ISEIS

Journal of Environmental Informatics 42(1), 65-74 (2023)

Journal of Environmental Informatics

www.iseis.org/jei

A Non-Parametric Approach for Change-Point Detection of Multi-Parameters in Time-Series Data

Y. M. Hu[1], C. X. Yang[1], Z. M. Liang[1 *], X. Y. Luo[1], Y. X. Huang[1], and C. Tang[1]

[1] *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*

Received 10 September 2021; revised 06 January 2022; accepted 12 March 2022; published online 05 May 2022

ABSTRACT. Change-point analysis of time-series data plays a vital role in various fields of earth sciences under changing environments. Most of the analysis approaches were usually designed to detect the change-point in the level of time-series mean. In this study, we aimed to propose a non-parametric approach to detect the change-point of different parameters of time-series data. In this approach, the Bootstrap method, coupling with Kernel density estimation, was first used to estimate the probability distribution function (pdf) of a parameter before and after any potential change-points. Second, the *Ar*-index based on the uncross area of the two pdfs was designed to quantify the difference of the parameter before and after each potential change-point. Finally, the potential change-point owning the largest *Ar*-index value was determined as the locations of the change-point of the parameter. The hydrological extreme series from four stations in the Hanjiang basin were used to demonstrate this approach. The Pettitt test method commonly used in hydrology was employed as a comparison to indirectly analyze the reliability of the proposed approach. The results show that change-point detected by the proposed approach in the four stations are identified with those detected by the Pettitt approach in the level of time-series mean. But in comparison with the Pettitt test, the proposed approach can provide more detection information for other parameters, such as coefficient of variation (*Cv*) and coefficient of skewness (*Cs*) of the series. The results also show that the degree of change in the series mean is greater than its *Cv* and *Cs*, while the degree of change in series *Cv* is greater than its *Cs*.

Keywords: time-series data, change-point analysis, multi-parameters, bootstrap method, kernel density estimation

1. Introduction

Time-series data nonstationarity analysis, including series temporal trend and change-point detection, has been an important research content in various fields of earth science under changing environments (Burn and Elnur, 2002; Zhang et al., 2007; Milly et al., 2009; Engström and Waylen, 2017; López et al., 2017; Hu et al., 2021). In the field of hydrology, numerous published studies from around the world have demonstrated that the hydrological series exhibited significant nonstationarity characteristics, such as increasing/decreasing trend, upward/downward shift or a combination of them, due to the impact of climate change and human activities over the past decades (Perreault et al., 1999; Xiong and Guo, 2004; Narisma et al., 2007; Wang et al., 2016). It is necessary to mine the hydrological time-series nonstationarity characteristics, which is useful for analyzing climate change (Zhang et al., 2007; Murumkar and Arya, 2014), understanding the evolution laws of the hydrological event (Yang and Tian, 2009; Ben et al., 2014; Sarhadi et al., 2016; Hu et al., 2019) and guiding water resource engineering planning and operation in changing environment (Pina et al., 2016; Ferguson et al., 2018; Volpi et al., 2018).

* Corresponding author. Tel.: +86 10 83786475; fax: +86 10 8378647.
E-mail address: zmliang@hhu.edu.cn (Z. M. Liang).

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.202200478.

To analyze the change-point in hydrological time-series, several typical approaches have been commonly applied, such as Lepage Test (Lepage, 1971; Azizabadi and Khalili, 2013), Wilcoxon rank-sum test (Wilcoxon, 1945; Hao et al., 2016), Petitt test (Petitt, 1979; Liang et al, 2018; Hu et al., 2021), and Bayesian model (Perreault et al.,1999; Xiong and Guo, 2004). The Lepage test is a non-parametric and two-sample approach, which can be used to detect whether there is a significant difference between the studied two samples in the mean of the data series (Azizabadi and Khalili, 2013). The Lepage test requires that the sample size of the studied data series should be not less than ten and assumes that the Lepage statistic follows the Chi-squares distribution with two degrees of freedom (Lepage, 1971). Theoretically, the Lepage test can be applied to detect multiple change-points of the studied series. However, the location of the detected change-point is influenced by time windows of different lengths, which leads to uncertainty in the identification of change-points (Chen et al., 2009). The Wilcoxon rank-sum (WRS) test is a non-parametric alternative to the two samples t-test, which can be applied to identify whether two samples come from the same population (Wilcoxon, 1945). In comparison with the t-test which assumes that sample data is subject to a normal distribution, the WRS test can be used for unknown distributions and an assumed distribution is not necessary (Li et al., 2018). The WRS test is based solely on the rank in which the observations from the two groups of samples before and after the potential change-point fall (Hao et al., 2016). The WRS statistic follows approximately a standard normal distribution. Significance change-point indicated that the series mean before and after the change-point are significantly different. Pettitt test is also a non-parametric distribution-free approach, which has been widely reported in various published papers for change detection in both rainfall and streamflow data (Sagarika et al., 2014; Engström and Waylen, 2017; Liang et al., 2018). This approach can not only detect the location of the change-point but also test whether the detected change-point is significant at a given significant level (Pettitt, 1979). However, the change-point provided by the Pettitt approach is also in the level of the series mean (Ma et al., 2008). Another typical approach is the Bayesian test model, which is a parametric test approach. It allows for the detection of multiple change-points in a time series. Although the Bayesian model is highly robust in identifying change-point, it should be pointed out that some Bayesian models require normality of time series with a relatively light tail for the convenience of mathematics (Xiong and Guo, 2004; Chen et al., 2009). Thus, the Bayesian test should be carefully applied to the time series following the skewed distribution with a heavy-tailed. For non-normal time series, the Box-Cox transformation (Box and Jenkins, 1976) must be employed to transform the original data series into a new series with a normal distribution (Xiong and Guo, 2004). Other change-point analysis approaches include the Brown-Forsythe (Brown and Forsythe, 1974), Lee-Heghinian (Lee and Heghinian, 1977), Sequential Clustering (Ding, 1986), Lombard (Lombard, 1987) and fused lasso approach (Jeon et al., 2016). In recent years, the nonstationarity analysis for multivariate hydrological series receives more and more attention (Xiong et al., 2015). However, considering that our study is limited to univariate hydrological series, we do not discuss too much about the change-point analysis in multivariate scenarios.

