

RESEARCH LETTER

Open Access



Correlation-aided method for identification and gradation of periodicities in hydrologic time series

Ping Xie^{1*}, Linqian Wu¹, Yan-Fang Sang^{2*} , Faith Ka Shun Chan^{3,4}, Jie Chen¹, Ziyi Wu¹ and Yaqing Li¹

Abstract

Identification of periodicities in hydrological time series and evaluation of their statistical significance are not only important for water-related studies, but also challenging issues due to the complex variability of hydrological processes. In this article, we develop a “Moving Correlation Coefficient Analysis” (MCCA) method for identifying periodicities of a time series. In the method, the correlation between the original time series and the periodic fluctuation is used as a criterion, aiming to seek out the periodic fluctuation that fits the original time series best, and to evaluate its statistical significance. Consequently, we take periodic components consisting of simple sinusoidal variation as an example, and do statistical experiments to verify the applicability and reliability of the developed method by considering various parameters changing. Three other methods commonly used, *harmonic analysis method* (HAM), *power spectrum method* (PSM) and *maximum entropy method* (MEM) are also applied for comparison. The results indicate that the efficiency of each method is positively connected to the length and amplitude of samples, but negatively correlated with the mean value, variation coefficient and length of periodicity, without relationship with the initial phase of periodicity. For those time series with higher noise component, the developed MCCA method performs best among the four methods. Results from the hydrological case studies in the Yangtze River basin further verify the better performances of the MCCA method compared to other three methods for the identification of periodicities in hydrologic time series.

Keywords: Periodicity, Correlation analysis, Significance evaluation, Hydrologic time series analysis

Introduction

Hydrological processes are influenced by both deterministic and stochastic factors (Mehdizadeh et al. 2017; Rios and de Mello 2013; Stojkovic et al. 2017) along with uncertainty (Coulbaly and Baldwin 2005; McCuen 2003; Sang et al. 2017). Some observed hydrological time series usually include deterministic components (as “signals”), such as periodic fluctuation of the water level (or stream-flow) of a river in the annual, interannual and larger

timescales. They also include random fluctuations, just as “noise” (Sang et al. 2009). Detecting, extracting and evaluating those “signals” with useful information can help us to identify the variability of hydrological process with physical causes, and dealing with stochastic modeling (Bordi et al. 2004; Rao et al. 1992).

Periodicity is an important type of hydrological signals, and it is mainly caused by the Earth revolution and rotation, geological processes, human activities and other physical factors (Hao et al. 2016; Kottegoda 1980). According to the number of periodic components, if there is a periodicity only at one frequency, it will be called simple periodicity, and periodicities at two or more frequencies are namely, compound periodicities (Siegel 1980). Also, there are more complex periodic variations like quasi-periodicity (Nigmatullin et al. 2014).

*Correspondence: pxie@whu.edu.cn; sangyf@igsnr.ac.cn

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

² Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Full list of author information is available at the end of the article

Periodicity-related research is mainly concerned with two problems, identification of the periodic component and evaluation of its statistical significance.

Several methods have been applied in identifying the hydrological periodicities. They originate from the spectral analysis in signal processing, dealing with problems of signals and noise (Zhou and Sornette 2011). The harmonic analysis (Nuttle 1997; Steele 1982), as to perform the classical spectral analysis, was early probed to interpret the periodicity in time series. Developed from the Fourier analysis, the periodic component is represented by a set of sinusoidal functions which is an accurate mathematical concept, but it cannot avoid computational burden. The *fast Fourier transform* (FFT) (Cooley and Tukey 1965), aiming at a faster Fourier transform, is a relatively more powerful approach improving the time series to transfer from time domain to frequency domain. Periodogram investigates the periodicity by estimating the power spectral density (PSD) using the time series directly (Schuster 1898; Thomson 1982). Though there are attempts of smoothing the periodogram (Bartlett 1950; Kay 1988; Welch 1967), the incompatibility between high spectral resolution and low ‘*power leakage*’ still limits its application to yield the true spectrum of time series. Correlogram (Ghil and Taricco 1997), as another most commonly used power spectrum estimation based on autocorrelation function, has the same defects as those exposed by periodogram. All these conventional methods are limited by short sample length. Modern techniques of identifying and extracting periodic components have developed and some have been applied in the field of hydrological science. Continuous spectral analysis like *maximum entropy method* (MEM) (Burg 1975) was developed to overcome the preceding drawbacks. With high resolution and sharp peaks for shorter data length (Cao et al. 1997; Kay 1988), the MEM has been widely used, but its sensitivity to noise constrains its wide applications (Jaynes 1982). Apart from that, the algorithm is based upon assuming the data conform to the *AR* (autoregressive) model and determining the order by various criteria, such as the *final prediction error* criterion (FPE) (Akaike 1970), the *Akaike Information Criterion* (AIC) (Akaike 1974; Sakamoto and Kitagawa 1987) and the *Bayesian Information Criterion* (BIC) (Tamura et al. 1991). The choice of proper criteria must be treated with caution (Padmanabhan and Rao 1988), as a wrong order will cause potential influences on the accuracy of results.

Another primary problem in periodicity analysis is quantitatively assessing the statistical significance of the identified component. In recent decades, studies mainly focus on the improvement (Yuan et al. 2016), comparison (Yang 2015) and application of periodicity identification

methods (Stosic et al. 2016; Wu et al. 2016), but lesser focused on significance assessment. Popular methods of significance assessment are developed on statistical hypothesis tests, mainly as one-tailed tests (Siegel 1980), comparing the identified periodicity with a non-periodic component. They could only give qualitative evaluation like “*significant*” or “*not significant*” based on a certain significance level and statistical threshold. The periodic component being more significant outweighs other components in the whole series and has more contribution to the variability of hydrological process. Lacking precise classification of significance levels leads to insufficient understanding of the degree of periodicity variation, which is not favored to the assessment of impact and risk of potential consequences dominated by this periodic pattern. An intuitive index to reflect the significance of a periodic components is the amplitude, while the value of the amplitude varies theoretically from negative to positive infinity. Instead, the correlation coefficient (CC) changes within a certain range of -1 to 1 (McCuen 2003; Troch et al. 2013). And the correlation between the periodic component and the original series can generally represent the effect of this component on the whole original series. In such case, if the mathematical relationship between the amplitude and CC can be established, this can contribute to quantitative assessment of the significance of periodic components. The CC-aided idea has once been applied to jump points detection (Wu et al. 2019). As different variability types (like jump and periodicity) have completely different mathematical expressions, the application of CC-based method to the detection of periodicity still needs new derivation and demonstration.

Therefore, research on periodicities is still worth exploring. The main objective of this study is developing a new *moving correlation coefficient-based analysis* (MCCA) method for the identification of periodicities and evaluation of their significance levels with a more precise criterion. It is based on the correlations between the potential periodic component and the original time series. By deducing the relationship between the correlation coefficient and amplitude of periodicity mathematically, the MCCA method singles out the most probable periodicity by virtue of the correlation coefficient and characterizes the periodic component with necessary information like the cyclic period(s), the amplitude, the mean value of the observed data and other parameters and the significance level (Nuttle 1997). “*Methods*” section proposes the MCCA method through formula deduction, and gives the principle of periodicities identification and its significance gradation using correlation coefficient in detail. Besides, synthetic time series are used to verify the rationality and to investigate the

influence of several factors on the efficiency of MCCA method, with three other methods being compared. “Study area and data” section describes the annual runoff and precipitation data used in this study. The periodicities of runoff and precipitation in the Yangtze River basin are analyzed in “Results and discussion” section to further verify the MCCA method, and this manuscript will be ended by conclusion.

Methods

Relationship between the correlation coefficient and the half-amplitude of periodicity

To characterize the fluctuation degree of a periodic component using the correlation coefficient, a periodic process needs to be firstly constructed. Simple periodicity is of particular interest because of the simplicity to illustrate the rationale of the proposed method. Now taking the sinusoidal wave as an example, which is simple but general, we shall consider a time series $x(t)$ ($t = 1, 2, \dots, n$) be measured as

$$x(t) = A + B \sin\left(\frac{2\pi}{T}t + t_0\right) + \eta(t), \quad (1)$$

where T is the length of the periodicity, t_0 is the initial phase varying from 0 to 2π , $\eta(t)$ is a random residual, A is the mean value of time series $x(t)$, and B is the half-amplitude. If combining A and $\eta(t)$ as the random part of $x(t)$, denoted as $u(t)$, Eq. (1) can be expressed in linear superposition form:

$$x(t) = Bz(t) + u(t), \quad (2)$$

where $z(t) = \sin\left(\frac{2\pi}{T}t + t_0\right)$, $u(t) = A + \eta(t)$. The correlation coefficient for quantifying the relationship between the original time series $x(t)$ and the periodic component $y(t) = Bz(t)$ can be expressed as:

$$r = \frac{\sum_{t=1}^n (x(t) - \bar{x})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^n (x(t) - \bar{x})^2 \sum_{t=1}^n (y(t) - \bar{y})^2}}. \quad (3)$$

For specific half-amplitude B , periodicity length T and initial phase t_0 , the correlation coefficient (CC) in Eq. (3) can be rewritten as:

$$r = \frac{\sum_{t=1}^n (x(t) - \bar{x})(z(t) - \bar{z})}{\sqrt{\sum_{t=1}^n (x(t) - \bar{x})^2 \sum_{t=1}^n (z(t) - \bar{z})^2}}, \quad (4)$$

where $x(t)$ is the original hydrologic time series, $z(t)$ represents the periodic part, $\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t)$ and $\bar{z} = \frac{1}{n} \sum_{t=1}^n z(t)$ are the mean values of $x(t)$ and $z(t)$, respectively.

For hydrologic time series with an unknown periodicity, suppose a periodicity length T and an initial phase t_0 , when CC between the generated periodic component $z(t) = \sin\left(\frac{2\pi}{T}t + t_0\right)$ and the original time series $x(t)$ reaches its maximum, that is, the sinusoid comes closest to the real fluctuation of the periodicity in time series $x(t)$, and correspondingly the assumed periodicity length T and the initial phase t_0 are the best results expected. Finally, A and B could be obtained by the least square method:

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

$$\hat{A} = \bar{x} - \hat{B}\bar{z}. \quad (6)$$

Thus, Eq. (1) representing a simple periodicity can be determined as a result.

