



Regional detection of multiple change points and workable application for precipitation by maximum likelihood approach

Qianfeng Wang^{1,2} · Jia Tang^{1,2} · Jingyu Zeng^{1,2} · Song Leng³ · Wei Shui^{1,2}

Received: 26 December 2017 / Accepted: 4 September 2019 / Published online: 2 December 2019
© Saudi Society for Geosciences 2019

Abstract

Precipitation is an essential part of the hydrological cycle. Objective change-point detection plays an important role in the research on extreme climate events and risk assessment in the context of global climate change. The study can automatically identify and extract multiple change points over a lengthy time series and calculate trends over several segmentations by uniting the maximum likelihood approach and the nonparametric Mann-Kendall trend test which is also compared with ordinary least squares (OLS). The observed station precipitation data were compiled over Hebei Province in China from 1961 to 2014. Temporal-spatial characteristics were also investigated by using several indices: the number of change points, standard deviation, timing of change points, minimum and maximum trends for segmentations, and the generalization trend. Obvious change points were generally around 1974, 1981, and 1998, and occurred at few stations before 1974 and after 2000, indicating that precipitation was relatively stable in the study area during the periods 1961–1974 and 2000–2014; the minimum trend for segmentations decreased over all stations; the maximum trend for segmentations increased over all stations except Leting; and generalization trend weakened abrupt changes in specific time sections among the multiple segmentations. Change-point detection followed by trend analysis can detect an obvious increasing or decreasing trend over certain parts of the time series and the proposed method can serve as a management tool with proper measures to deal with climate change. The results for both segmentations and generalization can provide a workable reference for managing regional water resources and implementing strategies to mitigate meteorological risks.

Keywords Precipitation · Change point · Maximum likelihood approach · Mann-Kendall · Trend; Hebei

Introduction

Global climate change continues to influence our world's ecosystem, water cycle and soil condition, crop yield, and other aspects of the natural environment in addition to social and economic activities (Fischer et al. 2012; Panda et al. 2013; Zhang 2013; Wang et al. 2018). Research on global climate

change has grown particularly intense in recent years as the scientific community—and indeed, the public has grown increasingly aware of these effects. (Warner et al. 2009; Wu et al. 2010). Precipitation is an essential part of the hydrological cycle. Abnormal changes in precipitation contribute to a host of environmental problems (Djaman et al. 2016). Precipitation tends to rise in the tropical and high-altitude zones across the globe while decreasing in subtropical zones (Alexander et al. 2006; Frich et al. 2002). These changing trends vary by region (Mihailović et al. 2015). In China, study has revealed inconsistent trends at the regional scale due to the various analysis approaches and study periods to investigate them (Fan et al. 2012; Ip et al. 2011). Consequently, there is an urgent necessity for trend analysis methods that function independent of the study periods.

Within lengthy time series datasets, precipitation may change non-monotonically (Khaliq et al. 2007; Rahmani et al. 2015) or monotonically, however, with wide variations in amplitude across different periods (Ruggieri 2013). These

Responsible Editor: Abdullah M. Al-Amri

✉ Wei Shui
shuiwei@fzu.edu.cn

- ¹ College of Environment and Resource, Fujian Provincial Key Laboratory of Remote Sensing of Soil Erosion and Disaster Prevention, Fuzhou University, Fuzhou 350116, China
- ² Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education of China, Fuzhou 350116, China
- ³ Climate Change Cluster, University of Technology Sydney, Broadway, New South Wales 2007, Australia

phenomena reflect the existence of change points in such datasets (Chen et al. 2016; Fealy and Sweeney 2005). Traditional approaches to change-point detection assume that the change point is identical locations (or timing) and quantity of change points in the data across multiple time series; it is not flexible enough to apply for longer time series (Jamali et al. 2015). Moreover, when change points are recognized, at present, existing approaches cannot contain the necessary indices to verify the reliability and practicability of the recognized change points (Chen et al. 2016; Fealy and Sweeney 2005; Fu et al. 2015). Because change-point detection is mainly used to analyze data of lengthy time series in economy, environment, ecology, management, and demography, as well as meteorological research, its approach must meet new requirements for flexibility and simplicity.

Trend analysis approaches are primarily comprised of parametric or nonparametric trend test (Burn and Elnur 2002; Wang et al. 2014; Xu et al. 2004; Zenklusen Mutter et al. 2010). Parametric approaches are simple but require data series which follow normal distribution. The least squares method is a popular such approach, in which the best fit is a trend line with the minimized sum of squared residuals and the significance of linear fitting is considered the significance of the trend (Crujeiras and Van Keilegom 2010; Wang et al. 2014). Nonparametric approaches, conversely, do not require the data series to follow normal distribution. The most commonly used is a calculation process that combines the Mann-Kendall trend test with the Sen's method (Kendall 1948; Mann 1945; Sen 1968). This approach is more reliable in terms of significance test results than the least squares method and is widely applied for meteorology, hydrology, agriculture, ecology, and environment (Burn and Elnur 2002; Wang et al. 2015, 2017; Xu et al. 2004). The present study focuses on a novel nonparametric approach for analyzing precipitation trends over lengthy time series.

Most of the trend analysis that has been conducted has not considered the existence of a change point in the time series analysis (Suhaila and Yusop 2018), also according to a previous study that different methods often give different numbers and locations of change points (Sharma et al. 2016); however, the identification of change points in the previous studies only consider the location and number of change points; the trend of each segmentation between adjacent change points cannot be assessed jointly in the time series; in our study, trend tests can be used as a management tool with proper measures to evade climate change impacts (Nouri et al. 2017). We try to propose an optimized non-parametric method including the automatic identification of change points and calculation of segmentation trends for each station, combined with the trend approach and change point detection which are popular hot topics in hydrology climate and ecology fields (Razmi et al. 2017). Therefore, the combination of identifying change points and calculating segmentations trends is our innovation.

This study proposes a change-point detection approach that automatically identifies the characteristics of change points (locations and timing) and calculates the trend characteristics of multiple segments. The approach involves detecting the locations and timing of change points using the maximum likelihood approach, and assessing trends analysis and significance in the data in each segment using the Mann-Kendall trend test which is also compared with the ordinary least squares (OLS) method. Notably, we need to do a normal distribution test before we use OLS linear regression. The proposed approach was applied to a region scale using the number of change points, standard deviation, timing of change points, minimum and maximum trend for segmentations, and generalization trend across the entire study period as the index. The results presented here may provide a workable reference for regional water resource management and adaptation to extreme climate, as well as technical support for time series data analysis in other domains of social sustainability.