Although many approaches have been proposed to analyze the location of change-points of univariate time series, it is worth noting that most of these approaches usually cannot make clear which parameters in the hydrological series change, instead of just providing the year in which a time series break, or usually were designed to detect the change-point in the mean levels of a time series. Most approaches are impossible to detect whether other parameters change apart from the series mean, such as the coefficient of variation (*Cv*) and the coefficient of skewness (*Cs*) of a series. However, to identify the changing characteristics of different parameters of a series is essential for gaining a deep understanding of the changing features of the hydrological series. For example, an increase in the *Cv* of extreme series after a change-point usually indicates that the possibility of extreme events tends to increase in the future period. Besides, when employing a nonstationary hydrological frequency analysis model to analyze the return level of extreme precipitation or flood under changing environment, it should make clear which parameters of distribution change and how each parameter changes with some covariates (Strupczewski et al., 2001; Rootzen and Katz, 2013; Hu et al., 2018; Liang et al., 2018).

This study was aimed at developing a non-parametric approach to detect change-points of multi-parameters of a time series. This objective was threefold: (1) to describe the probability distribution characteristic of a parameter of interest before and after a potential change-point by using Bootstrap method coupling with the Kernel density estimation; (2) to design an index to quantify the difference of the parameter after and before the potential change-point; (3) to identify the most-likely locations of the change-points of multi-parameters of a series and assess its significance. Finally, the peak flow series from four stations of the Hanjiang basin in China were used to demonstrate this proposed approach.

The remainder of the paper was organized as follows: In section 2, we introduced the methodologies used in this study. In section 3, the study area and used data from four stations were described. In section 4, Results were presented. Finally, the main conclusions were summarized.

2. Methodologies

2.1. Bootstrap Method

The bootstrap method is a non-parametric resampling technique with replacement. It has been commonly used for inferring the expectation, standard deviation or distribution of a parameter estimate of interest. The idea behind the bootstrap method is that the sample series is the best guide to the underlying true distribution even when the information about the true distribution is lacking (Efron, 1992; Hu et al., 2013). It does not need the assumption of the true distribution followed by the sample series and only depends on the sample series itself. Here, the Bootstrap method will be used to generate larger numbers of estimates of a parameter of interest of a hydrological series before and after a potential change-point.

Taking series mean Ex and variance σ of a series as an example, for a given known distribution function $F(x)$, the statistic parameter Ex and σ can be calculated by (Hu et al., 2015):

$$Ex = \int_{-\infty}^{\infty} x dF(x)$$

$$\sigma = \int_{-\infty}^{\infty} (x - Ex)^2 dF(x) \tag{1}$$

Supposing a sample series $X = (X_1, X_2, \ldots, X_n)$ without knowing its true distribution function. The empirical cumulative distribution function (CDF) $\hat{F}_n$ of the sample series can be described as the following:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x) \tag{2}$$

where $I(\cdot)$ is the indicator function, $I(\cdot) = 1$ if $x_i \le x$, otherwise, $I(\cdot) = 0$.

Then, the mean $\hat{E}x$ and variance $\hat{\sigma}$ of the sample can be estimated by using its empirical CDF:

$$\hat{E}x = \int_{-\infty}^{\infty} x d\hat{F}_n(x)$$

$$\hat{\sigma} = \int_{-\infty}^{\infty} \left(x - Ex\right)^2 d\hat{F}_n\left(x\right) \tag{3}$$

Resampling N groups of new samples with replacement from the original time series $X = (X_1, X_2, \ldots, X_n)$, called Bootstrap samples. Based on these Bootstrap samples, N groups of estimates of Ex and σ can be calculated by using Equations (2) ~ (3). Finally, the distribution of the parameters E_x and σ are approximated by the empirical CDF of their N estimations, separately (Hu et al., 2013, 2015).

2.2. Kernel Density Estimation Method

Kernel density estimation (KED) is one of the non-parametric density estimators, which learns the shape of the density from the data automatically and does not need the assumption that the underlying probability density function (PDF) is from a parametric family (Parzen, 1962). Here, the Kernel density estimation will be used to fit the bootstrap-based parameter estimation samples to obtain the distribution of the parameter estimate.

Let $X_1, X_2, \ldots, X_n$ be an independent and identically distributed random sample from an unknown distribution $F(x)$ with probability density function $f(x)$, which is to be estimated, then the KED can be defined as (Parzen,1962):

$$\hat{f}_h\left(x\right) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4}$$

where $K(\cdot)$ is the kernel function, which controls the weight given to the observations $\{X_i\}$ at each point X based on their proximity; h is a smoothing parameter known as the bandwidth, and it controls the size of the neighborhood around X; n is the size of the sample series. In this study, the most widely used Gaussian kernel is used as the kernel function:

$$K(y) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right) \tag{5}$$

Thus, the KED of the variable X can be expressed:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right] \tag{6}$$

The optimal smoothing parameter h can be estimated by the following Equation (Nosratabadi et al., 2019):

$$h = \left\{\frac{\int\left[K\left(x\right)^2 dx\right]}{n\sigma_k^2\int\left[f''\left(x\right)\right]^2 dx}\right\}^{\frac{1}{5}} \tag{7}$$

where σ_k is the standard deviation of the kernel function $K(\cdot)$.

For a Gaussian kernel, the optimal smoothing parameter h can be simply calculated by the following Equation (Nosratabadi et al., 2019):

$$h = 1.06sn^{-1/5} \tag{8}$$

where σ is the standard deviation of sample data.