Since the amplitude of a periodic component reflects its significance and the correlation coefficient can quantify the significance level, the significance of a periodic component can be graded to different levels, if the relationship between the correlation coefficient and the amplitude is deduced. Substituting Eq. (5) into Eq. (3), then,

$$r = B \frac{\sigma_z}{\sigma_x}, \quad (7)$$

where σ_x and σ_z are the standard deviation of x_t and z_t , respectively. According to the theory of Stochastic Hydrology (Machiwal and Jha 2012; Sang et al. 2012), different components composed in hydrologic times series x_t conform to the linear superposition principle. Therefore, the random component and the periodic component are thought as independent, and σ_x^2 can be represented by the sum σ_z^2 and σ_u^2 as:

$$\sigma_x^2 = B^2 \sigma_z^2 + \sigma_u^2. \quad (8)$$

Substituting Eq. (8) into Eq. (7):

$$r^2 = \frac{1}{1 + \frac{\sigma_u^2}{B^2 \sigma_z^2}}, \quad (9)$$

where the standard deviation σ_z is influenced by the sample length n , the periodicity length T and the initial phase t_0 , which is expressed as:

$$\sigma_z^2 = \frac{1}{2} - \frac{\cos\left[\frac{2\pi}{T}(n+1) + 2t_0\right] \sin\frac{2\pi n}{T}}{2n \sin\frac{2\pi}{T}} - \frac{\left[1 - \cos\left(\frac{2\pi}{T}n + \frac{2\pi}{T} + 2t_0\right)\right] \left(1 - \cos\frac{2\pi n}{T}\right)}{2n^2 \left(1 - \cos\frac{2\pi}{T}\right)} \tag{10}$$

And σ_u is affected by the mean value \bar{u} and the variation coefficient Cv_u of the random component:

$$\sigma_u = \bar{u}Cv_u. \tag{11}$$

Given T and t_0 , the standard deviation σ_z^2 and σ_u^2 are known. Hence, the correlation coefficient r and the half-amplitude B show positive correlation with each other. The bigger absolute value of the correlation coefficient is, the bigger the amplitude of the periodic component is, which reflects more significant periodicity in the time series.

Correlation coefficient-based approach for the identification of periodicities

The specific steps of the identification of periodicity and its significance gradation by the proposed method are described as follows:

1. For the hydrologic time series $x(t)$ to be analyzed, construct a periodic component based on the sinusoidal function $z(t) = \sin\left(\frac{2\pi}{T}t + t_0\right)$;
2. Change the periodicity length T from 2 to $n/2$ by step l_1 , where n is the sample length. The initial phase is set to change from $-\pi$ to π by step l_2 , thus we get M sets of time series $z(t)$ ($M = \left(\frac{n/2 - 2}{l_1} + 1\right) \times \frac{2\pi}{l_2}$). Step length $l_1 = 1$ and $l_2 = 0.001\pi$ are usually set as defaults and will vary depending on the demanded accuracy.
3. Calculate the correlation coefficient r between $z(t)$ and $x(t)$ by Eq. (4). The periodicity length corresponding to the maximum absolute value denoted as $|r|_{\max}$ is the identification result.
4. Do the hypothesis test to evaluate the significance of the simulated periodic components (Xie et al. 2018). Given the significance level α and β , and $\alpha > \beta$, when $0 \leq |r| < r_\alpha$, the value of $|r|$ is not significant at level α and the null hypothesis that there is

no significant periodic component can be accepted; when $r_\alpha \leq |r| < r_\beta$, it indicates that $|r|$ is significant at level α but not at level β , then the significance of the periodic component in this interval is divided to “weak”. In the case where $|r|$ belongs to the range $r_\beta \leq |r| < 0.6$, it is categorized into moderate significance level. When $0.6 \leq |r| < 0.8$, the significance level is “strong”. Besides, when $0.8 \leq |r| \leq 1$, we use “dramatic” to describe the fact that the periodic component is the most significant. The CC thresholds for the significance gradation of periodicities are shown in Table 1.

5. When $x(t)$ contains multiple periodic components, loop step (1)–(4) for several times to find all significant periodic components in it. For round i , the identified periodic component z_i is removed by direct subtraction, and the left time series $x_i = x_{i-1} - z_i + \bar{x}_{i-1}$ ($i = 1, 2, 3, \dots, n$) is the new input series to identify other periodicities in it. The correlation coefficient between z_i and the original time series $x(t)$ is used to evaluate its significance level. The identification of periodicities can stop when no more significant periodicity can be found.

Verification of the proposed MCCA method

This section is subdivided into two parts. In the first part, we use the synthetic time series to validate the MCCA method, and in the second part we investigate the identification efficiency (IE) of the proposed MCCA method with several parameters’ changes.

Synthetic data analysis

Hydrologic time series are affected by various factors and contaminated with different kinds of noise, which is usually subject to the Pearson type III (PT-III) distribution (Singh 1998) in China. Therefore, the synthetic time series are generated by Monte Carlo method (Peres and Cancelliere 2016; Salas 1993) here by considering two parts: (1) the periodic component which need the parameters B , T and t_0 in the function $y(t) = B \sin\left(\frac{2\pi}{T}t + t_0\right)$; and (2) the stochastic component, which obeys the PT-III

Table 1 Thresholds of correlation coefficient r used for the gradation of significance level of the periodic component in hydrologic time series

Significance level	No	Weak	Moderate	Strong	Dramatic
Correlation coefficient	$0 \leq r < r_\alpha^a$	$r_\alpha \leq r < r_\beta^a$	$r_\beta \leq r < 0.6$	$0.6 \leq r < 0.8$	$0.8 \leq r \leq 1$

^a α and β are both significance levels and $\alpha > \beta$

distribution and is determined by the mean value \bar{u} the variation coefficient Cv_u and the skewness coefficient Cs_u .

The rationality of Eq. (9) needs to be confirmed first by the following simulated experiments. Statistical tests are conducted with 30 groups of half-amplitude increasing gradually. The procedures are explained as follows:

1. Generate 30 time series x_i with the sample length $n=100$, the mean value $\bar{u} = 100$, the variation coefficient $Cv_u = 0.2$ and the skewness coefficient $Cs_u = 0.4$. For the periodic component $B_i \sin \frac{\pi}{5}t$ ($B_i = i, i=1, 2... 30$), the periodicity length is set as $T=10$ and the initial phase $t_0 = 0$. With these parameters above, the standard deviation $\sigma_z^2 = 0.5$ and $\sigma_u^2 = 400$ can be determined by Eqs. (10) and (11), respectively.
2. Apply Eq. (4) to calculate the correlation coefficient r between $B_i \sin \frac{\pi}{5}t$ and x_i
3. Repeat each test for 10,000 times, then we get the series x_{ij} and the mean value $r_i = \frac{1}{10000} \sum_{j=1}^{10000} r_{ij}$ in each group, where $i = 1, 2, 3... 30, j = 1, 2, 3... 10,000$.

We use the significance levels $\alpha = 0.05$ and $\beta = 0.01$ in this paper, which are also widely used in hydrological time series analysis. When B is determined, we can get the theoretical correlation coefficient r_a by Eq. (9). Compare r_a with r_i by using the relative error $\delta = \frac{|r_i - r_a|}{r_a} \times 100(\%)$ as criterion. The experimental data are recorded in Table 2. It shows that among 30 groups of δ , 27 of them are within 1% and even the maximum value of δ is only 1.67%. The correlation coefficients got from the test and those from Eq. (9) are close to each other. It is thought that the results obtained from Eq. (9) are reliable, and the correlation coefficient can be used as an effective index to grade the significance levels of periodicities in hydrologic time series.

Then three sinusoidal functions and a random component are synthesized as the tested time series. This test is designed for two purposes: validating that the MCCA method can identify each periodic component and giving the correct significance gradation corresponding to the original setting. Parameters of the stochastic part $u(t)$ are the same as the previous statement, while the periodic component consisting of three true periodicities is set as

$$z_0(t) = 100 + 30 \sin \left(\frac{2\pi}{10}t + \frac{\pi}{2} \right) + 40 \sin \left(\frac{2\pi}{15}t + \frac{\pi}{6} \right) + 50 \sin \left(\frac{2\pi}{20}t + \frac{\pi}{3} \right), \tag{12}$$

and the synthetic time series $x_0(t) = u(t) + z_0(t), t \in [1, 100]$. For round i , the identified periodic component p_i is removed by direct subtraction and the left series $x_i = x_{i-1} - p_i + \bar{x}_{i-1}$ is the new input series to analyze the other periodicities of $x_0(t)$. We also define the relative error $\delta_i = \frac{|T_i - T'_i|}{T_i} \times 100(\%)$ ($i = 1, 2, 3, \dots, n$) to evaluate the accuracy of the results, where T is the theoretical value and T' is the calculated value.

