Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2018.23201

129 (2018) 332–336
October

Study on the periodic fluctuations of runoff with multi-time scales based on set pair analysis

Yan Ye[a], Jinping Zhang[b,*], Jiachun Yang[c], Jian Li[b,*]

[a]*College of Resources and Environment, Southwest University, Chongqing 400735, China; email: yye_china@126.com*
[b]*Institute of Water Resources and Environment, Zhengzhou University, Zhengzhou 450001, China, Tel. +86-371-60119629; emails: jinping2000_zh@163.com (J. Zhang), jianli@zzu.edu.cn (J. Li)*
[c]*Dehong Branch of Hydrology and Water Resources Bureau of Yunnan Province, Mangshi 678400, China; email: 1211165030@qq.com*

Received 7 August 2017; Accepted 22 August 2018

ABSTRACT

Knowledge of the complex characteristics of runoff is required to regulate water resources. To determine the complex characteristics of runoff, the natural annual runoff data from 1956 to 2005, taken from the Zhangjiashan hydrological station on the Jinghe River, China, were decomposed into multiple time scales using the empirical mode decomposition method. The results show that the natural runoff of Jinghe River has complex periodic fluctuations with multiple time scales; the short-, middle-, and long-term periodic fluctuations have periods of 2–4, 4–8, and 11–13 years, respectively. These fluctuations are consistent with the periodic variations of El Niño, air–sea interactions, and solar activity. A set pair analysis (SPA) shows that the relationships between the natural runoff of Jinghe River and its intrinsic mode function components predominantly fall into the "identity" aspect for the short-period fluctuations, the "contrary" aspect for middle-period fluctuations, and the "discrepancy" aspect for long-period fluctuations. Moreover, the SPA reveals that the short-period runoff fluctuations can adequately reflect the average state of the Jinghe River.

Keywords: Empirical mode decomposition; Multiple time scales; Periodic fluctuation; Runoff; Set pair analysis

1. Introduction

Runoff fluctuations are affected by factors including hydrological and meteorological factors, underlying surface conditions, and human activities [1,2], and therefore exhibit highly complex, random, and nonlinear characteristics [3] over long timescales. Laws that define runoff fluctuations and trend characteristics are key to improving runoff prediction accuracy [4,5] and are crucial in guiding regional water resource management and in the efficient protection and utilization of water resources.

Although considerable challenges exist, many researchers have used analytical methods for runoff predictions. These approaches can be divided into two categories: traditional statistical methods and artificial intelligence methods. Traditional statistical methods include hydrology techniques such as Mann-Kendall trend tests [6], regression analyses, and the exponential smoothing method [7,8]. Traditional statistical methods are simple and direct, but their availability, accuracy, and representativeness are dependent on the historical data. Owing to the nonlinear and complex characteristics of runoff fluctuations, artificial intelligence methods, including artificial neural networks (ANNs) [9,10], least squares support vector machines (LSSVM) [11], support vector machines (SVM) [12], and wavelet analyses [13], provide useful alternatives to conventional statistical techniques. However, ANNs

* Corresponding authors.

Presented at 2017 Qingdao International Water Congress, June 27–30, 2017, Qingdao, China.

are restricted by slow learning, local convergence, and poor generalization ability. LSSVM can achieve a global optimum by solving a quadratic optimization problem, and must only solve a set of linear, rather than quadratic, equations [14]. SVM can be used for small samples because SVM are characterized by non-linear kernels, high generalization ability, and sparse solutions. The wavelet analysis method is more commonly used for hydrological time series [15,16]. This method finds the variation characteristics and influencing factors for runoff data over multiple timescales, improving the runoff forecast accuracy and water regulation process [17,18].

Although the wavelet analysis method provides high-resolution results in both frequency and time domains, many false harmonics still exist. Moreover, the selection of wavelet base functions has a significant impact on the results [19–21]. Huang et al. [22,23] proposed a new signal analysis method called empirical mode decomposition (EMD). EMD is more effective than the wavelet analysis method, especially for local frequency and time domains [24]. EMD has successfully been applied to many hydrological time series.

Uncertainty contributes to the complexity of runoff analyses. In 1989, Zhao and Xuan [25] proposed the set pair analysis (SPA) method to measure this uncertainty. SPA explains the uncertainty relationships between variables from three aspects: "identity," "discrepancy," and "contrary" [26]. The SPA method gradually is widely used in the field of hydrology. Many scholars have used the SPA method to explain, for example, variations in annual runoff [27], the relationship between flood peak and volume [28], etc., showing that the SPA method is effective, objective, and reasonable in hydrologic uncertainty correlation analyses.