2.3. *Ar*-Index-Based Approach for Parameter Change-Point Detection

Supposing the series $X_1, X_2, \ldots, X_n$ is divided into two different subseries based on a potential change-point at the location of τ, the before-point series $x_1, x_2, \ldots, x_\tau$ is called $S1$ and the afterpoint series $x_{\tau+1}, x_2, \ldots, x_n$ is called $S2$. Resampling from the original sample $S1$ and $S2$ for N repeats by using the Bootstrap method, separately. Then N groups of Bootstrap samples of $S1$ and $S2$ can be obtained. Based on each of the N groups of Bootstrap samples, one corresponding estimation of a parameter of interest of $S1$ and $S2$ can be obtained, such as mean (Ex), coefficient of deviation (Cv) or the coefficient of skewness (Cs). Therefore, N estimations of the given parameter can be obtained. In this study, the Linear-Moment method (Hosking and Wallis, 1997; Liang et al., 2014) is applied for estimating the parameters of Ex, Cv and Cs. Finally, the KED is employed to calibrate the estimation samples of the above three parameters of the series $S1$ and $S2$ to obtain their continuous pdfs. The similarity degree of the parameter before and after the change-point τ can be quantified by using the cross area of the two pdfs of $S1$- and $S2$-based estimation of the given parameter (Figure 1).

If there is only one cross point of the two different pdfs of the parameter related to $S1$ and $S2$ (Figure 1(a)), the cross area (CAr) is calculated by:

$$CAr = \int_{-\infty}^{a} f_B\left(y\right)dy + \int_{a}^{+\infty} f_A\left(y\right)dy \tag{9}$$

where $f_A(\cdot)$ and $f_B(\cdot)$ are the pdfs of the parameter related to the after-point and before-point series estimated by using the Bootstrap method and Kernel density estimation method, separately; a is the location of the cross point.

If there are two cross points (Figure 1(b)), the cross area (CAr) is calculated by:

$$CAr = \int_{-\infty}^{a} f_B\left(y\right)dy + \int_{a}^{b} f_A\left(y\right)dy + \int_{b}^{+\infty} f_B\left(y\right)dy \tag{10}$$

where a and b are the first and second locations of the two cross points. For other cases with more than two cross points, the calculation procedure is similar.

The uncrossed area Ar was calculated by using Equation (11), which was employed to quantify the difference between the before-point pdf and after-point pdf, and further to assess the parameter change degree before and after the change-point:

$$Ar = 1 - CAr \tag{11}$$

The uncrossed area Ar is located in the interval [0, 1] as

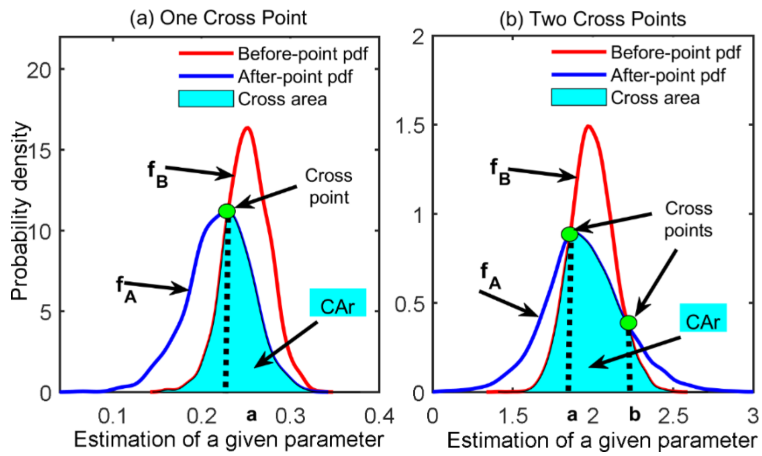

Figure 1. Diagram for calculating the cross area of the two different PDFs of a parameter of interest before and after the change-point: (a) one cross point case, and (b) two cross points case.

(CAr) ∈ [0, 1] calculated by Equation (9) or (10). Ar = 1 indicates that the two pdfs are completely different from each other, while $Ar = 0$ indicates that the two PDFs are completely consistent with each other. The larger the uncrossed area Ar, the more significantly different the parameter before and after the change-point.

Setting a series of potential locations of the change-points of the observation series, i.e., $\{\tau_1, \tau_2, \cdots, \tau_m\}$, then for a parameter of interest, the Ar-index value corresponding to each potential change-point can be calculated, i.e., $\{Ar_1, Ar_2, \cdots, Ar_m\}$. The location of the change-point of the parameter can be determined by choosing the one owning the biggest Ar value:

$$\tau = \tau\left(\max\left\{Ar_1, Ar_2, \cdots, Ar_m\right\}\right) \tag{12}$$

2.4. Significance Test of Parameter Change-Point

After using the Ar-index to identify the most-likely change-point location of a parameter of interest, it is necessary to determine whether the change-point is significant. After obtaining the location of the change-point, the empirical cumulative distribution function of the parameter related to the after-point and before-point series can be calculated by using the Bootstrap method mentioned in section 2.1. If there is a significant difference between the after-point and before-point distribution of the parameters, it means that the change-point of the parameter at this location is significant. In this study, the Kolmogorov-Smirnov (K-S) test was used to analyze whether the before-point and after-point distribution of the parameter changes significantly, and then to evaluate whether the parameter has a significant change at the most-likely change-point position. The K-S test is a non-parametric statistic, which defines the largest absolute difference between the two cumulative distribution functions as a measure of disagreement. The K-S test can be described as the follows (Wang et al., 2011):

$$D_n = max\,|\hat{F}_A - \hat{F}_B| = \max_{1\le i\le n} \delta_i \tag{13}$$

where $\hat{F}_A$ and $\hat{F}_B$ are the after-point and before-point distribution of the parameter, respectively; δ_i is the absolute difference between the two distributions.