Figure 1 illustrates the time-varying characteristics of the synthetic series as well as the input series and the periodic component in each round. It is shown in Fig. 1a that due to the synthesis of three periodic components and the addition of the random term, no obvious periodicity can be seen intuitively from the curve of the synthetic time series. After the MCCA processing, in Fig. 1b–d, each periodic component can be observed clearly. There are periodic variations of 20.4, 15.1 and 10, respectively, and the correlation coefficient r between p_i and x_0 grows with increasing amplitude. Compared with initial settings, the results in each round are close to the real one with small relative errors 2%, 1.3% and 0, and the accuracy is within the allowable range for the time interval of 1. Besides, it is obvious

Table 2 The theoretical value r_a , the calculated value r_i and the relative error δ (%) under different half-amplitudes B_i ($i = 1, 2, \dots, 30$)

B_i	r_a	r_i	δ (%) ^a	B_i	r_a	r_i	δ (%)	B_i	r_a	r_i	δ (%)
1	0.0349	0.0353	1.16	11	0.3637	0.3625	0.34	21	0.5983	0.5961	0.37
2	0.0713	0.0705	1.13	12	0.3926	0.3906	0.51	22	0.6157	0.6140	0.28
3	0.1064	0.1055	0.90	13	0.4187	0.4176	0.25	23	0.6324	0.6309	0.23
4	0.1407	0.1400	0.48	14	0.4447	0.4436	0.24	24	0.6493	0.6470	0.36
5	0.1770	0.1741	1.67	15	0.4706	0.4685	0.44	25	0.6637	0.6623	0.22
6	0.2072	0.2075	0.17	16	0.4954	0.4924	0.61	26	0.6782	0.6768	0.21
7	0.2419	0.2402	0.69	17	0.5173	0.5152	0.42	27	0.6933	0.6905	0.40
8	0.2724	0.2722	0.09	18	0.5385	0.5369	0.30	28	0.7064	0.7035	0.41
9	0.3037	0.3032	0.16	19	0.5598	0.5576	0.39	29	0.7189	0.7159	0.42
10	0.3350	0.3333	0.49	20	0.5796	0.5774	0.39	30	0.7295	0.7276	0.26

^a The relative error $\delta = \frac{|r_i - r_a|}{r_a} \times 100(\%)$

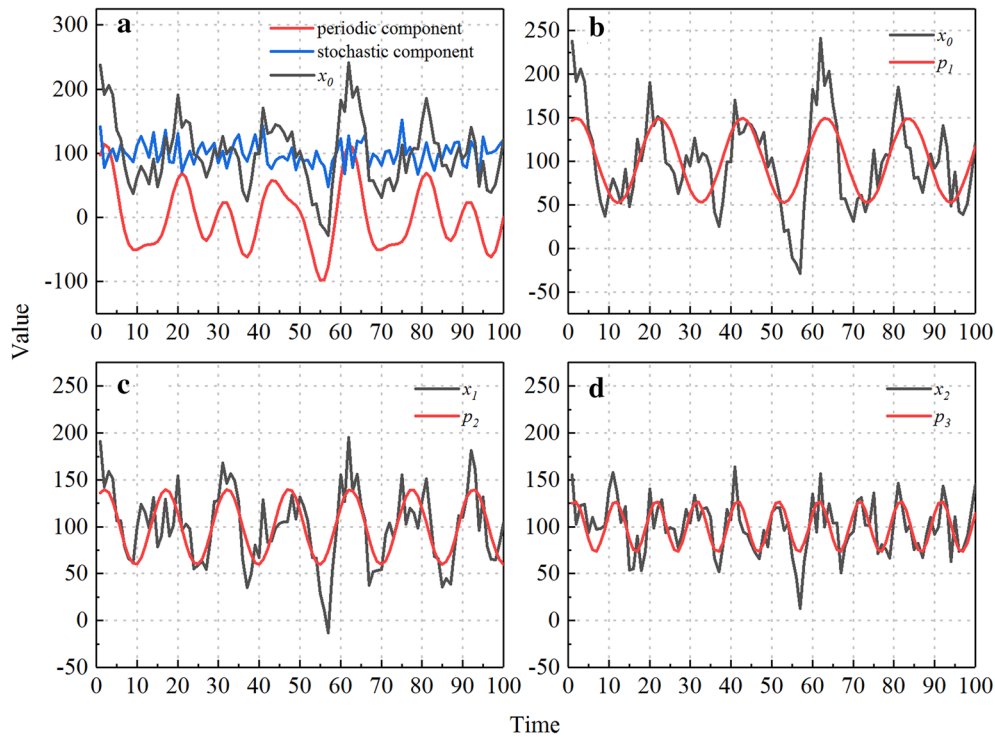


Fig. 1 The input time series and periodicity identification result in each round. **a** The original synthetic series x_0 mixed with periodic and stochastic components; **b** the input series x_0 and the periodic component p_1 identified in the first round; **c** the input series x_1 and the periodic component p_2 identified in the second round; **d** the input series x_2 and the periodic component p_3 identified in the third round

Table 3 The experimental data in each round, including the theoretical value T , the calculated value T' , the relative error δ (%), the theoretical half-amplitude B and the correlation coefficient r between x_0 and p_i ($i = 1, 2, 3$)

Round	T	T'	δ^a	B	r^b
1	20	20.4	2%	50	0.654
2	15	15.1	1.3%	40	0.513
3	10	10	0	30	0.406
4	–	3.4	–	–	0.105

^a The relative error $\delta_i = \frac{|T_i - T'_i|}{T_i} \times 100(\%)$ ($i = 1, 2, 3, \dots, n$)

^b The four thresholds of correlation coefficient are set as $r_{0.05} = 0.195$, $r_{0.01} = 0.254$, 0.6 and 0.8

the correlation coefficient $r = 0.105$ in the fourth round is less than the lower critical value $r_\alpha = 0.195$. That is the periodic component in this round is insignificant and the identification procedure can stop at round 4. The results are summarized in Table 3. It can be concluded that the MCCA method is able to detect and evaluate the periodicity in these synthetic time series.

Influences of several factors on the efficiency of MCCA method

Through the deduction of Eq. (9), it is known that the correlation coefficient between the original time series and the simulated periodic component may be affected by the following factors: the sample length n , the mean value \bar{u} and the coefficient of variation Cv_u of the stochastic component; the half-amplitude B , the periodicity length T and the initial phase t_0 of the periodic component. In this case, the change rules of the correlation coefficient and the effectiveness of the proposed method are further discussed. By varying the values of the above parameters, each test is correspondingly divided into several groups and each group is repeated for 100 times. The parameters are outlined in Table 4. Three other frequently used methods, power spectrum method (PSM), harmonic analysis method (HAM) and maximum entropy method (MEM) are also tested for comparison.

Denote T as the theoretical value and T' as the identified value of the periodicity length, then the allowable error of the method can be expressed as $\Delta T = |T - T'| = 1$, where “1” is the unit time interval of the data. If there are totally N groups of simulated time

Table 4 Groups of six parameters, the sample length n , the mean value \bar{u} and the variation coefficient Cv_u of the stochastic component, the half-amplitude B , the periodicity length T and the initial phase t_0 of the periodic component, used in statistical tests for investigating their influences on the identification of periodicities

Test	N	\bar{u}	Cv_u	B	T	t_0
1	100, 150, 200, 250, 300, 350, 400, 450, 500	100	0.2	30	25	0
2	200	100, 150, 200, 250, 300, 350, 400, 450, 500	0.2	30	25	0
3	200	100	0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45	30	25	0
4	200	100	0.2	1.0, 1.5, 2.0, 2.5, 3.0 ^a	25	0
5	200	100	0.2	30	5, 10, 15, 20, 25, 30	0
6	200	100	0.2	30	300	0, $\pm 0.25\pi$, $\pm 0.5\pi$, $\pm 0.75\pi$, π^b

^a The half-amplitude should be comparable with the stochastic fluctuation. Therefore, let $B = A \times M$, where $A = |\text{quantile}(x(t), 3/4) - \text{quantile}(x(t), 1/4)|$. The quantile $(x(t), 1/4)$ and $\text{quantile}(x(t), 3/4)$ here are the upper and lower quartiles of time series $x(t)$, respectively. M is a coefficient representing multiple relation between B and A

^b As $t_0 = \pi$ and $t_0 = -\pi$ represent the same phase for the sinusoidal function, only $t_0 = \pi$ is tested

Table 5 The periodicity identification efficiency (IE) (%) of four methods power spectrum method (PSM), harmonic analysis method (HAM), maximum entropy method (MEM) and moving correlation coefficient-based analysis (MCCA) under different sample lengths

Sample length	PSM	MEM	HAM	MCCA
100	0	30	100	98
150	0	64	99	100
200	49	78	100	100
250	100	86	100	100
300	92	91	100	100
350	96	95	100	100
400	34	96	100	100
450	100	100	100	100
500	100	100	100	100

series and M groups of them are identified to have the results within the allowable error, the identification efficiency (IE) is defined as $IE = \frac{M}{N} \times 100\%$.

(1) Sample length

For a certain periodicity length T , with a larger sample size, the number of a complete periodic fluctuation will be larger as well; therefore, the identification will be more effective as the periodic component weights more in the whole series. It can be seen from Table 5 that as the sample length grows from 100 to 400, the IE of MEM increases as expected and for sample length larger than 450, the IE can reach 100%, which shows that MEM is

affected by sample length. The IE of PSM is also affected by the sample length, but the linear rule is not obvious due to the impact of the maximum time lag m (Wang and Me 1990). The HAM and MCCA method have higher IE for different sample lengths, which shows the reliability of the MCCA method and its stability with sample length changing.

(2) Mean value

It is obvious in Fig. 2a that the correlation coefficient decreases with the increase of the mean value, and the IE values of the four methods also shows a descending trend. The IE of MEM and PSM drop greatly from 95 to 10% and from 85 to 5%, respectively, when \bar{u} is larger than 150. However, the IE values of the HAM and MCCA method is more stable, but when the mean value is larger than 300, the IE of these two methods start to decrease and it can be noticed that the correlation coefficient is also smaller than the critical value. When $\bar{u} = 500$, the IE of HAM is 10% lower than that of MCCA.

(3) Coefficient of variation

The PSM and MEM methods both shift down significantly with the increase of the coefficient of variation Cv_u (Fig. 2b). When Cv_u is larger than 0.2, the IE of PSM is less than 50%. By contrast, the MCCA and HAM methods show good stability and the MCCA method is the best among four methods. After $Cv_u > 0.25$, the IEs of the four methods all show a downward trend, especially the PSM and MEM drop significantly to lower than 10%.