The main objective of this study was to apply EMD to the annual runoff data from 1956 to 2005 from the Zhangjiashan hydrological station (ZHS) on the Jinghe River in China. EMD decomposed the data into multiple temporal scales to obtain fluctuation components. The SPA method was then applied to study the uncertainty of the fluctuation components for different periods.

2. Data and methods

2.1. EMD method

As proposed by Huang et al. [22], the intrinsic mode functions (IMFs) must satisfy the following two conditions to obtain a meaningful instantaneous frequency: (1) over the whole data range, the number of local extrema and the number of zero-crossings must be equal, or at most have a difference of one; and (2) at any point, the mean value of the upper envelope formed by all local maxima, and the lower envelope formed by all local minima, must be zero.

The key step of EMD is to extract the IMF from the given time series $x(t)$. First, the upper envelope and the lower envelope are constructed using identified local maxima and local minima with a cubic spline interpolation application. Then, the mean value of the two envelopes is calculated as m_1. A new time series with the lower frequency removed is achieved [22]:

$$h_1(t) = x(t) - m_1 \tag{1}$$

Usually, $h_1(t)$ is not an expected IMF, so this shifting process must be repeated k times until the first IMF component, $c_1(t) = h_{1k}(t)$, is achieved from the data $x(t)$. The residue $r_1 = x(t) - c_1$ will be decomposed further until it satisfies the stopping standard deviation (SD). In this way, the IMF modes and one residue r_n are obtained. The original series $x(t)$ is then rewritten as follows:

$$x(t) = \sum_{i=1}^{n} c_i + r_n \tag{2}$$

The SD criterion is defined as follows:

$$\mathrm{SD} = \sum_{k=1}^{T} \left(\left| h_{1(k-1)}(t) - h_{1k}(t) \right|^2 / h_{1(k-1)}(t)^2 \right) \tag{3}$$

where T is the length of the time series and the SD value is between approximately 0.2 and 0.3.

2.2. Set pair analysis

A set pair $H(A,B)$ refers to a couple that consists of two interrelated sets: A and B. The properties of these two sets include identity degree, discrepancy degree, and contrary degree. Additionally, the total number of characteristics in H is assumed to be N. The amounts of identical, discrepant, and contrary characteristics are then assumed to be S, F, and P, respectively. Based on the definition of SPA, the relationship between S, F, P, and N is $N = S + F + P$. The calculation algorithm for the connection degree is then [29] expressed as follows:

$$\mu = S/N + (F/N)i + (P/N)j \tag{4}$$

where μ is the connection degree of the set pair, and i and j are the coefficients of the discrepancy degree F/N and of the contrary P/N, respectively. S/N denotes the identity degree, i is an uncertain value between –1 and 1 (i.e., $i \in [-1,1]$), and j is specified as –1.

Given $a = S/N$, $b = F/N$, and $c = P/N$, Eq. (4) can be rewritten as follows:

$$\mu = a + bi + cj \tag{5}$$

In general, the steps of an SPA are as follows: (1) to construct two sets A and B and one set pair $H(A,B)$; (2) to quantify each element of A and B with a certain classification standard; (3) to compare these quantified values with each other and calculate the values of S, F, and P; (4) to choose an appropriate value of i to obtain a, b, c, and μ.

2.3. EMD and SPA analysis

The natural runoff data series $Y = (y_1, y_2, y_3, \ldots, y_n)$ is denoted as set A, its IMF components $X_m = (x_{m1}, x_{m2}, x_{m3}, \ldots, x_{mn})$ are specified as set $B_m (m = 1, 2, 3, \ldots M)$, and M is the quantity of IMFs. The set pairs are then defined as (A, B_1),

(A, B_2), …, (A, B_M), and are formed by the decomposed IMF components and the natural runoff series. The mean SD method is adopted to classify these IMF components into three states: "rich" (I), "normal" (II), and "poor" (III). EX is the mean value of a set in a set pair, and d is the SD, so the corresponding classified ranges are, respectively, described as ($-\infty$, $EX - 0.5d$), [$EX - 0.5d$, $EX + 0.5d$], and ($EX + 0.5d$, $+\infty$).