For a given significance level α, the corresponding threshold value of the K-S statistical is calculated:

$$D_\alpha = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}\sqrt{\frac{n+m}{nm}} \tag{14}$$

where n and m are the number of samples used to calculate the after-point and before-point empirical distribution of the parameter.

If D_n is greater than or equal to D_α, it indicates that there is a significant difference between the two distributions at the significance level α. That is, the change-point of the parameter at this location is significant. Otherwise, the change-point is not significant.

3. Study Area and Data

The peak flow series from four stations in the Hanjiang basin in China were used in this study. The four stations were XinDianPu, GuoTan, HuangLongTan and HuangZhuang. The period and size of the four peak flow series were listed in Table 1. Figure 2 shows the location of the four stations.

Figure 3 presents the time series diagrams of the four peak flow series, respectively. It seems that the peak flow series in XinDianPu, GuoTan and HuangZhuang have a possible change-point at about 1985 while the HuangLongTan has it at about 1990.

4. Results

For the three parameters Ex, Cv and Cs, the Ar-index value related to each of the given potential change-points were calculated. The search range of potential change-points of the four

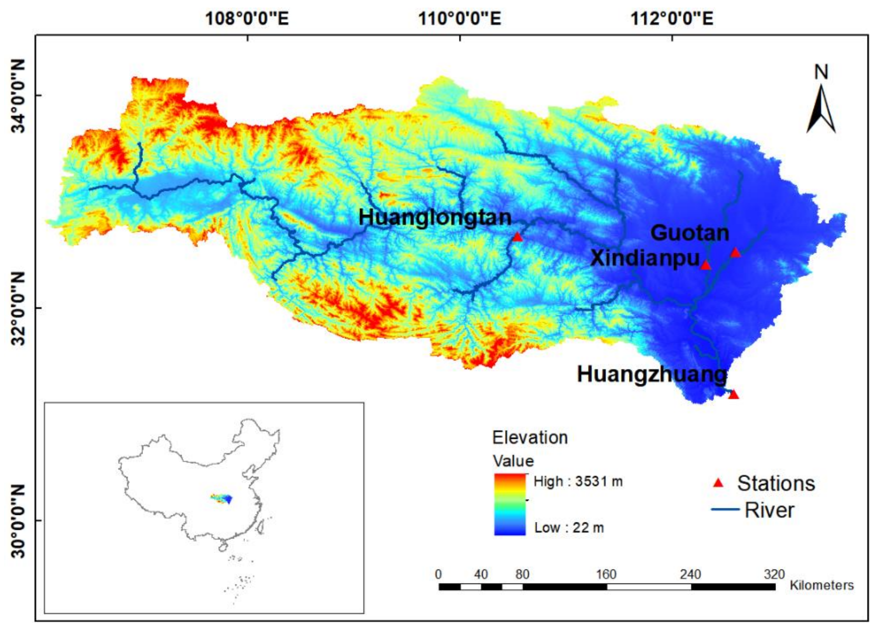

Figure 2. The location of the four stations in the Hanjiang basin in China.

peak flow series was listed in Table 1. The calculation results of *Ar*-index values, corresponding to the three parameters in different potential change-points, were presented in Figure 4. It can be seen that for different potential change points, the corresponding *Ar*-index values of the three parameters were different.

Table 1. Basic Information of the Four Stations and the Set of Search Range of Potential Change-Points

Names	Observed period	Sample size	Search range
XinDianPu	1953 ~ 2014	62	[1963, 2004]
GuoTan	1957 ~ 2014	58	[1967, 2004]
HuangLongTan	1953 ~ 2014	62	[1963, 2004]
HuangZhuang	1954 ~ 2014	61	[1964, 2004]

Table 2. Results of the Change-Point Detection Using the *Ar*-Index-Based Approach and Pettitt Approach

Names	Ar-based method		Pettitt method
	Parameter	Location	Location
XinDianPu	*Ex*	1983	1983
	Cv	1994	
	Cs	1964	
GuoTan	*Ex*	1984	1984
	Cv	1974	
	Cs	1968	
HuangLongTan	*Ex*	1991	1991
	Cv	1977	
	Cs	1999	
HuangZhuang	*Ex*	1984	1984
	Cv	1966	
	Cs	1991	

Table 2 presents the change-point detection results obtained by the *Ar*-index-based approach. It can be seen that for the XinDianPu, GuoTan, HuangLongTan and HuangZhuang station, the location of the change-point of the three parameters (*Ex*, *Cv*, *Cs*) was at (1983, 1994, 1964), (1984, 1974, 1968), (1991, 1977, 1999) and (1984, 1966, 1991), respectively. This indicates that the change-points of the three parameters of each of the four series are at the different locations, which cannot be detected if using the commonly used change-point test, i.e., Pettitt test, because they only can detect the location of the change-point in the level of the series mean instead of to detect other parameters changes. Moreover, it can be also found that the change degree of the parameter *Ex* was more obvious than the parameters *Cv* and *Cs*, while the change degree of the parameter *Cv* was more obvious than the parameter *Cs*.

Given that the Pettitt approach has been commonly used for detecting the change-point in hydrology, the detection results of the change-point using the Pettitt approach was provided as a comparison with those obtained by the *Ar*-index-based approach in the level of series mean. As can be seen in Figure 5 that the Pettitt-based change-point locations of the four peak flow series were in 1983, 1984, 1991 and 1984, respectively (Table 2), which were in accordance with the *Ar*-index-based detection results in terms of the parameter *Ex*. However, compared to the *Ar*-index-based approach, the Pettitt approach cannot detect if other parameters change other than the series mean.