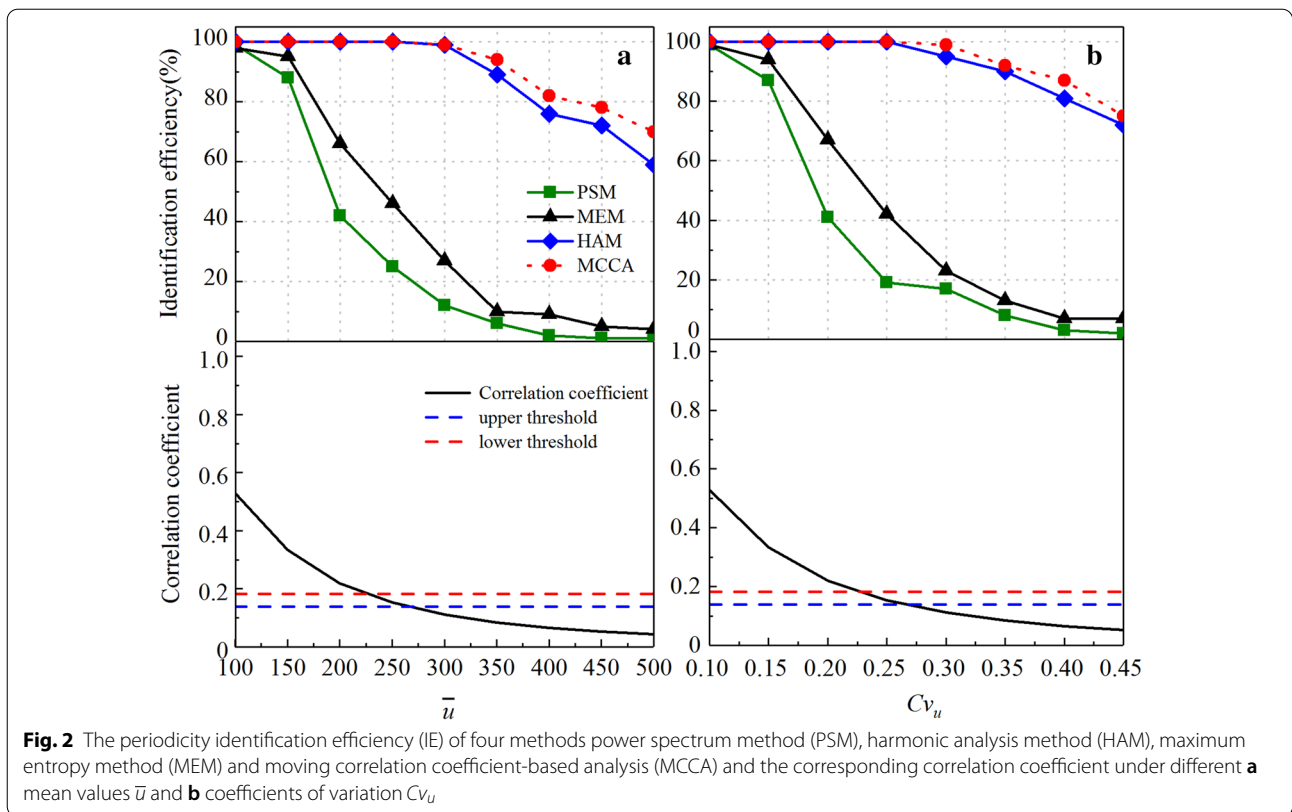


Fig. 2 The periodicity identification efficiency (IE) of four methods power spectrum method (PSM), harmonic analysis method (HAM), maximum entropy method (MEM) and moving correlation coefficient-based analysis (MCCA) and the corresponding correlation coefficient under different **a** mean values \bar{u} and **b** coefficients of variation Cv_u

By comparing Fig. 2a, b, it is obvious that these two figures show a consistent pattern of change, which is the result that the mean value \bar{u} and the coefficient of variation Cv_u both have impact on the dispersion degree of the time series. The more obvious the random fluctuation is, the less significant the periodic component is, which will cause difficulty in the identification and lead to low IE of the methods used.

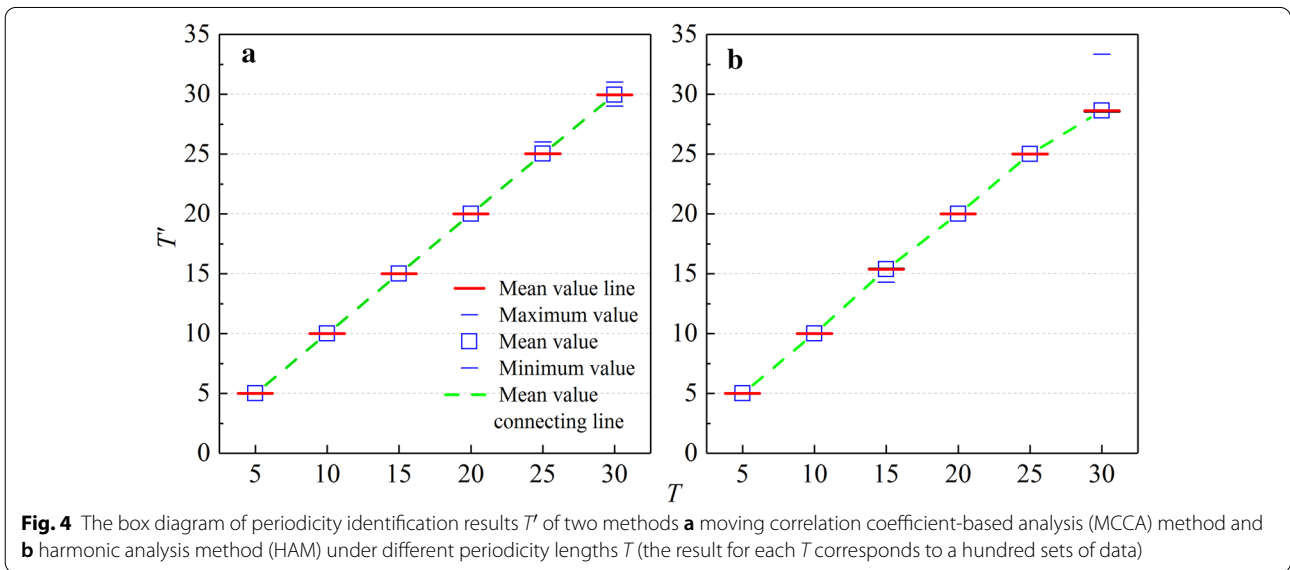
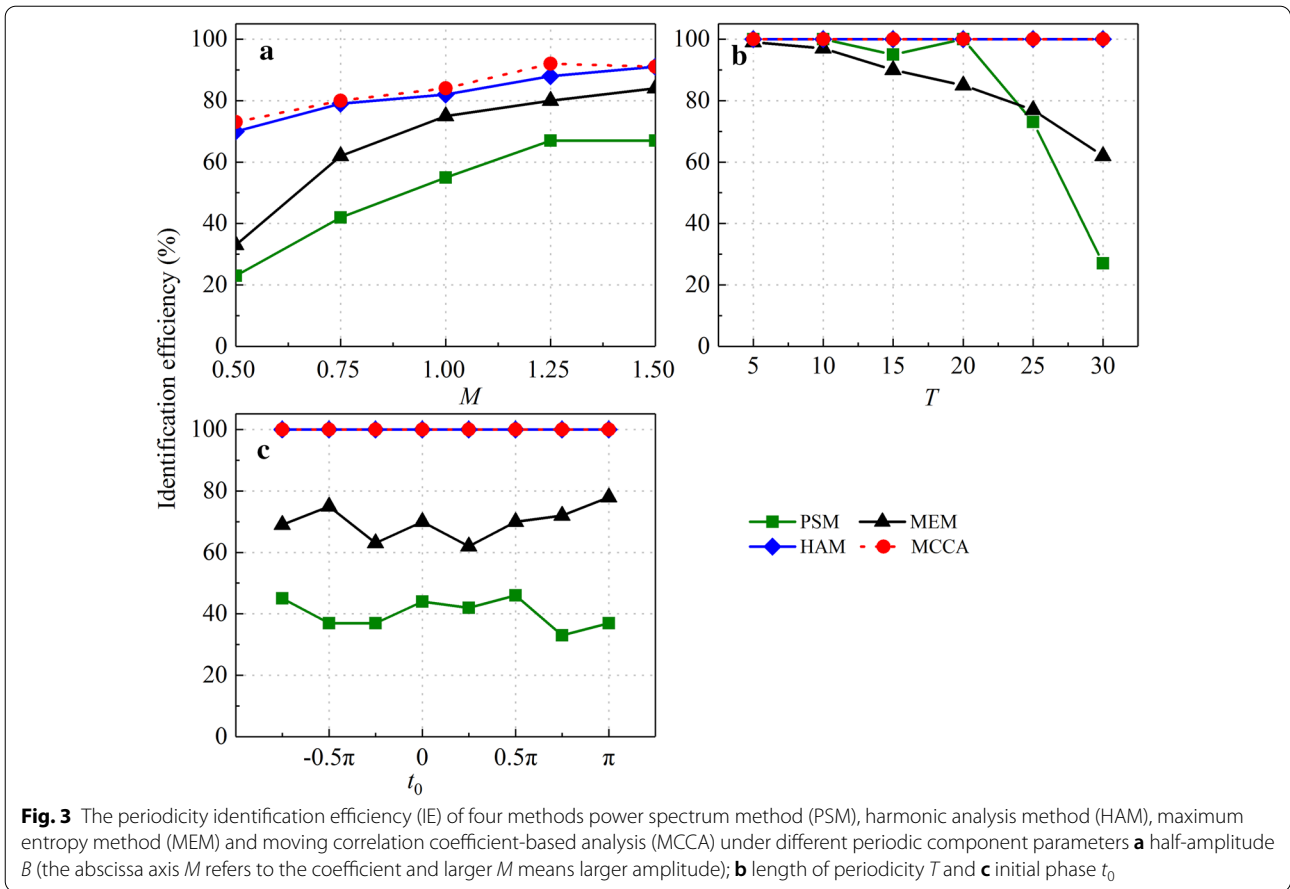
(4) Amplitude

It can be seen from Fig. 3a that the IE of each method increases with the increase of half-amplitude. PSM is the worst among the four methods. MEM has low IE when the half-amplitude is small, but it gets better with the half-amplitude increasing to 1.5A ($M=1.5$), which is approximate to the results of MCCA and HAM. MCCA has the best performance among the four methods, and the correlation coefficient is positively correlated with the amplitude. The half-amplitude represents the significant degree of periodic fluctuation in the time series. With the increase of half-amplitude, the proportion of periodic components in the series increases, which makes it easier to be identified.

(5) Periodicity length

In Fig. 3b, The IE of PSM decreases with the increase of the periodicity length T except when $T=20$. The IE of MEM has the same variation as that of PSM, but with more moderate extent of change. The common defect of PSM and MEM is that consideration cannot be given to both the high and low frequency. The IE is higher in short T while longer T will lead to the identification of pseudo-periodic components. For HAM and MCCA, the IE is not affected by T and both are 100%.