If the classified results of the set pairs (A, B_1), (A, B_2), …, (A, B_M) are in same state (i.e., rich [I], normal [II], or poor [III]), they are called identical, and S is then the amount of identical characteristics. If the classified results of the set pairs (A, B_1), (A, B_2), …, (A,B_M) are in different states (i.e., rich [I] and normal [II], or normal [II] and poor [III]), they are called discrepant, and F is then the amount of discrepant characteristics. If the classified results of the set pairs (A, B_1), (A, B_2), …, (A, B_M) are in opposing states (i.e., rich [I] and poor [III]), they are called contrary, and P is then the amount of contrary characteristics.

3. Application

3.1. Data series

Jinghe River, the largest tributary of the Weihe River (one of the tributaries of the Yellow River in China), originates from Liupanshan Mountain in the Ningxia Hui Autonomous Region of China. It is 451 km long with a catchment area of 45,400 km^2. Jinghe River is known for its violent floods and large sediment volumes. Its average annual runoff is 2.14 billion m^3 and its average annual sediment transport is 3.09 t. Therefore, it is a major source of floodwater and sediments for the Weihe River and the Yellow River. The ZHS, the main control station covering the lower reaches of Jinghe River, is located at the river mouth where it flows into the Weihe River. Fig. 1 shows the natural runoff data series from ZHS between 1956 and 2005.

3.2. Multiple timescale analyses of natural runoff

Using the EMD method, we analyzed the natural annual runoff for multiple timescales with an evaluated SD value of 0.25. Border issues were addressed using the boundary extension method. The decomposed results obtained for rainfall using the EMD method, as shown in Figs. 2–5, include three IMF components (shown in Figs. 2–4) and one residue (Fig. 5). These results show that the fluctuation period of the first IMF (IMF1) component is short, the IMF2 component shows a middle period, and the IMF3 component

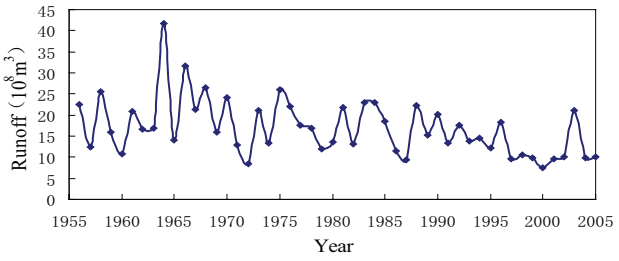

Fig. 1. Natural runoff data from the Zhangjiashan hydrological station (ZHS).

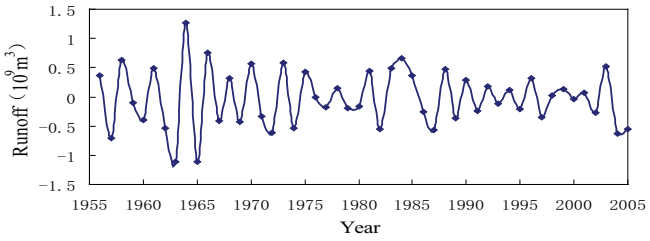

Fig. 2. The first IMF component (IMF1) for the runoff data.

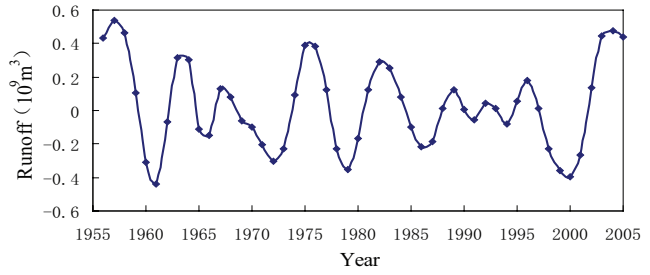

Fig. 3. The second IMF component (IMF2) for the runoff data.

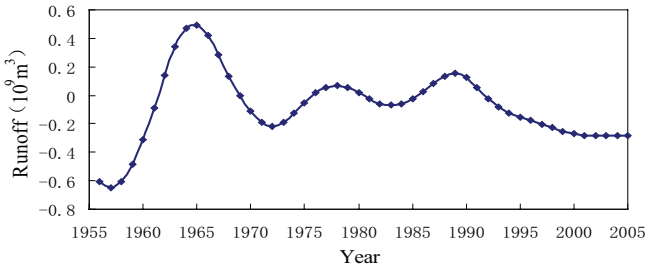

Fig. 4. The third IMF component (IMF3) for the runoff data.