Based on the locations of the change-point detected by the *Ar*-index-based approach and the Pettitt approach (Table 2), the corresponding pdf of the parameters *Cv* and *Cs* before and after the change-point, called them as before-point pdf (BPP) and after-point pdf (APP), were estimated by using the Bootstrap method coupling with Kernel density estimation mentioned in section 2. It should be noted that the Pettitt approach cannot detect the

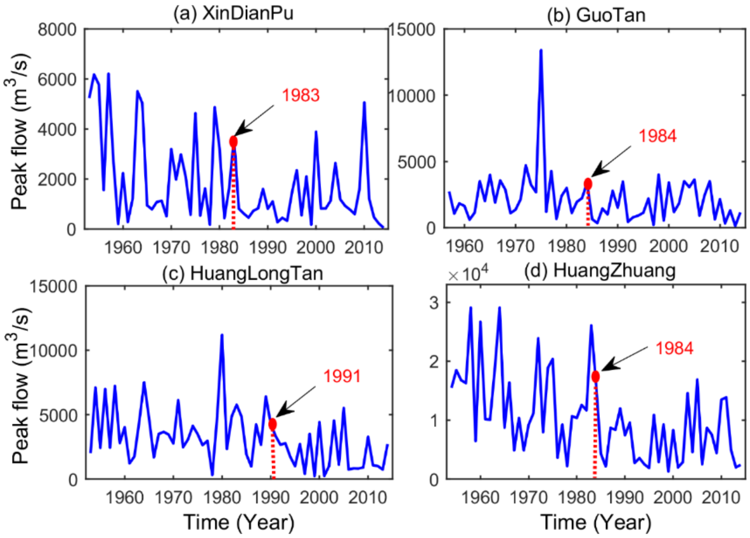

Figure 3. Time series diagrams of the peak flow in the four stations: (a) XinDianPu station, (b) GuoTan station, (c) HuangLongTan station, and (d) HuangZhuang station.

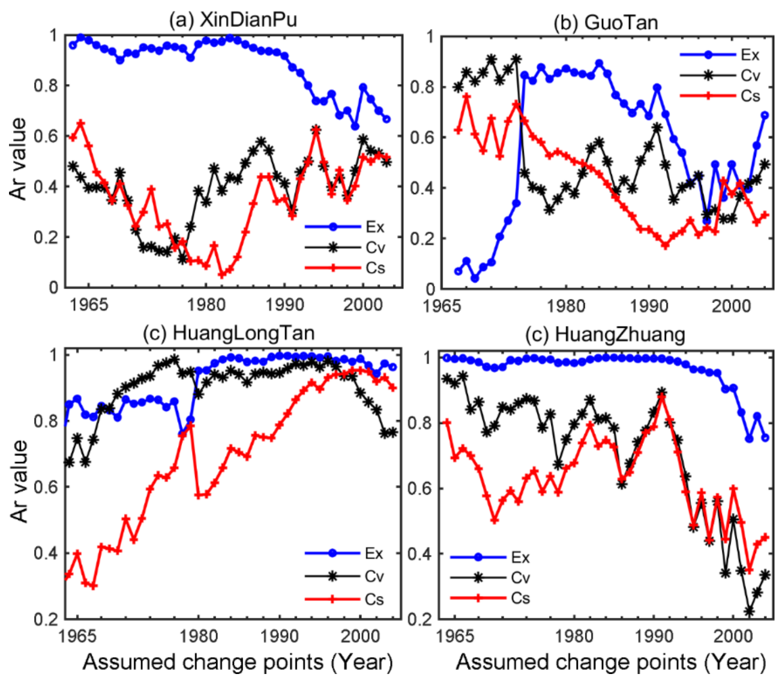

Figure 4. Results of the *Ar*-index values for the parameters *Ex*, *Cv* and *Cs* at the different potential change-points for the four stations: (a) XinDianPu station, (b) GuoTan station, (c) HuangLongTan station, and (d) HuangZhuang station.

change-point of parameters *Cv* and *Cs*, here for comparison purposes, assuming that the change-point of *Cv* and *Cs* are the same as the series mean provided by the Pettitt approach. As can be seen in Figure 6. that for a given parameter in each of the four stations, the cross area of BPP and APP at the location of the *Ar*-index-based change-point was less than that at the location of the Pettitt-based change-point. That is, the difference between the BPP and APP provided by the *Ar*-index-based approach was more significant than that provided by the Pettitt approach. This indicates that the change-point detected by the *Ar*-index-based approach is more reasonable than that obtained by the Pettitt approach.

For the parameters of *Cv* and *Cs* of the four stations, the Pettitt-based and *Ar*-index-based uncrossed area values between BPP and APP were presented in Table 3. Taking the XinDianPu station as an example, in terms of the parameter *Cv*, the uncrossed area between BPP and APP was 0.44 in the Pettitt-based point (1983) and 0.63 in the *Ar*-index-based point (1994), while in terms of the parameter *Cs*, the uncrossed area was 0.07 in the Pettitt -based point (1983) and 0.65 in the *Ar*-index-based

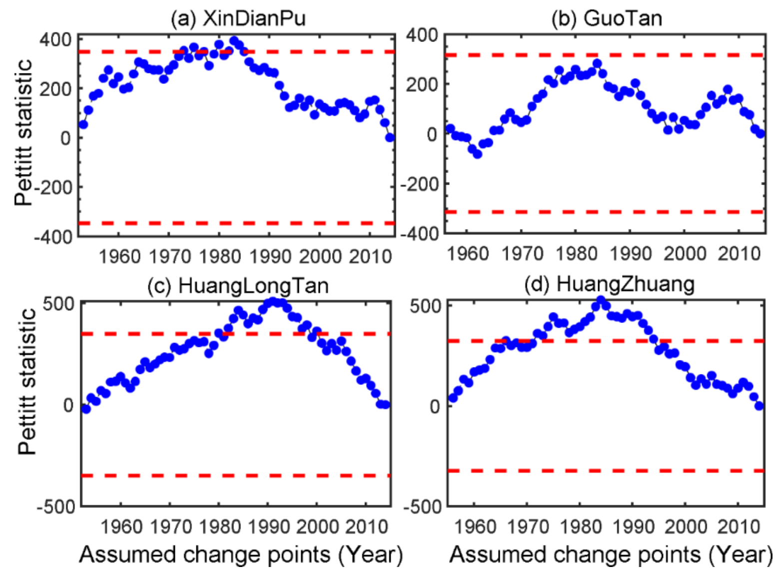

Figure 5. Results of the Pettitt statistic at the different potential change-points for the four stations: (a) XinDianPu station, (b) GuoTan station, (c) HuangLongTan station, and (d) HuangZhuang station.