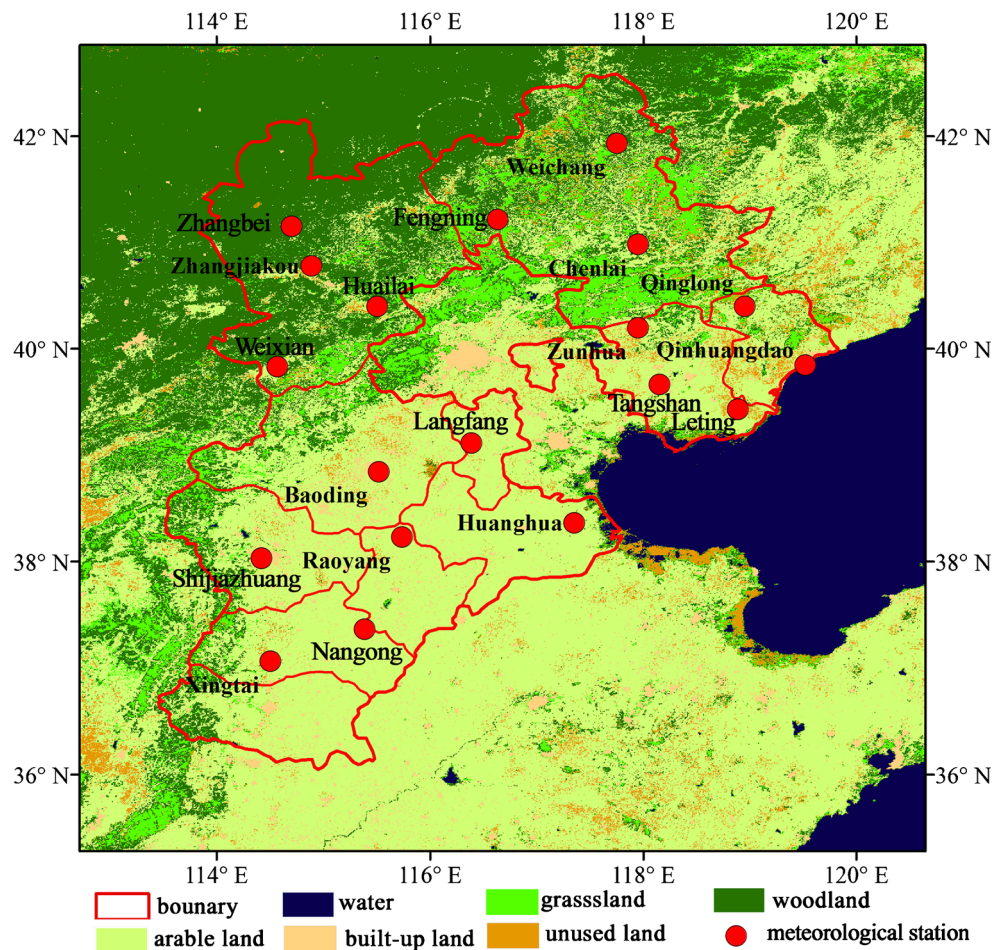
Study area and data sources

Hebei Province in northern China (113.45–119.83° E, 36.05–42.58° N) includes 11 cities with 188,800 km² in total area. It features complex and varied landforms and is the only Chinese province encompassed with plateaus, mountains, hills, plains, lakes, and coasts altogether. The terrain declines from northwest to southeast.

The study area consists of three landform units, Bashang Plateau, Yanshan Mountains and Taihang Mountains, and Hebei Plain respectively. Hebei has a temperate continental monsoon climate characterized by four distinctive seasons in most areas, with greater precipitation in the southeast than in the northwest. Nineteen meteorological stations are provided data for this study; they represent arable land, woodland, and grassland cover (Fig. 1). These stations are evenly distributed across the province and effectively represent the climatic and spatial patterns of Hebei.

The series of daily precipitation values (mm) over the 19 stations during the period 1961–2014 were sourced from the China Meteorological Administration (<http://data.cma.cn/en>). The discontinuities with series of daily precipitation can lead to misinterpretations of the studied climate. In order to avoid errors and obtain homogeneous climate time series, non-natural irregularities in climate data series must be detected and removed prior to its use (Ribeiro et al. 2016). From March 2011 to June 2012, the China Meteorological Administration carried out the construction of ground basic meteorological data. Repeated quality testing and control were carried out on the continuous observation data in the monthly ground report data files of national stations from 1951 to the present, which corrected a large number of wrong data, and made up for the digital missing data, so that the data quality has been significantly improved. So, we can analyze the

Fig. 1 Study area



data directly with no need to do extra homogeneity assessment. Several stations with missing values were excluded to ensure the integrity and continuity of the data series across the entire study period. MODIS Land Cover product data with a 500-m spatial resolution for the year 2013 were adopted (available online from the United States Geological Survey website). The land cover in the study area was classified into six categories: water, grassland, woodland, arable land, built-up land, and unused land, pursuant to the classification standards (Liu et al. 2003). The spatial distribution of these land cover types is shown in Fig. 1.

Methodology

Multiple change-point detection approach

A change-detection approach identifies the time node where the data element in a lengthy time series changes abruptly. It also allows the researcher to divide time series data into discrete segments. In traditional detection, the locations and quantity of change points may be identical across multiple time series (Jamali et al. 2015). Although the fitting with the least squares

method is popular in the present research on the detection of multiple change points, it is prone to error due to the heavy-tailed distribution of the data series (Zou et al. 2014). Segmentation at arbitrary change point also fails to reveal optimal points in terms of their locations and quantity. However, the distribution of the data in a lengthy time series within individual segmentations is typically unknown. This type of parametric method can sustain model misspecifications. If the data with time series have no parametric assumption, we consider the determined change point based on the independent data $\{X_i\}_{i=1}^n$:

$$X_i \sim F_k(x) \quad \tau_{k-1} \leq i \leq \tau_k - 1, k = 1, \dots, K_n + 1, \quad i = 1; \dots, n \quad (1)$$

where K_n is the true number or quantity of change points; the locations or date (for time series data) of the change points with the convention of $\tau_0=1$ and $\tau_{K_n+1}=n+1$, F_k denotes cumulative distribution function (CDF) for the segment k satisfying $F_k \neq F_{k+1}$; the number of change points can be allowed to alter according to the sample size n .

We assume that Z_1, \dots, Z_n are mutually independent and following the same distribution F_0 , and let \hat{F}_n denote empirical

CDF of sample, then $n\hat{F}_n(u) \sim \text{binomial}(n, F_0(u))$. If we consider the sample as the binary data with a probability of success $\hat{F}_n(u)$, in the context for formula (1), the joint-likelihood for the candidate set for change points $(\tau'_1 < \dots < \tau'_L)$ can be written as follow:

$$\xi_u(\tau'_1, \dots, \tau'_L) = \sum_{k=0}^L (\tau'_{k+1} - \tau'_k) \left\{ \begin{aligned} &\hat{F}_{\tau'_k}^{\tau'_{k+1}}(u) \log \left(\hat{F}_{\tau'_k}^{\tau'_{k+1}}(u) \right) \\ &+ \left(1 - \hat{F}_{\tau'_k}^{\tau'_{k+1}}(u) \right) \log \left(1 - \hat{F}_{\tau'_k}^{\tau'_{k+1}}(u) \right) \end{aligned} \right\} \quad (2)$$

where $\hat{F}_{\tau'_k}^{\tau'_{k+1}}$ denotes the empirical CDF of subsample time series $\{X_{\tau'_k}, X_{\tau'_{k+1}-1}\}$ with $\tau'_0 = 1$ and $\tau'_{L+1} = n + 1$. To estimate change points, we maximize the formula (1) in an integrated form as follow:

$$R_n(\tau'_1, \dots, \tau'_L) = \int_{-\infty}^{\infty} \xi(\tau'_1, \dots, \tau'_L) dw(u) \quad (3)$$

where $w(\cdot)$ denotes the positive weight function; thus, R_n can be finite, and the integral is used to combine all the information across u .

The log-likelihood function formula (2) is essentially related to two-sample data goodness of fit statistic test depending on nonparametric likelihood ratio (Einmahl and McKeague 2003); in order to conduct a balance or stopgap between the number of change points and the likelihood, we employ the Schwarz's Bayesian information criterion (BIC) by considering the penalty for large L ; we identify the value of L by minimizing L as follows:

$$BIC_L = -\max R_n(\tau'_1, \dots, \tau'_L) + L\zeta_n \quad (4)$$

where ζ_n represents the proper sequence going to infinity. The BIC has been used in $\zeta_n = \log n$ to determine the number of change points and to show its consistency (Yao and Au 1989).