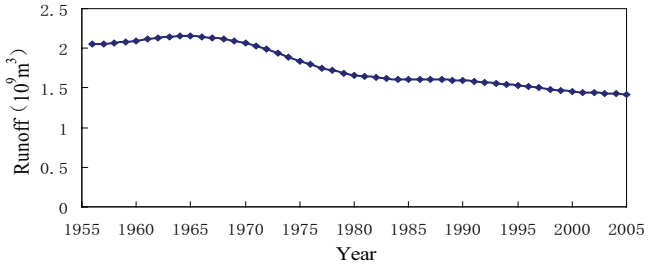

Fig. 5. The residue component for the runoff data.

shows a long period. The residue shows no obvious periodic fluctuation.

A quasi-periodic fluctuation of 2–4 years is present in the IMF1 component. The largest fluctuations appeared in the mid-1960s, but from the 1970s to the late 1980s, the fluctuations were relatively stable, and exhibited little change from the beginning of the 1990s.

A quasi-periodic fluctuation of 4–8 years is clearly present in the IMF2 component. The fluctuations can be clearly seen over the majority of the observation period. Larger fluctuations occurred in the late 1950s, the mid-1970s, the mid-1980s

Table 1
Classification results for annual runoff and its IMF components

Classification	Y	X_1	X_2	X_3
Rich (I)	(–∞, 13.64]	(–∞, –0.27]	(–∞, –0.10]	(–∞, –0.20]
Normal (II)	(13.64, 20.25)	(–0.27, 0.22)	(–0.10, 0.16)	(–0.20, 0.05)
Poor (III)	[20.25, +∞)	[0.22, +∞)	[0.16, +∞)	[0.05, +∞)

Table 2
Values of S, F, and P for each set pair

Set pairs	S	F	P	N
AB_1	31	18	1	50
AB_2	28	16	6	50
AB_3	22	23	5	50

Table 3
SPA of annual runoff for multiple time scales

Set pairs	a	b	c	μ
AB_1	0.62	0.36	0.02	0.82
AB_2	0.56	0.32	0.12	0.84
AB_3	0.44	0.46	0.10	0.77

and in the early 21st century, compared with smaller fluctuations observed in the 1990s.

A quasi-periodic fluctuation of 11–13 years is present in the IMF3 component. The largest fluctuation appeared in the mid-1960s; subsequently, the fluctuation size began to decrease. From the 1990s, the fluctuations decreased further.

The residue component indicates overall variations in the annual runoff. The results show that the runoff from Jinghe River has decreased since the mid-1960s.

3.3. SPA of annual runoff for multiple timescales

The classification results for the Jinghe River annual runoff $Y = (y_1, y_2, y_3, \ldots, y_{50})$ and its IMF components $X_m = (x_{m1}, x_{m2}, x_{m3}, \ldots, x_{m50})$ (m = 1, 2, 3) are shown in Table 1.

The set pairs (A, B_1), (A, B_2), and (A, B_3) are formed by three decomposed IMF components and the natural runoff series. The values of S, F, and P are shown in Table 2.

Specifying i as 0.5, the identity degree a, the discrepancy degree b, the contrary degree c, and the connection degree μ are shown in Table 3.

Tables 2 and 3 show the following: (1) As the fluctuation period increases, the identity degrees of the natural runoff and IMF components decrease. For the identity degree, a maximum of 0.62 is seen for the short-period fluctuations. (2) Alongside the increase in the fluctuation period, the discrepancy degrees for the natural runoff and IMF components decrease initially before increasing. The largest discrepancy degree (0.46) exists in the long-period fluctuations. (3) Alongside the increase in the fluctuation period, the contrary degrees for the natural runoff and IMF components initially increase before decreasing. For the contrary degree, a maximum of 0.12 is seen for the middle-period fluctuations. (4) The connection degree initially increases and then decreases. For the middle-period fluctuations, the maximum connection degree increases to 0.84. This is consistent with the contrary degree.

