* **108** 205-15
- [6] Wang Y, Feng Q, Chen L and Yu T 2013 Significance and effect of ecological rehabilitation project in inland river basins in northwest China *Environ. Manag.* **52** 209-20
- [7] Guo Q, Feng Q and Li J 2009 Environmental changes after ecological water conveyance in the lower reaches of Heihe River, northwest China *Environ. Geol.* **58** 1387-96
- [8] Zhang Y, Yu J, Qiao M and Yang H 2011 Effects of eco-water transfer on changes of vegetation in the lower Heihe River basin *J. Hydraul. Eng.* **42** 757-65 (in Chinese)
- [9] Cheng G, Li X, Zhao W, Xu Z, Feng Q, Xiao S and Xiao H 2014 Integrated study of the water–ecosystem–economy in the Heihe River Basin *Nat. Sci. Rev.* **1** 413-28
- [10] Wang G, Cheng G and Shen Y 2002 Dynamic tendency of arid oasis under the influence of water resources decrease—a case study of Ejina oasis in Heihe River Basin *J. Appl. Ecol.* **13** 564-8 (in Chinese)
- [11] Qin D, Zhao Z, Han L, Qian Y, Ou Z and Wang M 2012 Determination of groundwater

- recharge regime and flowpath in the Lower Heihe River Basin in an arid area of Northwest China by using environmental tracers: Implications for vegetation degradation in the Ejina Oasis *Appl. Geochem.* **27** 1133-45
- [12] Zhang Y, Huang Q, Zhang H, Sun X and Gao F 2010 Research on joint operation of cascade reservoirs in the upstream Heihe River *J. Hydroelec. Eng.* **29** 52-6 (in Chinese)
- [13] Chou C 2014 Application of set pair analysis-based similarity forecast model and wavelet denoising for runoff forecasting *Water* **6** 912-28
- [14] Zhao X and Chen X 2015 Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting *Water Resour. Manag.* **29** 2913-26
- [15] Qu Y, Zhou J, Liu F, Yang J and Li Y 2006 Medium-and long-term runoff forecasting based on improved elman neural network *J. China Hydrol.* **26** 45-50 (in Chinese)
- [16] Shoaib M, Shamseldin A, Melville B and Khan M 2016 A comparison between wavelet based static and dynamic neural network approaches for runoff prediction *J. Hydrol.* **535** 211-25
- [17] Chen R, Kang E, Yang J and Zhang J 2002 Nonlinear features of the runoff from mountain areas of the Heihe River, Qilian Mountains *J. Glaciol. Geocryol.* **24** 292-7 (in Chinese)
- [18] Demirel M, Venancio A and Kahya E 2009 Flow forecast by SWAT model and ANN in Pracana basin, Portugal *Adv. Eng. Softw.* **40** 467-73
- [19] Lü H, Hou T, Horton R, Zhu Y, Chen X, Jia Y, Wang W and Fu X 2013 The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method *J. Hydrol.* **480** 102-14
- [20] Suliman A, Jajarmizadeh M, Harun S and Darus I 2015 Comparison of semi-distributed, GIS-based hydrological models for the prediction of streamflow in a large catchment *Water Resour. Manag.* **29** 3095-110
- [21] Jiang X and Liu C 2004 Radial basis function networks based on wavelet analysis for the annual flow forecast *J. Appl. Sci.* **22** 411-4 (in Chinese)
- [22] Chu Y, Lan Y, Li X, Zhang Y, Hu X and Liu Q 2005 Models for long-term forecast of annual runoff at Yinluoxia station of Heihe River *J. Des. Res.* **25** 869-73 (in Chinese)
- [23] Zhang J and Ding H 2005 Application of the gray topological method in predicting runoff volume at the debouchure of the Heihe River *Arid Land Geogr.* **28** 751-5 (in Chinese)
- [24] Hou H, Yang L, Li F and Wang L 2010 Annual runoff forecast of the mainstream of Heihe river based on time series analysis method *Yellow River* **32** 49-53 (in Chinese)
- [25] Li J, Zan M and Li B 2014 Grey prediction of out-mountainous annual runoff of Heihe River based on R/S analysis *Areal Res. Dev.* **33** 127-31 (in Chinese)
- [26] Shi Y, Shen Y, Kang E, Li D, Ding Y, Zhang G and Hu R 2007 Recent and future climate change in northwest China *Climatic Change* **80** 379-93
- [27] Wang L and Zhang X 2010 Effect of the recent climate change on water resource in Heihe River basin *J. Arid Land Res. Environ.* **24** 60-65 (in Chinese)
- [28] Yue S and Wang C 2004 The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series *Water Res. Manag.* **18** 201-8
- [29] Tabari H, Marofi S, Aeini A, Talaei P and Mohammadi K 2011 Trend analysis of reference evapotranspiration in the western half of Iran *Agr. Forest Meteorol.* **151** 128-36
- [30] Chen Y, Lu B, Zhang Y and Zhou X 2011 Improvement of sequential cluster analysis method for extracting turning point of time series *J. China Hydrol.* **31** 41-4 (in Chinese)
- [31] Sang Y, Wang Z and Liu C 2012 Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis *J. Hydrol.* **424** 154-64
- [32] Tang C, Guo L and Wang X 2007 Application of periodic mean superposition in the middle-to-long term hydrological prediction in Cuntan station *China Rur. Water Hydrop.* **8** 15-7 (in Chinese)
- [33] Kantelhardt J, Bunde E, Rybski D, Braun P, Bunde A and Havlin S 2006 Long-term persistence and multifractality of precipitation and river runoff records *J. Geophys. Res-Atmospheres* **111** D1

- [34] Jiang T and Deng 2004 Some problems in estimating a Hurst exponent-a case study of applicatings to climatic change *Scient. Geog. Sin.* **24** 177-82 (in Chinese)
- [35] Hurst H 1951 Long-term storage capacity of reservoirs *T. Am. Soc. Civil Eng.* **116** 770–99
- [36] Xie Y, Huang Q, Chang J, Liu S and Wang Y 2016 Period analysis of hydrologic series through moving-window correlation analysis method *J. Hydrol.* **538** 278-92
- [37] Yang T, Yu P and Chen C 2005 Long-term runoff forecasting by combining hydrological models and meteorological records *Hydrol. Process.* **19** 1967-81
- [38] Zhang K, Yang Y, Qin Z and Meng Z 2013 Analysis and prediction of annual runoff for Fen river basin *Res. J. Appl. Sci. Eng. Technol.* **6** 872-7
- [39] Wang J and Meng J 2008 Characteristics and tendencies of annual runoff variations in the Heihe River Basin during the past 60 years *Scient. Geog. Sin.* **28** 83-8 (in Chinese)
- [40] Huang Z, Wang C and Ma J 2002 Adaptability of Pearson III type distribution and analysis for the error of hydrologic design value *Hydrology* **22** 21-4 (in Chinese)