An efficient nonparametric maximum likelihood approach was adopted in this study to detect multiple change points. The proposed approach does not assume that the data series has any specific parameter structure or distribution characteristics. Therefore, it is suitable for the detection of any changes in data distribution. The number of optimal change points is determined by BIC, and locations of the points can be estimated with the dynamic programming algorithm and maximum likelihood function. This approach has been performed well in previous studies on the recognition of multiple change points

Fig. 2 The number of change points for annual precipitation at each station

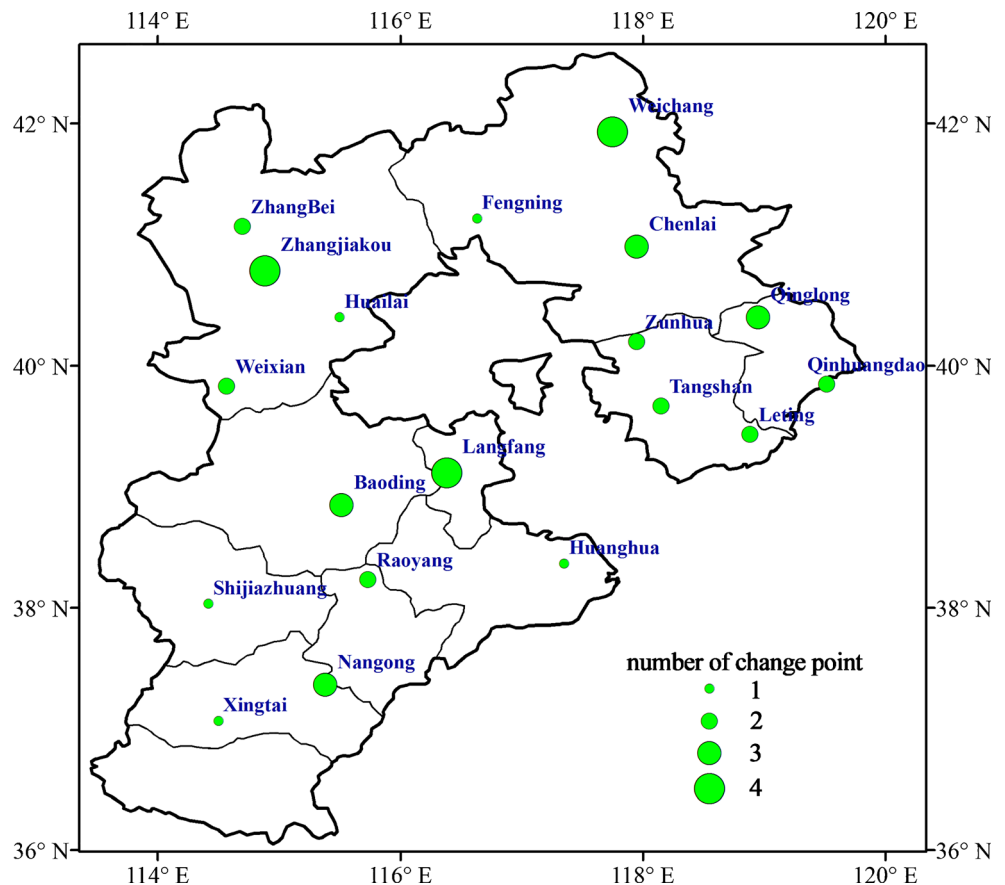
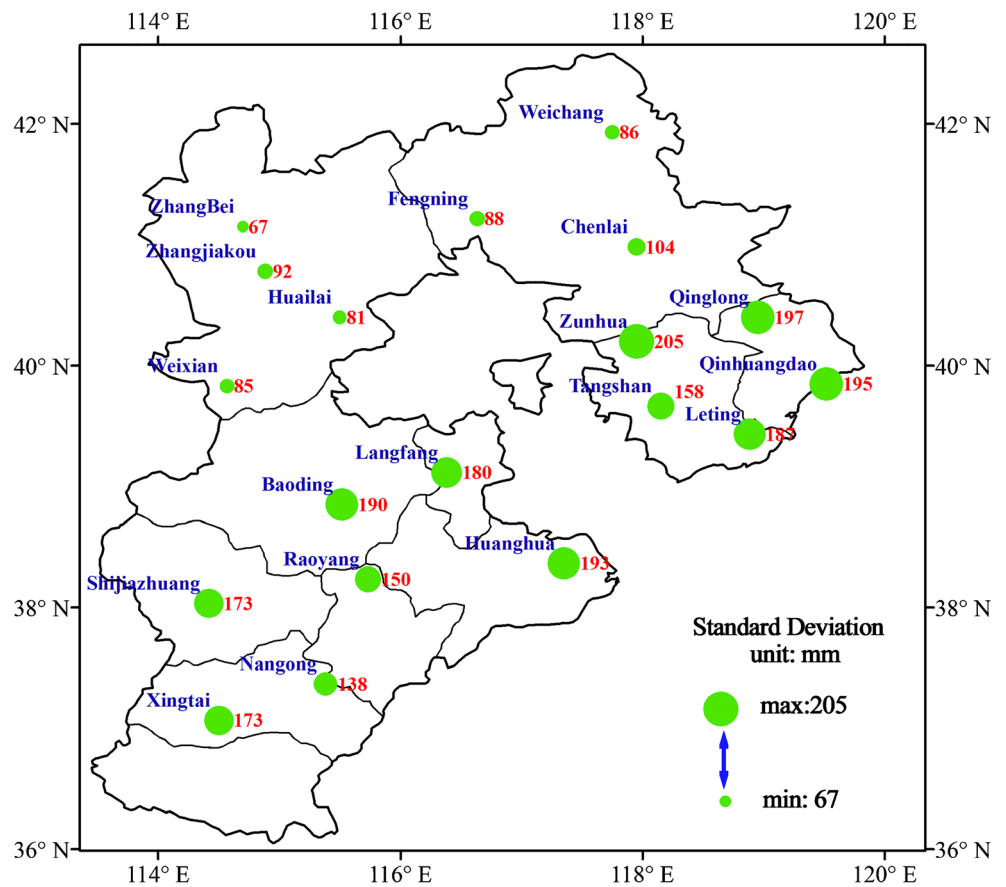


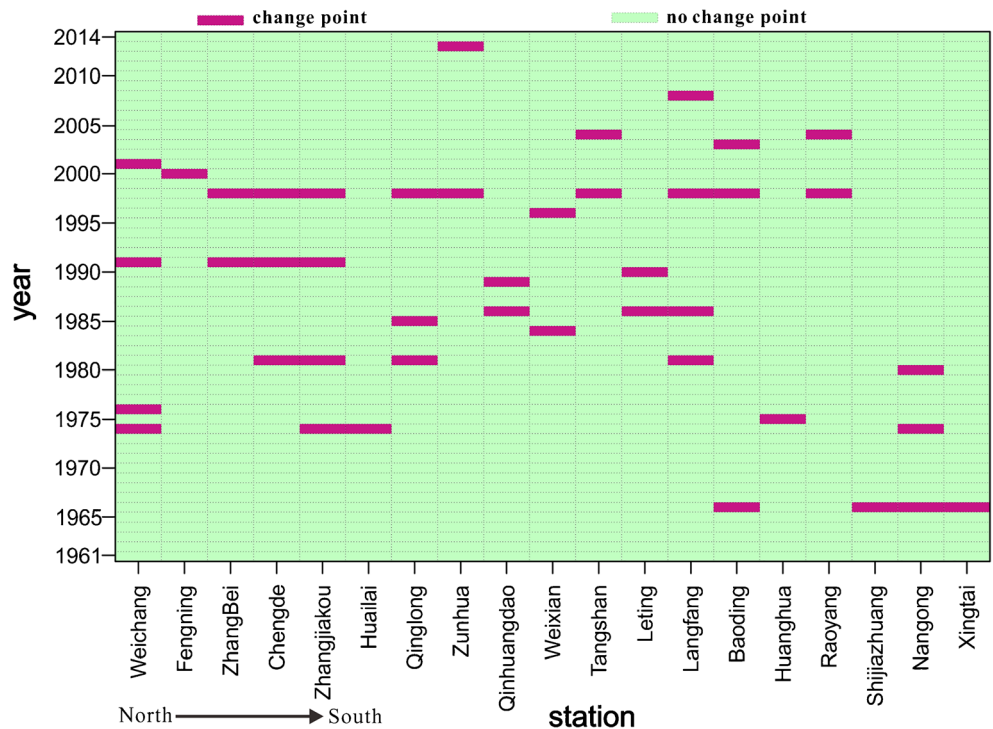
Fig. 3 Standard deviation for annual precipitation at each station



(Zou et al. 2014). The number of change points detected at each station is P ($P > 0$). Using the location of each change

point, the total annual precipitation data series during the period 1961–2014 is divided into $P + 1$ segments.