4. Discussion and conclusions

The EMD method was used to determine the complex periodic fluctuations over multiple timescales for the natural runoff from Jinghe River from 1956 to 2005. Quasi-periodic fluctuations of 2–4, 4–8, and 11–13 years were found for the short-, middle-, and long-period fluctuations, respectively. The overall variation in the runoff shows a gradual decrease. Some researchers [30,31] have found quasi-periodic fluctuations of 3.5 years and 4–8 years in the runoff. These periodic variation characteristics are almost identical to the periods of the El Niño-Southern Oscillation phenomenon and of sunspot activity [31], which illustrates that climate change has a major influence on local runoff variation characteristics. Therefore, the fluctuation periods of the IMF1 and IMF2 components show marked similarities to the period of the El Niño phenomenon. This demonstrates a close relationship between El Niño and the runoff of Jinghe River. Mei et al. [32] concluded that the daily values of solar radio flux and sunspot areas for the time interval from February 1, 1947 to September 30, 2016, both solar radio flux and sunspot areas are governed by the domain 11-year solar cycle. Solar magnetic activity indicators, from the lower atmosphere (photosphere and chromosphere) to the upper atmosphere (transition region and corona), are coupled in various styles of dynamical processes operating in the solar dynamo. Moreover, the IMF3 component may be related to air–sea interaction and long-period solar activity.

The SPA of the annual runoff for multiple timescales shows that the relationships between the natural runoff of Jinghe River and its IMF components predominantly fall into the "identity" aspect for the short-period fluctuations, the "contrary" aspect for middle-period fluctuations, and the "discrepancy" aspect for long-period fluctuations. These results also reveal that short-period runoff changes can adequately reflect the average state of Jinghe River. Therefore, the short-term forecast accuracy for runoff changes should be enhanced to aid regulation and planning for Jinghe River's water resources.

Due to limited data, only runoff is analyzed, which is the shortcoming of the paper; however, the obvious periodicity and correlation have been analyzed through EMD-SPA method. In the later stage, with the acquisition of more hydrological and meteorological data, the change rule of the hydrological and meteorological elements and their correlations will be further revealed.

Acknowledgment

This research was supported by the National Key R&D Program of China (Grant No. 2018YFC0406501), Outstanding Young Talent Research Fund of Zhengzhou University (Grant No. 1521323002), Program for Innovative Talents (in Science and Technology) at University of Henan Province (Grant No. 18HASTIT014), Foundation for University Key Teacher by Henan Province of China in 2017, and also by "Fundamental Research Funds for the Central Universities" (SWU115081).