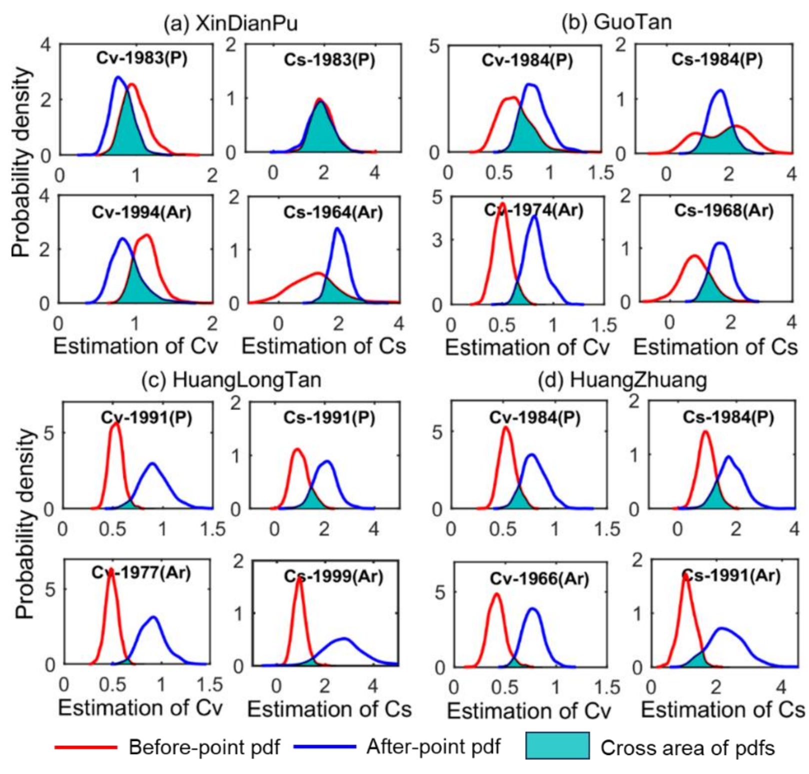

Figure 6. The pdf of the parameters *Cv* and *Cs* before and after the change-point related to the *Ar*-index-based approach marked with the symbol "Ar" and the Pettitt approach marked with the symbol "P". The smaller the cross area of the two pdf, the greater the difference of the parameter before and after the change-point. The subfigures are for (a) XinDianPu station, (b) GuoTan station, (c) HuangLongTan station, and (d) HuangZhuang station, respectively.

point (1964). The larger the uncrossed area, the more obvious the difference of the parameter between the BBP and APP.

The non-parametric Kolmogorov-Smirnov test was used to analyze whether the change-points of the parameters *Ex*, *Cv* and *Cs* of each peak flow series are significant at the 5% significance level. The results from Table 4 show that the *Ar*-index-based change-points of the three parameters are significant at the 5% significance level.

Table 3. Pettitt-Based and *Ar*-Index-Based Uncrossed Area of the Parameters *Cv* and *Cs* of the Four Stations

Station	Parameter	Pettitt-based uncrossed area	*Ar*-index-based uncrossed area
XinDianPu	*Cv*	0.44	0.63
	Cs	0.07	0.65
GuoTan	*Cv*	0.58	0.91
	Cs	0.46	0.76
HuangLongTan	*Cv*	0.95	0.99
	Cs	0.82	0.95
HuangZhuang	*Cv*	0.81	0.94
	Cs	0.75	0.88

Table 4. Results of the Change-Point Significance Test by Using the K-S Method

Names	*Ar*-index-based method		5% significance level
	Parameter	Location	if significance
XinDianPu	*Ex*	1983	Yes
	Cv	1994	Yes
	Cs	1964	Yes
GuoTan	*Ex*	1984	Yes
	Cv	1974	Yes
	Cs	1968	Yes
HuangLongTan	*Ex*	1991	Yes
	Cv	1977	Yes
	Cs	1999	Yes
HuangZhuang	*Ex*	1984	Yes
	Cv	1966	Yes
	Cs	1991	Yes

5. Conclusions and Discussion

In this study, a non-parametric approach was proposed to detect the change-points of multi-parameters of time-series data. In this approach, the Bootstrap method was first employed to generate larger numbers of estimations of a parameter of interest before and after a potential change-point. Then the Kernel density estimation was used to fit the Bootstrap-based parameter estimation samples to obtain the probability density function (pdf) of the parameter before and after the given potential change-point, called before-point pdf and after-point pdf. Based on the two pdfs, an *Ar*-index was designed to quantify the difference between the before-point pdf and after-point pdf related to a potential change-point. The potential change-point owning the largest *Ar*-index value was determined as the most-likely change-point of the parameter of interest. Finally, the Kolmogorov-Smirnov (K-S) test was used to evaluate whether the parameter has a significant change at the most-likely change-point position.

To demonstrate the applicability of this proposed approach, the annual maximum peak flow series from the four different stations were used. The results show that the change-points detected by the *Ar*-index-based method are the same as those detected by the commonly used Pettitt approach in the level of series mean. But in comparison with the Pettitt approach, the *Ar*-index-based approach can provide more detection information for any parameters, not limited to the parameters mentioned in this study, which is of great importance for more fully understanding the change characteristic of hydrology series.