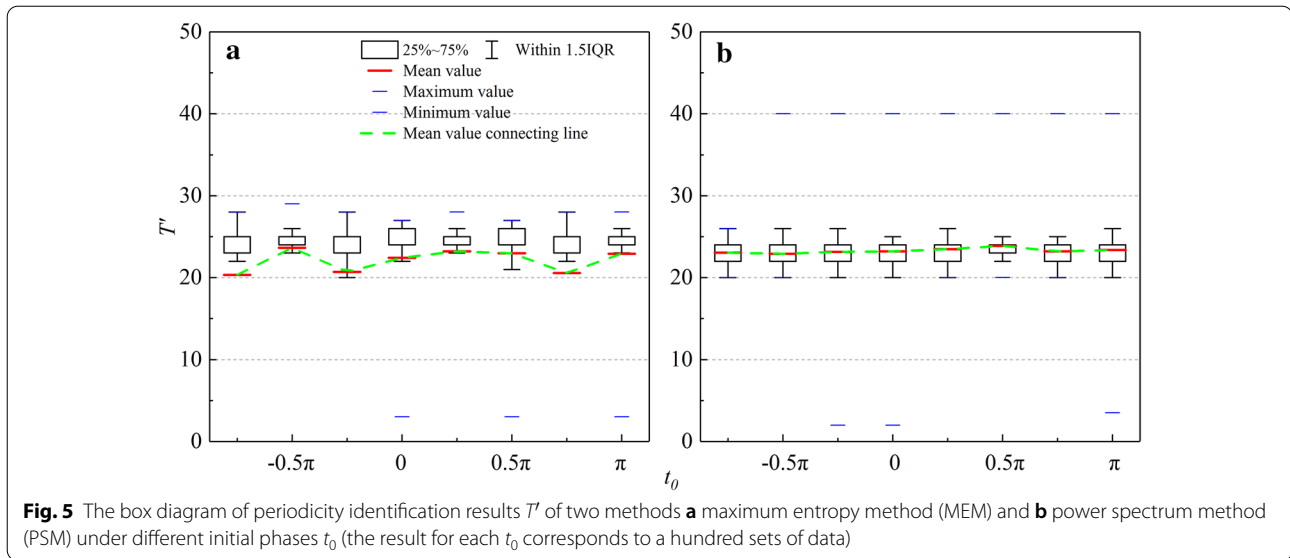
In order to analyze the performance of the two methods in detail when T changes, a box diagram with 100 sets of data of each group is given in Fig. 4. It shows that the mean value connecting line in Fig. 4a is smoother than that in Fig. 4b. The mean value lines in Fig. 4a are exactly corresponding to the theoretical values T' and the maximum and the minimum line also have no deviation or small deviation. While in Fig. 4b, when the theoretical value $T = 15$, the mean value line of T' is higher than 15; when $T = 30$, it is lower than 30 and the maximum line points to $T' = 33$. The overall comparison indicates that the identification results of the MCCA method are more accurate than the HAM method.



(6) Initial phase

From the results shown in Fig. 3c, it is obvious that the change of t_0 has little influence on IE. The IE of MCCA

and HAM both reach 100% with different initial phases while the IE of MEM and PSM are around 70% and 45%, respectively. To explain this difference more clearly, the



test data of MEM and PSM are given in the form of box diagram in Fig. 5. The mean value connecting line in Fig. 5a represents that T' identified by MEM are slightly higher than 20, while those of PSM are generally smaller than 25 in Fig. 5b. If the range of the allowable error is extended to $\Delta T = |T - T'| = 2$, the IE of MEM can increase to 85–90% and the IE of PSM can reach about 60%, which indicates that the IE values of these two methods are interfered by PT-III noise and the results are not accurate enough. Besides, there are minimum values lower than 5 and maximum values $T' = 40$ in Fig. 5b, and the existence of these pseudo periodicities also indicates the distortion of the identification results when methods are disturbed by noise.

In summary, the result shows that among these four methods, PSM and MEM have the worst performances; HAM and MCCA have similarly higher IEs (identification efficiency), especially for the MCCA method with the best performance. As the tests are on the synthetic time series, both the parameters of pure random component and periodic component will have impacts on the IE. When the periodic component gets insignificant due to the change of parameters, correspondingly, the IE of each method decreases. Specifically, the IE is positively correlated with the amplitude and sample length while negatively correlated with the mean value, coefficient of variation of stochastic components and length of periodicity, and almost independent of the initial phase when other factors are fixed. Based on the correlation coefficient criterion, the IE of the MCCA method decreases when the correlation coefficient becomes lower, especially when it is less than the critical value. When the periodicity is buried in much noise, the MCCA method

still shows its superiority compared with other three methods.

Study area and data

The Yangtze River is the largest river in China and the third largest river in the world. The Yangtze River basin (YRB, excluding Taihu Lake basin) includes 11 sub-basins linking southwest, central and eastern China (shown in Fig. 6). They are upper reaches of Jinsha River, lower reaches of Jinsha River, Mintuo River, Jialing River, Wu River, reaches from Yibin to Yichang, Dongting Lake system, Han River, Poyang Lake system, reaches from Yichang to Hukou and below Hukou, respectively.

We use the observed annual precipitation and annual runoff data from 1956 to 2017 to investigate the periodicities in the Yangtze River basin (YRB). The observed data is far more complicated than the generated synthetic time series because of the environmental and anthropological influences. Mixed with jump, trend, dependence or other types of variation, results of the periodicity identification will be interfered (Sang et al. 2009). For instance, a downward jump might be a section of a trough in the periodic fluctuation. Therefore, the jump or trend components in these runoff and precipitation time series are already subtracted before periodicity identification. We take Jialing River sub-basin as an example to illustrate the subtraction process. As plotted in Fig. 7, the mean value of series before 1993 and after 1993 (the red solid line in Fig. 7a, also defined as “jump”) are not at the same level. This downward jump at 1993 could be removed by first subtracting the value of jump component from the original series, and then adding the mean value of the series before 1993 to the whole series. Finally, the series after

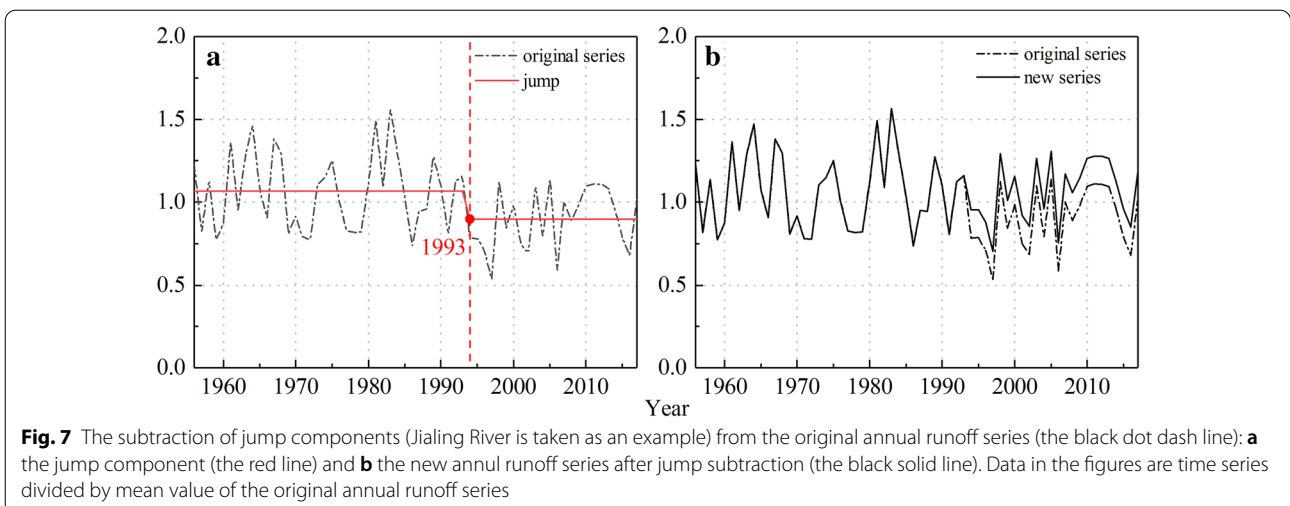
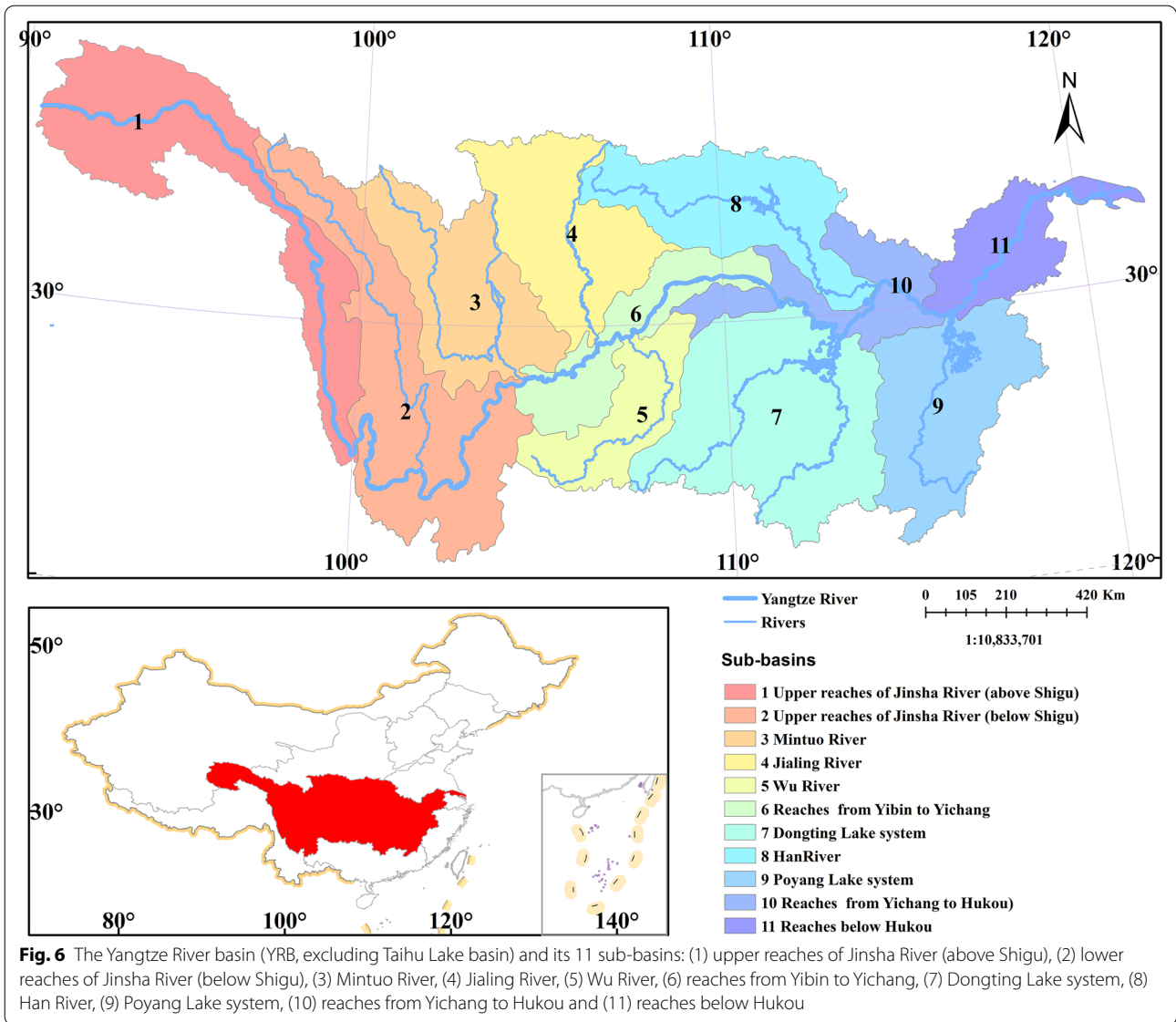