References

[1] J.X. Chang, H.X. Zhang, Y.M. Wang, Y.L. Zhu, Assessing the impact of climate variability and human activities on streamflow variation, Hydrol. Earth Syst. Sci., 20 (2016) 1547–1560.
[2] J.T. Barge, H.O. Sharif, An ensemble empirical mode decomposition, self-organizing map, and linear genetic programming approach for forecasting River streamflow, Water, 8 (2016) 247.
[3] F.H.S. Chiew, N.J. Potter, J. Vaze, C. Petheram, L. Zhang, J. Teng, D.A. Post, Observed hydrologic non-stationarity in far south-eastern Australia: implications for modelling and prediction, Stochastic. Environ. Res. Risk Assess., 28 (2014) 3–15.
[4] S.H.W. Wang, B.J. Fu, S.H.L. Piao, Y.H. Lü, P. Ciais, X.M. Feng, Y.F. Wang, Reduced sediment transport in the Yellow River due to anthropogenic changes, Nat. Geosci., 9 (2015) 1–5.
[5] L. Chen, Y. Zhang, J. Zhou, V.P. Singh, S. Guo, J. Zhange, Real-time error correction method combined with combination flood forecasting technique for improving the accuracy of flood forecasting, J. Hydrol., 521 (2015) 157–169.
[6] E. Stonevicius, G. Valiuskevicius, E. Rimkus, J. Kazys, Climate induced changes of Lithuanian rivers runoff in 1960–2009, Water Resour., 14 (2014) 592–603.
[7] S.H.L. Lu, D.L. Li, J. Wen, Analysis on periodic variations and abrupt change of air temperature over Qinghai-Xizang plateau under global warming, Plateau Meteorol., 29 (2010) 1378–1385.
[8] H.C. Lloyd, S.W. Tommy, Runoff forecasting for an asphalt plane by Artificial Neural Networks and comparisons with kinematic wave and autoregressive moving average models, J. Hydrol., 397 (2011) 191–201.
[9] D.P. Solomatine, K.N. Dulal, Model trees as an alternative to neural networks in rainfall-runoff modeling, Hydrol. Sci. J., 48 (2003) 399–411.
[10] R.B. Mohammad, Z. Zahra, K. Sungwon, Soft computing techniques for rainfall-runoff simulation: local non–parametric paradigm vs. model classification methods, Water Resour. Manage., 31 (2017) 3843–3865.
[11] O. Kisi, C. Ozkan, A new approach for modeling sediment-discharge relationship: local weighted linear regression, Water Resour. Manage., 31 (2017) 1–23.
[12] J.-H. Jeon, C.-G. Park, A. Bernard, Comparison of performance between genetic algorithm and SCE-UA for calibration of SCS-CN surface runoff simulation, Water, 6 (2014) 3433–3456.
[13] N. Vahid, H.B. Aida, A. Jan, G. Mekonnen, Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling, J. Hydrol., 407 (2013) 28–40.
[14] X. Zhao, X. Chen, Y. Xu, An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting, Water, 9 (2017) 153.
[15] R.H. Compagnucci, S.A. Blanco, M.A. Figliola, P.M. Jacovkis, Variability in subtropical Andean Argentinean Atuel river: a wavelet approach, Environmetrics, 11 (2015) 251–269.
[16] C. Gaucherel, Use of wavelet transform for temporal characterization of remote watersheds, J. Hydrol., 269 (2002) 101–121.
[17] C.H.M. Liu, L. Cheng, Analysis on runoff series with special reference to drying up courses of Lower Huanghe River, J. Geogr. Sci., 55 (2000) 57–265.
[18] L. Chen, V.P. Singh, Entropy-based derivation of generalized distributions for hydrometeorological frequency analysis, J. Hydrol., 577 (2018) 699–712.
[19] A.H. Tewfiki, D. Sinaha, P. Jorgensen, On the optimal choice of a wavelet for signal representation, IEEE Trans. Inf. Theory, 38 (1992) 747–765.
[20] D. Labat, J. Ronchail, J. Callede, J.L. Guyot, Wavelet analysis of Amazon hydrological regime variability, Geophys. Res. Lett., 31 (2004) 33–45.
[21] J. Wang, J.J. Meng, Research on runoff variations based on wavelet analysis and wavelet neural network model: a case study of the Heihe River drainage basin, J. Geogr. Sci., 17 (2007) 327–338.
[22] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, CH.CH. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. London, Ser. A, 8 (1998) 903–995.
[23] N.E. Huang, Z. Shen, S.R. Long, A new view of nonlinear water waves: the Hilbert spectrum, Annu. Rev. Fluid Mech., 31 (1999) 417–457.
[24] M. Li, X. Wu, X. Liu, An improved EMD method for time-frequency feature extraction of telemetry vibration signal based on multi-scale median filtering, Circuits Syst. Signal Process., 34 (2015) 815–830.
[25] K.Q. Zhao, A.L. Xuan, Set pair theory-a new theory method of non-define and its applications, Syst. Eng., 14 (1989) 18–23.
[26] K.Q. Zhao, Set Pair Analysis and Its Elementary Application, Science and Technology Publishing House of Zhejiang, Hangzhou, 2000.
[27] Q. Zou, J.ZH. Zhou, CH. Zhou, L.X. Song, J. Guo, Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP, Stochastic Environ. Res. Risk Assess., 27 (2013) 525–546.
[28] B. Zhu, H.F. Wang, W.S.H. Wang, Y.Q. Li, Analysis of relation between flood peak and volume based on set pair analysis, J. Sichuan Univ., 39 (2007) 29–33.
[29] P. Feng, R.G. Han, Z.H.H. Ding, Multiple time-scale SPA analysis on uncertainty relationship between rivers' runoff time series, J. Sci. Eng., 17 (2009) 716–726.
[30] D.R. Zhang, C.H. Xue, Relationship between the El Nino and precipitation patterns in China since 1500 AD, Q. J. Appl. Meteorol., 5 (1994) 168–175.
[31] D. Liu, Q. Fu, T. Li, W. Li, Wavelet analysis of the complex precipitation series in the Northern Jiansanjiang Administration of the Heilongjiang land reclamation, China, J. Water Clim. Change, 7 (2016) 796–809.
[32] Y. Mei, H. Deng, F. Wang, On midrange periodicities in solar radio flux and sunspot areas, Astrophys. Space Sci., 363 (2018) 84.