The proposed approach was a non-parametric-distribution-free method, which has no requirement on which distribution function the time series follows. The approach can be used to detect the location of change-point of any parameters of a time series, not limited to the mentioned parameters in this study including *Ex*, *Cv* and *Cs*. The approach was developed based on the probability density function of a parameter of interest. Thus, to a certain extent, the approach can consider the impact of limited sample length on the reliability of detection results, because it depends on the distribution of the parameter instead of its point estimates, which may be useful for reducing the effect of series length on detection results.

In addition, in our demonstrative examples, the most widely used Gaussian kernel with high robustness was applied in Kernel density estimation for obtaining the continuous pdfs of the estimation samples of the parameters *Ex*, *Cv* and *Cs*. In the practical application, considering the possible influence of different kernel density function selection on diagnostic results, it would be better to use different kernel functions with different smoothing parameters to fit the parameter estimation samples and chose the optimal one based on some appropriate goodness index.

Acknowledgements. This study was supported by the National Key R&D Program of China (2018YFC1508001), the National Natural Science Foundation of China (51709073), and Fundamental Research Funds for the Central Universities of China (B220202031).

References

Azizabadi, F.M. and Khalili, D. (2013). Seasonality characteristics and spatio-temporal trends of 7-day low flows in a large, semi-arid watershed. *Water Resour. Manag.*, 27(14), 4897-4911. https://doi.org/10.1007/s11269-013-0445-6

Ben Aissia M.A., Chebana F., Ouarda T.B.M.J., Roy, L., Bruneau, P. and Barbet, M. (2014). Dependence evolution of hydrological characteristics, applied to floods in a climate change context in Quebec. *J. Hydrol.*, 519, 148-163. https://doi.org/10.1016/j.jhydrol.2014.06.042

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control* (revised edn). Holden-Day, SanFrancisco, California, USA. ISBN: 1118674928, 9781118674925

Brown, M.B. and Forsythe, A.B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16(1), 129-1321. https://doi.org/10.1080/00401706.1974.10489158

Burn, D.H. and Elnur, M.A.H. (2002). Detection of hydrologic trends and variability. *J. Hydrol.*, 255, 107-122. https://doi.org/10.1016/S0022-1694(01)00514-5

Chen, Y.D., Zhang, Q., Xu, C.Y., Yang, T., Chen, X.H. and Jiang, T. (2009). Change-point alterations of extreme water levels and underlying causes in the Pearl River Delta, China. *River. Res. Appl.*, 25(9), 1153-1168. https://doi.org/10.1002/rra.1212