Table 6 The time period and the gradation standards of annual runoff and annual precipitation time series in Yangtze River basin (excluding Taihu Lake basin)

Time period	Significance level				
1956–2017	$0 \leq r < 0.250^a$	$0.250 \leq r < 0.325^a$	$0.325 \leq r < 0.6$	$0.6 \leq r < 0.8$	$0.8 \leq r < 1$
	No (N)	Weak (W)	Moderate (M)	Strong (S)	Dramatic (D)

^a For the annual time series with the length of 62 years, the thresholds of correlation coefficient are set as $r_{0.05} = 0.250$, $r_{0.01} = 0.325$, 0.6 and 0.8 to give five-level assessment

1993 is raised to the same mean value level as the series before it, which eliminates the impact of jump (as shown in Fig. 7b). The correlation coefficient thresholds for evaluating the significance of periodicities in the data are shown in Table 6.

Results and discussion

Identification of periodicities in runoff

Because of the important role in the distribution and management of water resources at regional scales and even the whole country, several studies have focused on the periodicities of runoff and precipitation in the YRB, where different temporal and spatial scales were concerned (Dai and Zhang 2013; Zhou et al. 2014). It has been confirmed that for precipitation in the YRB, it has periodicities of 4–7a, which is connected to the ENSO (El Niño–Southern Oscillation) (Yang et al. 2016), and also periodicities of 16a and about 20a, distributed along the lower reaches of Jinsha River and the upper reaches of the Yangtze River (Sun et al. 2012; Wang 2009; Yang et al. 2016). Runoff has a periodicity of 7–9a in the YRB, and a periodicity of 3–5a and about 20a in the upper reaches of the YRB (Chen et al. 2010; Wang and He 2004; Yang et al. 2016). Not many researches are on the analysis of the periodicity and its significance in the whole YRB.

In this study, the first dominant periodicity in the annual runoff time series in each sub-basin in the YRB is identified by the MCCA method, and other three commonly used methods PSM, HAM and MEM are also used in this section for comparison and verification. Given the poor performance of PSM in statistical experiments, the possible dominant periodicities are obtained by being calculated under several maximum lag m values ranging from $1/10n$ to $1/4n$ (n is the sample length 62 and m ranges from 6 to 15). The scoring criterion is set for MCCA method, 1 point if the periodicity identified by the MCCA has its counterpart in possible periodicities given by other three methods and otherwise 0 point. This scoring standard is to confirm the results of MCCA through the same results identified by other methods. A summary of results is shown in Table 7.

First of all, Table 7 shows that the performances of four methods are consistent with the conclusion of statistical experiment overall. To be specific, as for the periodicity

Table 7 The first dominant periodicities detected in the time series of annual runoff in 11 sub-basins in the Yangtze River basin (YRB, excluding Taihu Lake basin) by power spectrum method (PSM), maximum entropy method (MEM), harmonic analysis method (HAM) and moving correlation coefficient-based analysis (MCCA)

No. ^a	PSM	MEM	HAM	MCCA	Score ^b
1	2, 9, 10, 11, 12, 13	2, 4, 11	8.9, 12.4	9.2	1
2	4, 5, 6, 7, 11, 12	–	31	29.6	1
3	6, 7, 8, 9, 10, 11, 12	–	4.8, 7.8, 12	7.5	1
4	7, 8, 9, 10, 12	–	8.9	9.3	1
5	12–16, 18, 20, 22	–	3.1, 15.5	24.6	0
6	2, 3, 8, 9, 10, 12	–	3.1, 15.5	3.1	1
7	2, 24, 26, 28, 30	2	3.6	24.1	1
8	6, 7, 8, 9	–	6.9, 8.9	6.8	1
9	3, 4	2	–	21.8	0
10	2	2	6.9	6.7	1
11	2	2	–	2	1

^a The numbers from top to bottom in “No.” column refer to (1) upper reaches of Jinsha River, (2) lower reaches of Jinsha River, (3) Mintuo River, (4) Jialing River, (5) Wu River, (6) reaches from Yibin to Yichang, (7) Dongting Lake system, (8) Han River, (9) Poyang Lake system, (10) reaches from Yichang to Hukou and (11) reaches below Hukou, respectively

^b Give 1 point to MCCA for having corresponding result to at least one method and 0 point for none

identification of runoff, we can see that MEM only give results of sub-basin No. 1, 7, 9, 10 and 11 and the periodicities of them are all 2 years except sub-basin No. 1. Since the observed data are discretely sampled time series, in this paper, we tend to regard the periodicity of 2 years as random component in the case of annual time scale. It is also noteworthy that the results of PSM corresponding to different time lag m values are different. Multiple m values need to be tested to get reliable results, which increase its computation burden and the uncertainty of the results conversely. This reflects that these two methods are more inclined to be influenced by the stochastic characteristic of the time series than HAM and MCCA.

For most sub-basins, the results given by MCCA can be confirmed by other methods with 9 points for annual runoff series. To be specific, for sub-basin No. 11, all these four methods reach a consensus that there is no

significant periodic component but time series close to pure random component. For sub-basin No. 1, 3, 4, 6 and 8, all three methods except MEM have the same or similar results. Besides, at least one method is approximate to the result of MCCA for sub-basin No. 2, 7 and 10. Only for sub-basins No. 5 and 9, the MCCA method does not reach an agreement with other three methods. For the results that there is no consistent conclusion, we can verify the MCCA method on the basis of previous study results. The global periodicity of 25a of sub-basin No. 5 is given in (Xiong et al. 2010). As for sub-basin No. 9, according to (Liu et al. 2009; Ye et al. 2012), there are first dominant periodicity of 25a and secondary periodicity of 3–4a in sub-basin No. 9. Although PSM and MCCA both

have their corresponding results, not only the value but also the significance assessment of MCCA matches better with the known one. Besides, the annual runoff series of these two sub-basins discussed above are plotted in Fig. 8 fitted with the dominant periodic component identified by MCCA method. It can be seen intuitively that the periodic components (red line) of MCCA have good fit with the fluctuation of annual runoff series.

Characteristics and spatial distribution of periodicities of YRB runoff series

After verifying the application of MCCA method in the observed hydrological series, we next give a summary of complete results of YRB runoff series by MCCA

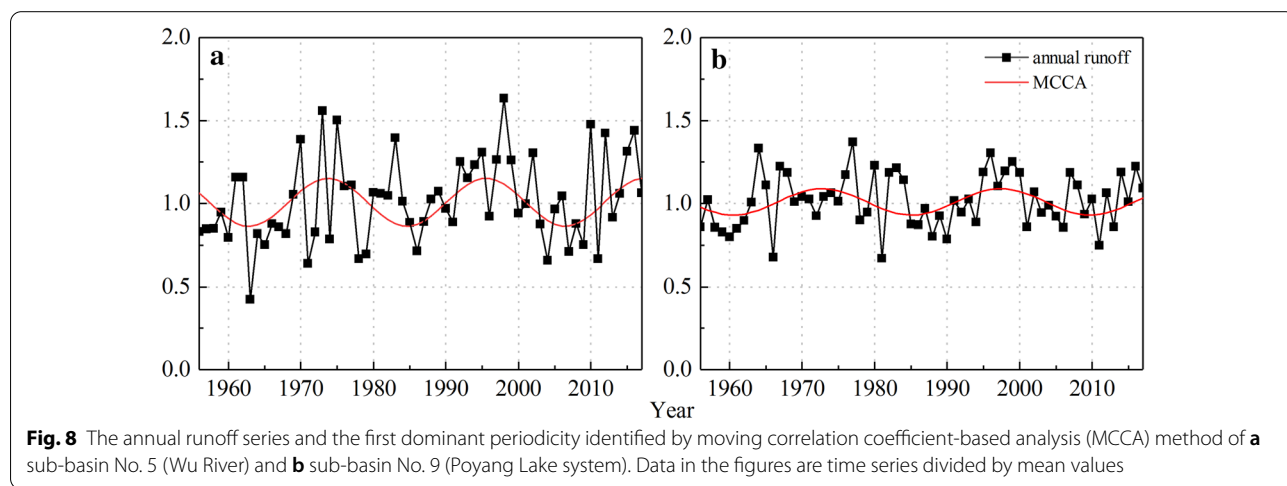


Fig. 8 The annual runoff series and the first dominant periodicity identified by moving correlation coefficient-based analysis (MCCA) method of **a** sub-basin No. 5 (Wu River) and **b** sub-basin No. 9 (Poyang Lake system). Data in the figures are time series divided by mean values