Fig. 4 Hovmöller map for timing of change points at each station



Trend analysis approach

The Mann-Kendall (MK) trend analysis is a nonparametric test method (Kendall 1948; Mann 1945; Haktanir and Citakoglu 2014; Haktanir and Citakoglu 2015) which does not require the data series to be normally distributed. MK is a proven and effective approach to analyze monotonous change in data series (Burn and Elnur 2002; Xu et al. 2004; Ay and Kisi 2015). The null hypothesis H0 states that the data series $x_k(k = 1, 2, 3 \dots n)$ are independent from one another and have the same distribution; the alternative hypothesis H1 states that there is a monotonic trend in the data series. The MK test statistics are calculated as follows:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \tag{5}$$

where x_j represents the sequential data values, n is the length of the dataset, and sgn is calculated as follows:

$$\text{sgn}(x_j - x_i) = \begin{cases} 1 & \text{if } x_j > x_i \\ 0 & \text{if } x_j = x_i \\ -1 & \text{if } x_j < x_i \end{cases} \tag{6}$$

According to Mann (1945) and Kendall (1948), when $n \geq 8$, the test statistics S is approximately normally distributed with the mean and variance as follows:

$$E(S) = 0$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^n t_m m(m-1)(2m+5)}{18} \tag{7}$$

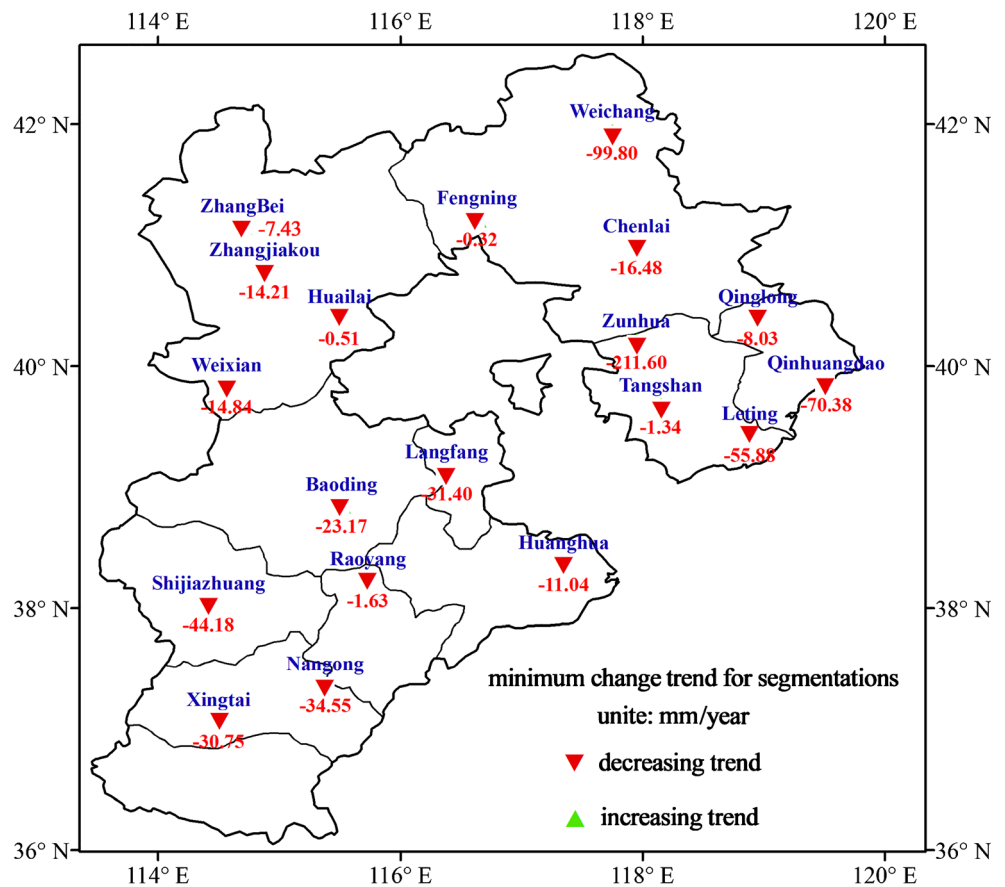
where t_m is the extent m . The standardized test statistics Z are calculated as follows:

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & S < 0 \end{cases} \tag{8}$$

$|Z_{\alpha}| = 1.65, 1.96, \text{ and } 2.58$, which correspond to the critical values at the significance level $P = 0.1, 0.05, \text{ and } 0.01$, respectively. If $|Z| > |Z_{\alpha}|$, the null hypothesis H0 is rejected; $P = 0.05$ and 0.01 were considered in the present research.

To analyze the time series variable trend, we used the robust estimator for the amplitude of trend slopes proposed by Sen (1968):

Fig. 5 minimum trends for segmentations at each station



$$\text{slope} = \text{Median}\left(\frac{x_j - x_i}{j - i}\right) \quad (1 \leq i < j \leq n) \quad (9)$$

where slope is the monotonic increase or decrease rate (i.e. the linear slope) of the entire data series $x_k(k = 1, 2, 3 \dots n)$ or any segmentation $x_w(w = i, i + 1, 3 + 2 \dots j)$. If it is positive, the series monotonically increases; if it is negative, the series monotonically decreases. *Median* indicates that the function takes the median value. Sen’s trend calculation was performed on the entire series and the segmentations (divided by the results of change-point detection) of annual precipitation data at the stations during the period 1961–2014, and conducted significance tests on the Sen’s trend results using the MK approach.

The ordinary least squares (OLS) linear regression analysis is used to derive relationships between variables for time series data that is believed to be uniformly applicable across the study area (Foody 2003). The OLS technique can quantitatively explore the effects of a collection of independent variables on one dependent variable and OLS also can extract change trends of the time series data (Yang et al. 2016). However, if the data do not satisfy some of assumptions, then results can be misleading, especially, outliers violate the assumption of normally distributed residuals in the least-squares regression. Therefore, it is important to assess whether the

time series data at each station conform to the law of normal distribution before the OLS is used; here, we adopt the Q-Q plot and parametric test (such as coefficient of Skewness, coefficient of Kurtosis, and *P* value (Roozbeh and Arashi 2016) to determine the normality of time series data. The U-statistics was constructed based on coefficient of skewness and coefficient of kurtosis respectively, and *P* values were calculated by the K-S test (Liu and Zhang 2016). then U_{Skewness} , U_{Kurtosis} , and *P* values are applied for testing. At least two of the three indicators can prove the normal distribution ($U_{\text{Skewness}} < U_{0.05} = 1.96, U_{\text{Kurtosis}} < U_{0.05} = 1.96, P > 0.05$) (Liu and Zhang 2016); the goodness of fit between the precipitation dataset and the uniform distribution on (0, 1) at each station are also revealed through Q-Q plots and histogram in our study. The OLS fitting formula is as follows:

$$\hat{y}_i = a + bx_i \quad (10)$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (11)$$