Ding J. (1986). Statistical detection for transition point in flood time

sequences. *Journal of Wuhan University of Hydraulic and Electric Engineering*, 5, 36-41.
Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*. Springer, New York, NY, USA, pp. 569-593. https://doi.org/10.1007/978-1-4612-4380-9_41
Engström, J. and Waylen, P. (2017). The changing hydroclimatology of Southeastern U.S., *J. Hydrol.*, 548, 16-23. https://doi.org/10.1016/j.jhydrol.2017.02.039
Ferguson, C.R., Pan, M. and Oki, T. (2018). The effect of global warming on future water availability: CMIP5 synthesis. *Water Resour. Res.*, 54(10), 7791-7819. https://doi.org/10.1029/2018WR022792
Hao, Y.H., Zhang, J., Wang, J.J., Li, R.F., Hao, P.M. and Zhan, H.B. (2016). How does the anthropogenic activity affect the spring discharge?. *J. Hydrol.*, 540, 1053-1065. https://doi.org/10.1016/j.jhydrol.2016.07.024
Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis*, Cambridge University Press.
Hu, Y.M., Liang, Z.M., Li, B.Q. and Yu, Z.B. (2013). Uncertainty assessment of hydrological frequency analysis using bootstrap method. *Math. Probl. Eng.*, 2013, 1-7. https://doi.org/10.1155/2013/724632
Hu, Y.M., Liang, Z.M., Liu, Y.W., Zeng, X.F. and Wang, D. (2015). Uncertainty assessment of estimation of hydrological design values. *Stoch. Env. Res. Risk. A.*, 29 (2), 501-511. https://doi.org/10.1007/s00477-014-0979-z
Hu, Y.M., Liang, Z.M., Singh, V.P., Zhang, X.B., Wang, J., Li, B.Q. and Wang, H.M. (2018). Concept of equivalent reliability for estimating the design flood under non-stationary conditions. *Water Resour. Manag.*, 32(3), 997-1011. https://doi.org/10.1007/s11269-017-1851-y
Hu, Y.M., Liang, Z.M., Xiong, L.H., Sun, L., Wang, K., Yang, J., Wang, J. and Li, B.Q. (2019). Assessment on annual precipitation change in the headwater source of the middle route of China's South to North Water Diversion Project. *Theor. Appl. Climatol.*, 137(3-4), 2529-2537. https://doi.org/10.1007/s00704-018-02756-x
Hu, Y.M., Liang, Z.M., Solomatine, D.P., Wang, H.M. and Liu, T. (2021). Assessing the impact of precipitation change on design annual runoff in the headwater region of Yellow River, China. *J. Environ. Inform.*, 37(2), 122-129. https://doi.org/10.3808/jei.201900421
Jeon, J., Sung, J.H., and Chung, E. (2016). Abrupt change-point detection of annual maximum precipitation using fused lasso. *J. Hydrol.*, 538, 831-841. https://doi.org/10.1016/j.jhydrol.2016.04.043
Lee, A.F.S. and Heghinian, S.M. (1977). A shift of the mean level in a sequence of independent normal random variable: A Bayesian approach. *Technometrics*, 19(4), 503-506. https://doi.org/10.1080/00401706.1977.10489592
Lepage, Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika*, 58, 213-217. https://doi.org/10.1093/biomet/58.1.213
Li, Z.W., Xu, X.L., Xu, C.H., Liu, M.X. and Wang, K.L. (2018). Dam construction impacts on multiscale characterization of sediment discharge in two typical karst watersheds of southwest China. *J. Hydrol.*, 558, 42-54. https://doi.org/10.1016/j.jhydrol.2018.01.034
Liang, Z.M., Hu, Y.M., Li, B.Q. and Yu, Z.B. (2014). A modified weighted function method for parameter estimation of Pearson type three distribution. *Water Resour. Res.*, 50(4), 3216-3228. https://doi.org/10.1002/2013WR013653
Liang, Z.M., Yang, J., Hu, Y.M., Wang, J., Li, B.Q. and Zhao, J.F. (2018). A sample reconstruction method based on a modified reservoir index for flood frequency analysis of non-stationary hydrological series. *Stoch. Env. Res. Risk. A.*, 32(6), 1561-1571. https://doi.org/10.1007/s00477-017-1465-1
Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74, 615-624. https://doi.org/10.1093/biomet/74.3.615
López, I., Rodríguez, C., Gámez Cámara, M.A., Varga, M. and Garay, J. (2017). Change-point method applied to the detection of temporal variations in seafloor bacterial mat coverage. *J. Environ. Inform.*, 29(2), 122-133. https://doi.org/10.3808/jei.201500330
Ma, Z.M., Kang, S.Z., Zhang, L., Tong, X. and Su, X.L. (2008). Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. *J. Hydrol.*, 352(3), 239-249. https://doi.org/10.1016/j.jhydrol.2007.12.022
Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. (2008). Stationarity is dead: whither water management. *Science*, 319, 573-574. https://doi.org/10.1126/science.1151915
Murumkar, A.R. and Arya, D.S. (2014). Trend and periodicity analysis in rainfall pattern of Nira Basin, Central India. *A. J. Climate*, 3, 60-70. https://doi.org/10.4236/ajcc.2014.31006
Narisma, G.T., Foley, J.A., Licker, R. and Ramankutty, N. (2007). Abrupt changes in rainfall during the twentieth century. *Geophys. Res. Lett.*, 34(6), L06710. https://doi.org/10.1029/2006GL028628
Nosratabadi, H., Mohammadi, M. and Kargarian, A. (2018). Nonparametric probabilistic unbalanced power flow with adaptive kernel density estimator. *IEEE T. Smart Grid*, 10(3), 3292-3300. https://doi.org/10.1109/TSG.2018.2823058
Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065-1076.
Perreault, L., Hache, M., Slivitsky, M. and Bobee, B. (1999). Detection of changes in precipitation and runoff over eastern Canada and US using a Bayesian approach. *Stoch. Env. Res. Risk. A.*, 13(3), 201-216. https://doi.org/10.1007/s004770050039
Pettitt, A. (1979). A nonparametric approach to the change-point problem. *Appl. Stat.*, 28(2), 126-135. https://doi.org/10.2307/2346729
Pina, J., Tilmant, A. and Anctil, F. (2016). Horizontal approach to assess the impact of climate change on water resources systems. *J. Water Res. Plan. Man.*, 143(4), 04016081. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000737
Rootzen, H. and Katz, R.W. (2013). Design life level: quantifying risk in a changing climate. *Water Resour. Res.*, 49(9), 5964-5972. https://doi.org/10.1002/wrcr.20425
Sagarika, S., Kalra, A. and Ahmad, S. (2014). Evaluating the effect of persistence on long-term trends and analyzing step changes in streamflows of the continental United States. *J. Hydrol.*, 517, 36-53. https://doi.org/10.1016/j.jhydrol.2014.05.002
Sarhadi, A., Burn, D.H., Ausin, M.C. and Wiper, M.P. (2016). Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resour. Res.*, 52(3), 2327-2349. https://doi.org/10.1002/2015WR018525
Strupczewski, W.G., Singh, V.P. and Feluch, W. (2001). Non-stationary approach to at-site flood frequency modeling I. Maximum likelihood estimation. *J. Hydrol.*, 248(1-4), 123-142. https://doi.org/10.1016/S0022-1694(01)00397-3
Volpi, E., DiLazzaro, M., Bertola, M., Viglione, A. and Fiori, A. (2018). Reservoir effects on flood peak discharge at the catchment scale. *Water Resour. Res.*, 54(11), 9623-9636. https://doi.org/10.1029/2018WR023866
Wang, C., Zeng, B. and Shao, J. (2011). Application of bootstrap method in Kolmogorov-Smirnov test. In *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*. IEEE, 287-291. https://doi.org/10.1109/ICQR2MSE.2011.5976614
Wang, J., Liang, Z.M., Wang, D., Liu, T. and Yang, J. (2016). Impact of climate change on hydrologic extremes in the upper basin of the Yellow River Basin of China. *Adv. Meteorol.*, 2016, Article ID 1404290. https://doi.org/10.1155/2016/1404290
Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin*, 1(6), 80-83. https://doi.org/10.1007/978-1-4612-4380-9_16
Xiong, L.H. and Guo, S.L. (2004). Trend test and change-point detec-

tion for the annual discharge series of the Yangtze River at the Yichang hydrological station. *Hydrolog. Sci. J.*, 49(1), 99-112. https://doi.org/10.1623/hysj.49.1.99.53998

Xiong, L.H., Jiang, C., Xu, C.Y., Yu, K.X. and Guo, S.L. (2015). A framework of change-point detection for multivariate hydrological series. *Water Resour. Res.*, 51(10), 8198-8217. https://doi.org/10.1002/2015WR017677

Yang, Y. and Tian, F. (2009). Abrupt change of runoff and its major driving factors in Haihe River Catchment, China. *J. Hydrol.*, 374(3-4), 373-383. https://doi.org/10.1016/j.jhydrol.2009.06.040

Zhang, X.B., Zwiers, F.W., Hegerl, G.C., Lambert, F.H., Gillett N.P., Solomon, S., Stott, P.A. and Nozawa, T. (2007). Detection of human influence on twentieth-century precipitation trends. *Nature*, 448(7152), 461-465. https://doi.org/10.1038/nature06025