Table 8 Periodicities detected in the time series of annual runoff and annual precipitation in 11 sub-basins in the Yangtze River basin (YRB, excluding Taihu Lake basin), and the gradation of their significance levels based on the correlation coefficient *r*

No. ^a	Runoff				Precipitation			
	<i>T</i> ₁	Level ^b (<i>r</i>)	<i>T</i> ₂	Level (<i>r</i>)	<i>T</i> ₁	Level (<i>r</i>)	<i>T</i> ₂	Level (<i>r</i>)
1	9.2	M (0.466)	11.9	N (0.017)	11.7	M (0.410)	4.7	N (0.012)
2	29.6	M (0.353)	4.3	N (0.156)	4.3	M (0.353)	2.8	N (0.076)
3	7.5	M (0.399)	4.8	N (0.005)	4.8	M (0.393)	7.5	N (0.011)
4	9.3	M (0.405)	7.3	N (0.053)	9.4	M (0.360)	24.8	N (0.047)
5	24.6	M (0.351)	16.1	N (0.015)	4	W (0.288)	3.1	N (0.060)
6	3.1	M (0.397)	16.1	N (0.121)	16.2	M (0.373)	3.1	N (0.062)
7	24.1	M (0.420)	6.8	W (0.266)	23.2	M (0.343)	2.5	N (0.056)
8	6.8	M (0.344)	21.5	N (0.221)	3.1	W (0.314)	9	N (0.221)
9	21.8	M (0.399)	3.6	N (0.239)	21.4	M (0.365)	3.6	N (0.129)
10	6.7	M (0.392)	2.7	N (0.247)	6.7	M (0.369)	2.7	N (0.215)
11	6.7	M (0.355)	3.3	W (0.297)	6.7	W (0.324)	3.3	W (0.308)

^a The numbers from top to bottom in “No.” column refer to (1) upper reaches of Jinsha River (above Shigu), (2) lower reaches of Jinsha River (below Shigu), (3) Mintuo River, (4) Jialing River, (5) Wu River, (6) reaches from Yibin to Yichang, (7) Dongting Lake system, (8) Han River, (9) Poyang Lake system, (10) reaches from Yichang to Hukou and (11) reaches below Hukou, respectively

^b ‘N’ refers to no periodic variation; ‘W’ refers to weak periodic variation and ‘M’ refers to moderate periodic variation. For the annual time series with the length of 62 years (1956–2017), the thresholds of correlation coefficient are set as *r*_{0.05} = 0.250, *r*_{0.01} = 0.325, 0.6 and 0.8 to give five-level assessment

including the two dominant periodicities T_1 , T_2 , and their significance levels graded by the correlation coefficient r in Table 8. Considering little practical significance of the periodicity less than two years, we have filtered out this kind of result.

Overall, as for two main periodic components of annual runoff series, the shortest periodicity is 2.7a while the longest is 29.6a. It is noteworthy that sub-basins No. 1, 3, 4, 7, 8, 10, 11 all have significant (level W or M) periodic components in the range of 6.7–9.3a, which is consistent with the known fact that short periodicities of 7–9a are in most areas of YRB. Besides, there are also quite a few sub-basins (No. 2, 3, 6, 9, 10, 11) with a periodicity of 3–5a, but they are not statistically significant (level N) except for sub-basins No. 11 (level W).

The periodicities and their corresponding significance levels are further analyzed from a spatial perspective. First, as for runoff, it is obvious that runoff centralized in the south of YRB has longer periodicities than that of north and all periodicities are significant with level M. This result shows the difference of river runoff between north and south parts of YRB in terms of periodic characteristics. Sub-basins with periodicities of more than 5a (No. 1, 4, 5, 6, 7, 8) mostly distributed along the upper reaches of Yangtze River and the periodic components are not significant (level N) except for sub-basins No. 7 and 11.

Combined with the results of precipitation (as shown in Table 8), the periodicities of about 2.5–4.7a are in the whole basin except for sub-basin No. 4 and those of 6.7–9a (No. 3, 4, 8, 10, 11) are also common (Mao et al. 2014; Xiong et al. 2010). It is also obvious that periodicities of more than 5a are mostly distributed along upper reaches of Yangtze River. From this response, it can be concluded that there is a consistent one to one match between the periodicities of runoff and those of precipitation on the whole, and the precipitation mainly contributes to the periodic nature in runoff series in these regions (Zhang 2014).

There are some inconsistencies for the reason that the formation of runoff is also affected by many other factors such as underlying surface changes or human activities in addition to the hydrological processes. Many studies have shown that reservoir regulations and water withdrawal have a big impact on the runoff variability in the Yangtze River basin, which caused the impacts mainly reflected in the total runoff volume amount. The construction of reservoirs and the increase of water consumption that leading the annual runoff decline (Yang et al. 2010; Lei 2014; Zhang 2014; Tian 2016; Chen et al. 2018), usually shows as a trend or jump. This is also one of the reasons why the data in case study are pre-processed before periodicity analysis. As for the impact on periodicity, the storage

and discharge of reservoir mainly change the annual distribution of runoff. The reservoir regulation makes the runoff volume of upstream hydrological station decrease in flood season, and increase in non-flood season (Zhang and Yang 2014; Shu et al. 2016). Even for multi-year regulating reservoirs, this peaking cutting effect has little impact on the large timescale periodicities.

On the whole, the precipitation is still the main driving force for the interannual fluctuation of runoff (Zhang 2014). This is also in agreement with our conclusion. The issue on the runoff periodicity under various driving factors is still worth further study.

Conclusions

Extraction and quantitative evaluation of the significance of periodic components is important for hydrological time series analysis. In this regard, we proposed a new method, called MCCA, for the identification of periodicities, by utilizing the derived relationship between the correlation coefficient (CC) and the amplitude of periodicities. This correlation-aided method identified the significant periodicities and established a five-level criterion to evaluate different significance levels of periodicities.

Through investigating the influences of various statistical characteristics of data on the identification efficiency (IE) of the MCCA method, it was found that IE varied positively or negatively with some factors, and other three methods (PSM, HAM, and MEM) are used for comparison.

Specifically, as the mean value and the coefficient of variation of the time series gets larger, the IE of each method gets smaller, reflecting the impacts of stochastic term or noise on the identification of periodic component. By contrast, the IE of each method increases when the sample length and the amplitude get larger. The correlation coefficient was also positively related to IE, leading to the positive correlation between CC and the amplitude. This proves that CC can quantify the significance of the periodic components. PSM and MEM have the worst performances when the tested series are contaminated with much noise. HAM and MCCA had similarly better performances, especially for MCCA method with the highest IE. Indeed, these results generally suggest the superior accuracy and noise resilience characteristics of the MCCA method proposed.

The MCCA method was also performed over annual runoff series of 11 sub-basins of the Yangtze River basin (YRB, excluding Taihu Lake basin). The results found that annual runoff series have significant (level W or M) periodic components (6.7–9.3a) in 7 of 11 sub-basins, and periodicities of 3–5a are common in the rest sub-basins. We noticed that the sub-basins with a longer significant periodicity are mainly concentrated in the upper reaches

of the Yangtze River, and keep a good correspondence with the pattern of precipitation, indicating that precipitation has an important impact on the formation of runoff's periodicity. For some sub-basins, inconsistencies of periodicities between the runoff and the precipitation are probably a consequence of factors such as underlying surface changes or human activities in these areas. These results were consistent with previous studies, and the comparison with PSM, HAM and MEM also gave cogency to the results of MCCA. In this case, the proposed method is verified in the application to real hydrological data.

In conclusion, we confirmed that MCCA is a feasible scheme of identifying and evaluating hydrological periodicities. The advantage of the MCCA method is its simplicity of the principle and multi-level classification of the significance of the periodicity. Those commonly used methods can judge only whether the periodicity is significant or not at a certain confidence level, but no distinction based on the degree of significance. In terms of prediction accuracy, MCCA give a more detailed classification for all significant periods. In this case, the five-level criterion of MCCA has significant benefits for evaluating the impact of a periodic component on the time series.

However, studies probed into periodicity analysis are still moving forward, which also means the methods developed are not perfect. In terms of the periodicity pattern, MCCA is mainly tested for sinusoidal periodicities in statistical experiments. When extended to cases where the periodicities are non-sinusoidal, new techniques combining MCCA with some effective decomposition methods like *empirical mode decomposition* (EMD) (Huang et al. 1998; Huang and Wu 2008) may be operative. EMD can give adaptive *intrinsic mode functions* (IMFs) representing the underlying processes more effectively than pure sinusoids, and thus offer possibility for more reliable periodicity identification. Therefore, the MCCA method can be further improved in the future for its potential wide use ranges.

Abbreviations

HAM: Harmonic analysis method; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; CC: Correlation coefficient; EMD: Empirical mode decomposition; IMF: Intrinsic mode function; MEM: Maximum entropy; FFT: Fast Fourier transform; FPE: Final prediction error; PSD: Power spectral density; PSM: Power spectrum method; PT-III: Pearson type III; YRB: Yangtze River basin.

Acknowledgements

The authors gratefully acknowledge the valuable comments and suggestions given by the editors and the anonymous reviewers.

Authors' contributions

PX made substantial contributions to the conception, and supervised the work. LW analyzed and interpreted the experimental data and was a major contributor in writing the manuscript. YS supervised the work and

substantively revised it. FC revised the work. JC revised the work. ZW made major contributions to the methodology. YL revised the work. All authors read and approved the final manuscript.

Funding

This study was financially supported by the National Natural Science Foundation of China (Nos. 91547205, 41850410497, 41971040, 51579181 and 51779176), and the Youth Innovation Promotion Association CAS (No. 2017074).