$$a = \bar{y} - b\bar{x} \quad (12)$$

Fig. 6 Maximum trends for segmentations at each station

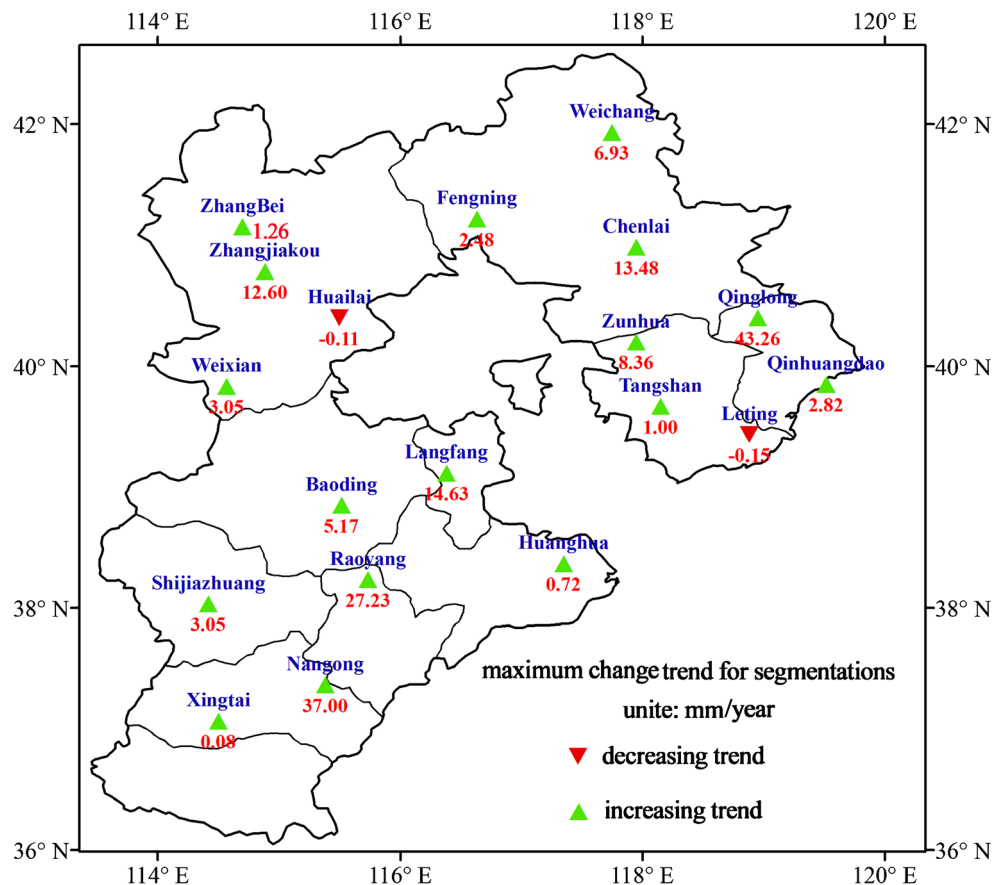


Fig. 7 Generalization trend at each station

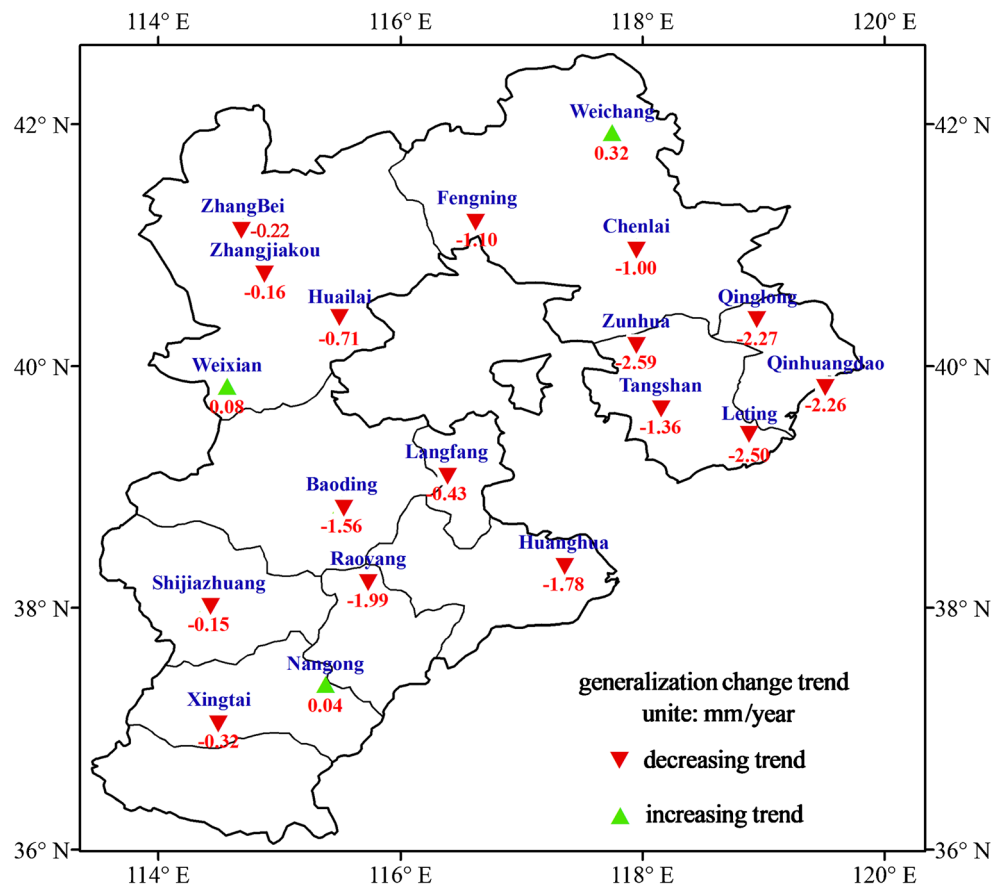


Table 1 The Mann-Kendall trend test at each station F means that significance level of trend is greater than 0.05, T means that significance level of trend is less than 0.05

Station name	Z value for segmentations for each segmentation(from 1961 to 2014)	P < 0.05	Z value for generalization	P < 0.05
Zhangbei	0.76, - 0.87, - 0.37	F	- 0.41	F
Weixian	0.30, - 1.77, 1.41	F	0.08	F
Shijiazhuang	- 0.75, 0.28	F	- 0.12	F
Xingtai	- 1.13, 0.37	F	- 0.36	F
Fengning	- 0.50, 0.79	F	- 1.43	F
Weichang	2.08, - 1.04,0.41, - 1.87, 0.44	T	0.60	F
Zhangjiakou	1.09, - 0.37,1.56, - 0.37,0.54	F	- 0.16	F
Huailai	0.44, - 0.08	F	- 0.88	F
Chenlai	- 0.21, 1.25, - 0.37, 0.29	F	- 1.01	F
Zunhua	- 0.05, 0.50, - 2.00	T	- 1.69	F
Qinglong	- 0.21, 1.22, 0.22, 0.21	F	- 1.50	F
Qinhuangdao	0.44, - 2.38,0.71	T	- 1.42	F
Langfang	0.39, 1.50, - 0.043, 1.71, - 2.10	T	- 0.36	F
Tangshan	0.16, 0.60, - 0.62	F	- 0.95	F
Leting	- 0.31, - 1.71, - 0.40	F	- 1.67	F
Baoding	- 0.75, 0.08, 0.75, 0.21	F	- 0.97	F
Raoyang	- 0.68, 1.80, - 0.47	F	- 1.55	F
Huanghua	0.79, 0.29	F	- 1.19	F
Nangong	- 0.75, 0.52, - 0.90, 0.97	F	- 0.06	F

where y is the dependent variable, x is the independent variable, a which represents the intercept and b which expresses the slope of the relationship between the two variables.

Spatial characteristics

With multiple change-point detection, the number and locations (or dates) of change points at the 19 stations were identified and the series were divided into different segmentations. The quantity of change points reflected the varying degrees of change in precipitation during the study period. This variable was compared against the standard deviation at each station to spatially assess the changes over the time series from different aspects. The location (or date) of each change point marked the year in which precipitation sustained an abrupt change in practice; this information supported engineers in developing appropriate strategies in response to such change. The Hovmöller map was used to illustrate the occurrence of change points at each station in each year (Hovmöller 1949; Wang et al. 2015). Different segmentations of the series have different amplitudes or trends of change.

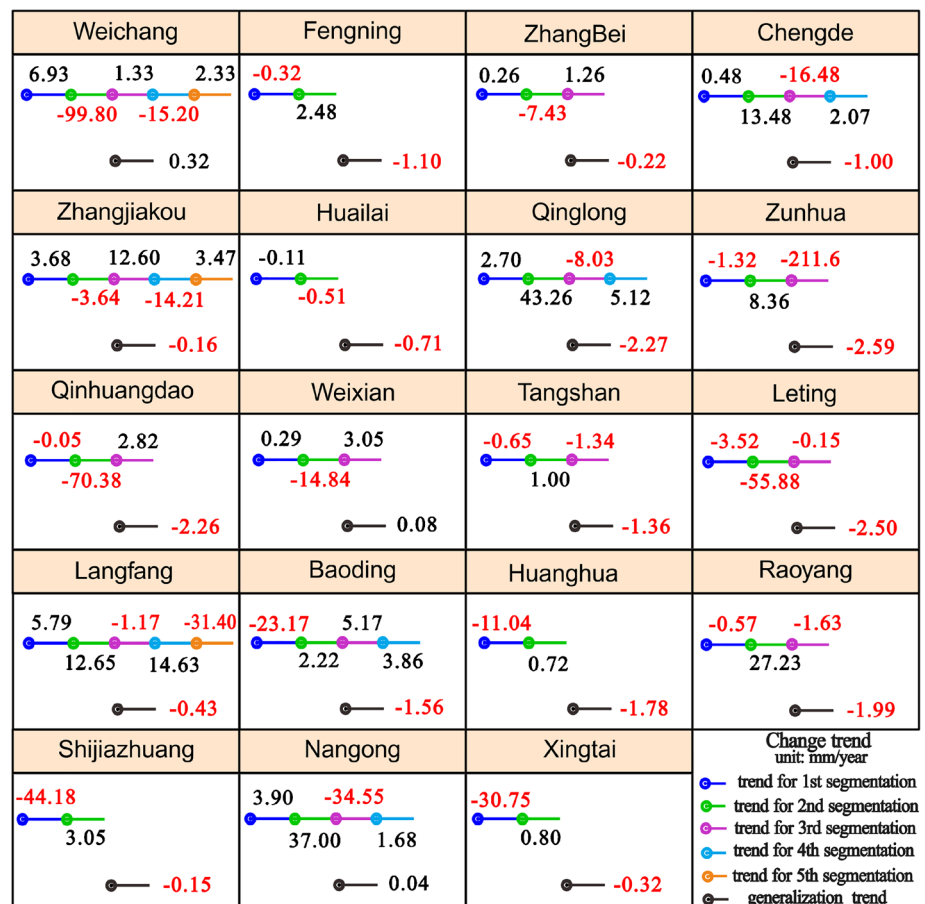
As discussed above, nonparametric MK tests and Sen's trend analysis are applied to each segmentation to determine their respective trend. Spatial comparison was performed on maximum trend, minimum trend, and generalization trend over the entire study period for each station. The automatic segmentation-based trend analysis and generalization trend analysis were compared against each other. Trends in all the segmentations and the entire period were depicted as overlapping on the same spatial-temporal diagram to make clear the magnitude of various changes at each station.

Results and discussion

Variability characteristics of annual precipitation

In this study, annual precipitation variability was characterized by the number of change points and the standard deviation. The number of change points at each station during the period 1961–2014 is shown in Fig. 2. The maximum number (4) occurred at three stations: Zhangjiakou, Weichang, and Langfang. This revealed that the precipitation at these stations

Fig. 8 Change trend for segmentations and generalization at each station



underwent abrupt changes of multiple times in five time segments of the study period. The minimum number of change points (1) occurred at the five stations: Fengning, Huailai, Shijiazhuang, Xingtai, and Huanghua, indicating that precipitation at these stations was relatively stable or underwent only one abrupt change during the study period. The standard deviation for annual precipitation at each station from 1961 to 2014 is shown in Fig. 3. Its spatial distribution is more representative than that shown in Fig. 2; that is, the deviation gradually decreased from more than 190 mm in the southeast to less than 70 mm in the northwest. This might be characteristically the higher precipitation in the southeastern regions and thereby the higher inter-annual change than that in the northwestern regions. The inconsistency between Figs. 2 and 3 in terms of spatial distribution was primarily attributable to the fact that the change points in Fig. 2 represent the precipitation variability across several time sections, while the standard deviation shown in Fig. 3 represents the variability across years. The number of change points and the standard deviation depicted the variability of precipitation on different time scales. Both, however, could reveal extreme climate events or abnormalities related to climate change.

The characteristics of change-point date

To illustrate the dates and locations of change points over different stations, the Hovmöller map was used to depict temporal-spatial characteristics (Fig. 4). The red marks on

the map indicated the occurrence of abrupt change points. Change points were generally obvious around 1974, 1981, and 1998 and occurred in a few stations before 1974 and after 2000. This suggested that precipitation changed at a relatively stable pace during the years 1961–1974 and 2000–2014, but changed abruptly in 1974–2000 during a period in which meteorological disasters were likely to occur. Identical dates in the change points were observed at some adjacent stations due to the proximity effect of local meteorological conditions in terms of precipitation. Figure 4 also shows that the change points were concentrated mostly within Zhangjiakou, Weichang, and Langfang. The Hovmöller map was also applied to the temporal-spatial analysis of other elements as it revealed “hot-spot years,” sites or areas of interest, and the continuity of changing characteristics in terms of a given area or time.

Characteristics of trends

To determine the effect of our multiple change-point detection, temporal-spatial comparison was established between the data in the segmentations divided by the multiple change points and the data over the entire time series; this allowed a clear comparison among minimum trends for segmentations (Fig. 5), maximum trends for segmentations (Fig. 6), and the generalization trend (Fig. 7) at each station. In addition, significance level is shown with the Z value of the MK test (Table 1).

Table 2 The table of normal distribution test at each station

Station name	Skewness	Kurtosis	U_{Skewness}	U_{Kurtosis}	P values	Normal distribution
Zhangbei	0.30	-0.20	0.92 < 1.96	0.31 < 1.96	0.72 > 0.05	T
Weixian	0.11	-0.19	0.34 < 1.96	0.30 < 1.96	0.93 > 0.05	T
Shijiazhuang	1.21	2.20	<i>3.72 > 1.96</i>	<i>3.44 > 1.96</i>	<i>0.00 < 0.05</i>	F
Xingtai	1.60	5.64	<i>4.93 > 1.96</i>	<i>8.83 > 1.96</i>	<i>0.00 < 0.05</i>	F
Fengning	0.42	0.10	1.30 < 1.96	0.15 < 1.96	0.45 > 0.05	T
Weichang	0.44	-0.28	1.37 < 1.96	0.43 < 1.96	0.20 > 0.05	T
Zhangjiakou	0.12	-0.46	0.37 < 1.96	0.72 < 1.96	0.50 > 0.05	T
Huailai	-0.11	-0.44	0.33 < 1.96	0.69 < 1.96	0.93 > 0.05	T
Chenlai	0.07	-0.82	0.21 < 1.96	1.28 < 1.96	0.40 > 0.05	T
Zunhua	0.42	-0.51	1.29 < 1.96	0.80 < 1.96	0.08 > 0.05	T
Qinglong	0.56	-0.70	1.73 < 1.96	1.09 < 1.96	<i>0.01 < 0.05</i>	T
Qinhuangdao	1.10	1.96	<i>3.38 > 1.96</i>	<i>3.07 > 1.96</i>	<i>0.00 < 0.05</i>	F
Langfang	0.88	0.16	<i>2.71 > 1.96</i>	0.26 < 1.96	<i>0.00 < 0.05</i>	F
Tangshan	0.46	-0.09	1.43 < 1.96	0.14 < 1.96	0.33 > 0.05	T
Leting	0.54	0.57	1.66 < 1.96	0.89 < 1.96	0.15 > 0.05	T
Baoding	0.54	-0.71	1.66 < 1.96	1.10 < 1.96	<i>0.01 < 0.05</i>	T
Raoyang	0.79	0.70	<i>2.43 > 1.96</i>	1.09 < 1.96	<i>0.04 < 0.05</i>	F
Huanghua	1.30	3.65	<i>4.00 > 1.96</i>	<i>5.72 > 1.96</i>	<i>0.00 < 0.05</i>	F
Nangong	0.45	0.69	1.38 < 1.96	1.08 < 1.96	0.20 > 0.05	T