Availability of data and materials

The runoff and precipitation data in the case study are available from the Yangtze & Southwest rivers water resources bulletin (<http://www.cjw.gov.cn/zwzc/bmgb/>). The bulletin was collected and compiled by Changjiang water resources commission of the Ministry of Water Resources according to the data provided by 20 provinces (including autonomous regions and municipalities directly under the central government) involved in the Yangtze River Basin and southwest rivers.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China. ² Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. ³ School of Geographical Sciences, Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo 315100, China. ⁴ Water@Leeds Research Institute, University of Leeds, Leeds LS2 9JT, UK.

Received: 17 November 2020 Accepted: 20 March 2021

Published online: 08 April 2021

References

- Akaike H (1970) Statistical predictor identification. *Ann Inst Stat Math* 22:203–217
- Akaike H (1974) A new look at statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Bartlett MS (1950) Periodogram analysis and continuous spectra. *Biometrika* 37:1–16
- Bordi I, Fraedrich K, Jiang JM, Sutera A (2004) Spatio-temporal variability of dry and wet periods in eastern China. *Theor Appl Climatol* 79:81–91
- Burg J (1975) Maximum entropy spectral analysis. Dissertation, Stanford University
- Cao H, Ellis BR, Littler JD (1997) The use of the maximum entropy method for the spectral analysis of wind-induced data recorded on buildings. *J Wind Eng Ind Aerodyn* 72:81–93
- Chen Y, Wang S, Wang G, Wang W (2010) Runoff variation characteristics analysis on Jinsha River. *Plateau Mt Meteorol Res* 30:27–30
- Chen K, Wang B, Xin M (2018) Impact of climate change and human activities on runoff variation of Yangtze River into sea. *Yangtze River* 49:36–40
- Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. *Math Comput* 19:297–301
- Coulibaly P, Baldwin CK (2005) Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *J Hydrol* 307:164–174
- Dai M, Zhang M (2013) Research on temporal and spatial distribution law of runoff in Yangtze River basin. *Yangtze River* 44:88–91
- Ghil M, Taricco C (1997) Advanced spectral analysis methods. In: Castagnoli GC, Provenzale A (eds) Past and present variability of the solar-terrestrial system: measurement data analysis and theoretical models. Proceedings of the international school of physics "Enrico Fermi". Società Italiana di Fisica/IOS Press, Bologna, pp 137–159
- Hao Y et al (2016) How does the anthropogenic activity affect the spring discharge? *J Hydrol* 540:1053–1065

- Huang N, Wu Z (2008) A review on Hilbert–Huang transform method and its applications to geophysical studies. *Rev Geophys* 46:RG2006
- Huang NE et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: *Proceedings of the Royal Society A: mathematical, physical, and engineering sciences*, vol 454. Royal Society, London, pp 903–995
- Jaynes ET (1982) On the rationale of maximum-entropy methods. *Proc IEEE* 70:939–952
- Kay SM (1988) *Modern spectral estimation: theory and application*. Prentice Hall, Englewood Cliffs, p 543
- Kottegoda NT (1980) *Stochastic water resources technology*. Springer
- Lei J (2014) Evolution trend and countermeasures of water resources in the Yangtze River basin under the influence of human activities. *Yangtze River* 45(7):7–10
- Liu J, Zhang Q, Xu C and Zhang Z (2009) Characteristics of runoff variation of Poyang Lake watershed in the past 50 years. *Trop Geogr* 29:213–218, 224
- Machiwal D, Jha MK (2012) *Structure of time series. Hydrologic time series analysis*. Springer
- Mao H, Chen G, Yang Q, Zhou C (2014) Comparative analysis on the application of precipitation anomaly percentage and Z index in Wujiang River Basin. In: *The 31st annual meeting of Chinese Meteorological Society*, Beijing, pp1–11
- McCuen RH (2003) *Modeling hydrologic change: statistical methods*. CRC Press, Boca Raton
- Mehdizadeh S, Behmanesh J, Khalili K (2017) A comparison of monthly precipitation point estimates at 6 locations in Iran using integration of soft computing methods and GARCH time series model. *J Hydrol* 554:721–742
- Nigmatullin RR, Khamzin AA, Tenreiro Machado J (2014) Detection of quasi-periodic processes in complex systems: how do we quantitatively describe their properties? *Phys Scr* 89:15201
- Nuttle WK (1997) Measurement of wetland hydroperiod using harmonic analysis. *Wetlands* 17:82–89
- Padmanabhan G, Rao AR (1988) Maximum entropy spectral analysis of hydrologic data. *Water Resour Res* 24:1519–1533
- Peres DJ, Cancelliere A (2016) Estimating return period of landslide triggering by Monte Carlo simulation. *J Hydrol* 541:256–271
- Rao AR, Jeong GD, Chang F (1992) Estimation of periodicities in hydrologic data. *Stoch Hydrol Hydraul* 6:270–288
- Rios RA, de Mello RF (2013) Improving time series modeling by decomposing and analyzing stochastic and deterministic influences. *Signal Process* 93:3001–3013
- Sakamoto Y, Kitagawa G (1987) *Akaike information criterion statistics*. Kluwer Academic Publishers
- Salas JD (1993) *Analysis and modeling of hydrologic time series*. In: Maidment DR (ed) *Handbook of hydrology*. McGraw-Hill, New York
- Sang Y, Wang D, Wu J, Zhu Q, Wang L (2009) The relation between periods' identification and noises in hydrologic series data. *J Hydrol* 368:165–177
- Sang Y, Wang Z, Liu C (2012) Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis. *J Hydrol* 424:154–164
- Sang Y, Xie P, Gu H, Li X (2017) Discussion on several major issues in the studies of hydrological. *Chin Sci Bull* 62:254–261
- Schuster A (1898) On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *J Geophys Res* 3:13–41
- Shu W, Li Q, Wang H, Wang X (2016) Impact analysis of climatic changes and human activities on characteristics of inflow runoff of three gorges reservoir. *Water Power* 42:29–33
- Siegel AF (1980) Testing for periodicity in a time series. *J Am Stat Assoc* 75:345–348
- Singh VP (1998) *Pearson type III distribution. Entropy-based parameter estimation in hydrology*. Springer, Dordrecht
- Steele TD (1982) A characterization of stream temperatures in Pakistan using harmonic analysis. *Hydrol Sci J* 27:451–467
- Stojkovic M, Kostić S, Plavšić J, Prohaska S (2017) A joint stochastic-deterministic approach for long-term and short-term modelling of monthly flow rates. *J Hydrol* 544:555–566
- Stosic T, Telesca L, Vicente DSFD, Stosic B (2016) Investigating anthropically induced effects in streamflow dynamics by using permutation entropy and statistical complexity analysis: a case study. *J Hydrol* 540:1136–1145
- Sun J, Lei X, Jiang Y, Wang H (2012) Variation trend analysis of meteorological variables and runoff in upper reaches of Yangtze River. *Water Resour Power* 30:1–4
- Tamura Y, Sato T, Ooe M, Ishiguro M (1991) A procedure for tidal analysis with a Bayesian information criterion. *Geophys J Int* 104:507–516
- Thomson DJ (1982) *Spectrum estimation and harmonic analysis*. Proc IEEE 70:1055–1096
- Tian Q (2016) *Impacts of climate change and human activity on water and sediment flux of the Yellow, Yangtze and Pearl River basins over the past 60 years*. Dissertation, East China Normal University
- Troch PA, Carrillo G, Sivapalan M, Wagener T (2013) Climate–vegetation–soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution. *Hydrol Earth Syst Sci* 17:2209–2217
- Wang S (2009) Changing pattern of the temperature, precipitation and runoff in Chuanjiang section of the Yangtze River. *Resour Sci* 31:1142–1149
- Wang P, He S (2004) The basic character on the process runoff and sediment discharge at Datong station of Yangtze River. *J East China Normal Univ (Nat Sci)* 2:72–80
- Wang Q, Me Z (1990) Discussions on some key problems in power spectra analysis. *Acta Geogr Sin* 45:363–372
- Welch PD (1967) The use of Fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15:70–73
- Wu X et al (2016) Variability of annual peak flows in the Beiji River basin, South China, and possible underlying causes. *Hydrol Res*. <https://doi.org/10.2166/nh.2016.228>
- Wu Z et al (2019) Moving correlation coefficient-based method for jump points detection in hydroclimate time series. *Stoch Environ Res Risk Assess* 33:1751–1764
- Xie P et al (2018) Evaluation of the significance of abrupt changes in precipitation and runoff process in China. *J Hydrol* 560:451–460
- Xiong Y, Zhang K, Yang G, Gu Z (2010) Periodic changes of the precipitation and runoff in Wu River watershed. *J Sichuan Agric Univ* 28:475–479
- Yang H (2015) *Research of hydrological time series cycle analysis method*. China Water Power Electr 5:63–66
- Yang S, Liu Z, Dai S et al (2010) Temporal variations in water resources in the Yangtze River (Changjiang) over the Industrial Period based on reconstruction of missing monthly discharges. *Water Resour Res* 46(10):W10516.1–W10516.13
- Yang L, Mei Y, Ye Y, Lin Y (2016) Temporal analysis of hydrological and meteorological factors in downstream of Jinsha River and three Gorges reach of Yangtze River. *Hydrology* 36:37–45
- Ye X, Zhang Q, Liu J, Xu L (2012) Natural runoff change characteristics and flood/drought disasters in Poyang Lake catchment basin. *J Nat Disasters* 21:140–147
- Yuan J, Chen Y, Gu S, Xu G (2016) Cycle identification of annual runoff time series based on Holt–Winters method. *Water Resour Power* 34:28–31
- Zhang X (2014) *Climate change and anthropogenic impacts on water discharge in the Yangtze River over the nearly 60 years*. Dissertation, East China Normal University
- Zhang X, Yang S (2014) Climatic and anthropogenic impacts on water discharge in the Yangtze River over the last 56 years (1956–2011). *Resour Environ Yangtze Basin* 23:1729–1739
- Zhou W, Sornette D (2011) Statistical significance of periodicity and log-periodicity with heavy-tailed correlated noise. *Int J Mod Phys C* 13:137–169
- Zhou N, Yang S, Shen X, Liu X (2014) Mutation and multi-scale characteristics analysis of rainfall series in Dongting Lake watershed. *J Tongji Univ (Nat Sci)* 42:867–872

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.