The items presented in italic are the time series data that do not conform to normal distribution, and F is abbreviation of FALSE, T is abbreviation of TRUE

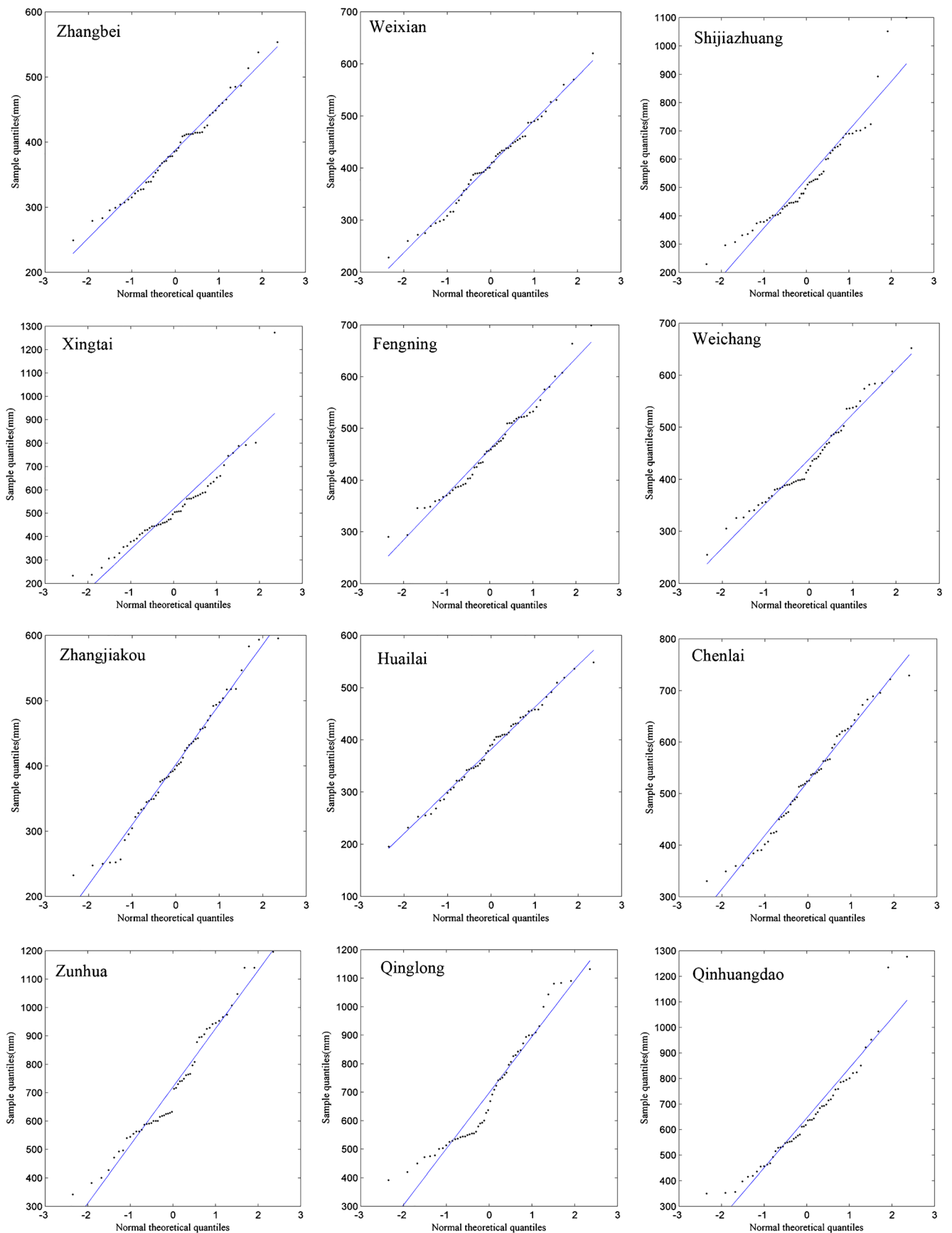


Fig. 9 Q-Q plots to present the fitting degree of the precipitation dataset relative to the uniform distribution on (0, 1) at each station

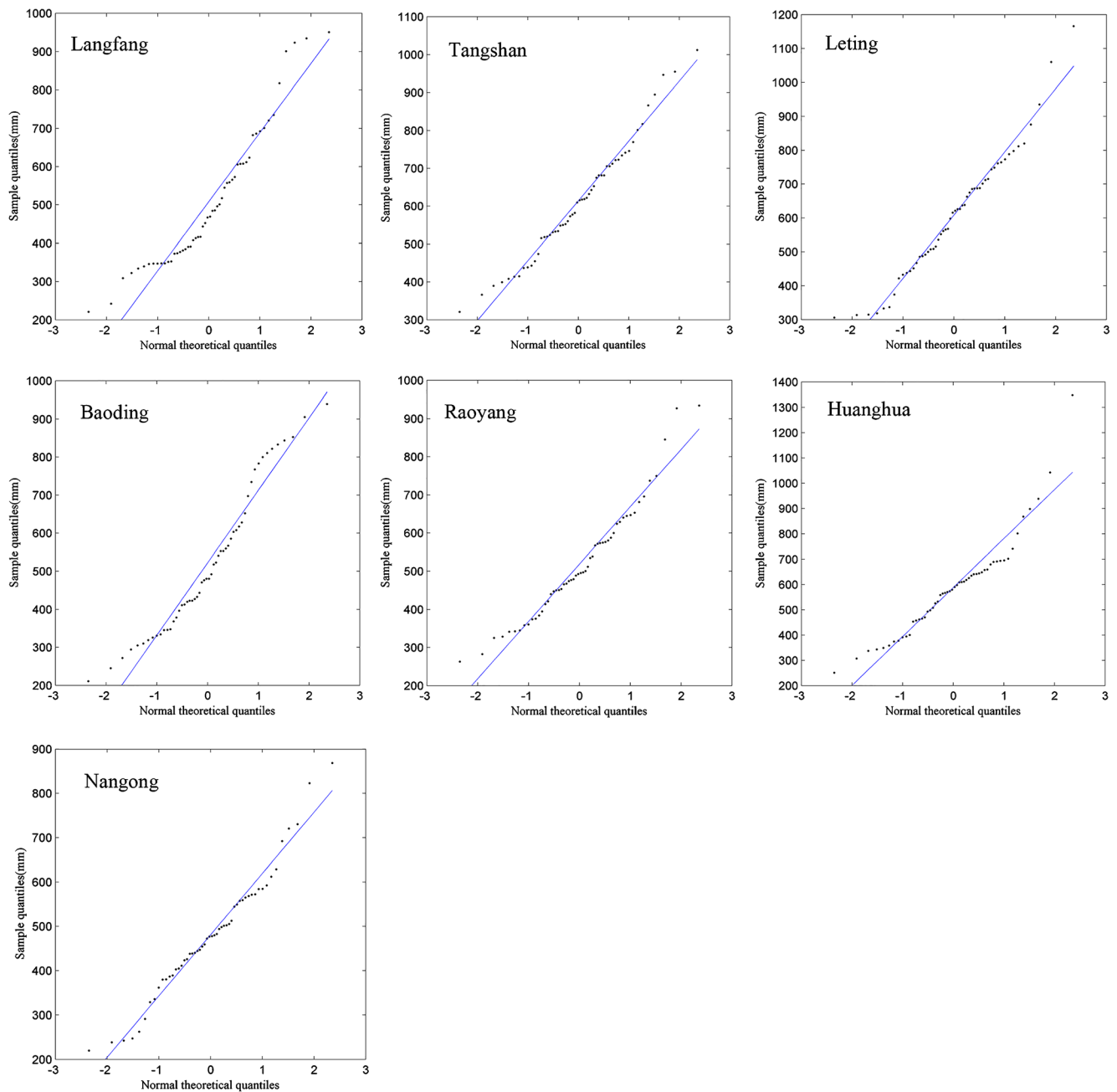


Fig. 9 (continued)

Table 1 reflected that the significance level between generalization trend and segmentation trend has some differences (T means P is less than 0.05, F means P is greater than 0.05). For generalization trend, there are no stations with a P value less than 0.05. However, four stations (Weichang, Zunhua, Qinhuangdao, Langfang) with a P value less than 0.05 for different segmentations were observed which means only the use of generalization trend may possibly mask abrupt change.

Figure 5 shows where the minimum trend for segmentations decreased at all stations. The most obvious decreased in amplitude or trend occurred at Zunhua (-211.60 mm/year), Weichang (-99.80 mm/year), Qinhuangdao (-70.38 mm/

year), and Leting (-55.85 mm/year) stations. In effect, to determine the most abrupt decrease rate in the trend of a given variant, it was feasible to derive the minimum trend indicator from the segmentation-based trend analysis as part of the multiple change-point detection process.

Figure 6 illustrates where the minimum trend for segmentations increased at all stations other than Leting, and most obviously at Qinglong (43.26 mm/year), Nangong (-37.00 mm/year), and Raoyang (27.23 mm/year). Again, deriving the maximum trend indicator from the segmentation-based trend analysis as part of multiple change-point detection appeared to reveal the most abrupt increase rate in the trend of a variant.

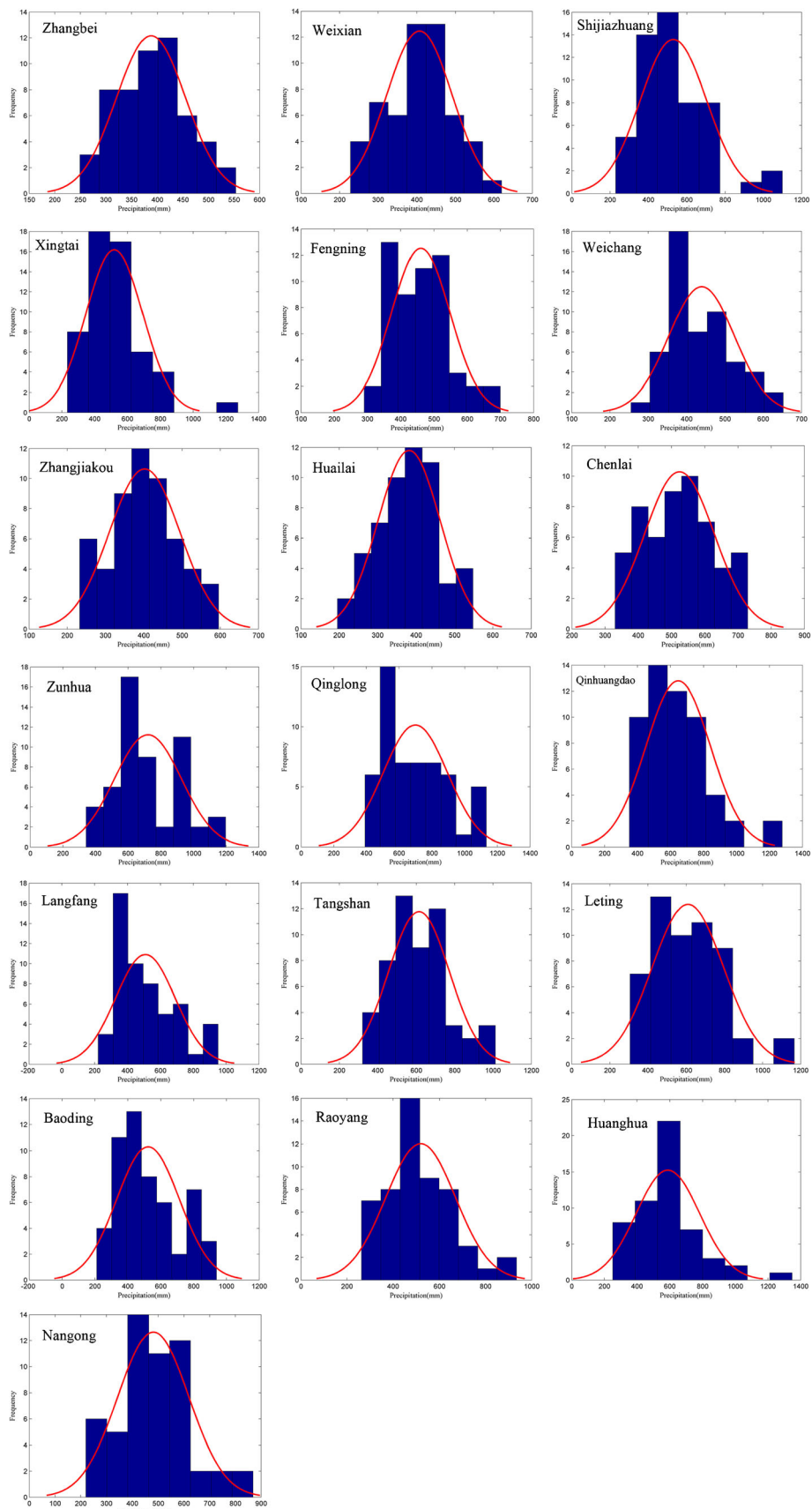


Fig. 10 Histograms with normal curves at each station

Decreasing generalization trends are shown in Fig. 7 over the unsegmented long time data series at all stations other than Weichang, Weixian, and Nangong. The amplitude of any increase or decrease was insignificant at any of 19 stations. The greatest amplitude of increase occurred at Weichang station (0.32 mm/year), and the greatest amplitude of decrease occurred at Zunhua (− 2.56 mm/year), Leting (− 2.5 mm/year), Qinglong (− 2.27), and Qinhuangdao (− 2.26). The absolute change was less than 3 mm/year at all stations. The generalization-based trend analysis effectively smoothed the trend of abrupt changes in specific time sections.

Generalization trends over full time series and for different segmentations are shown in Fig. 8; change points occurred at all 19 stations. There was one change point at Fengning, Huailai, Huanghua, Shijiazhuang, and Xingtai stations, and four (i.e., the maximum) at Weichang, Zhangjiakou, and Langfang. As shown in Fig. 8, the segmentation trends at all stations other than Leting and Qinhuangdao increased and decreased alternatively. The generalization trends at all stations other than Huailai, Tangshan, and Raoyang fell between the minimum and maximum trends on multiple segmentations. This revealed that generalization trends covered the trends characterized by abrupt changes in specific time sections among multiple segmentations. Such a change might be abrupt at a certain time even if the generalization trend was mild.

To compare MK trend tests and Sen's trend analysis, we use the OLS linear regression method to calculate the trends. Table 2 presents the results of the normal distribution test at each station in the study. T represents accept the hypothesis of the normal distribution and F is refuse the hypothesis of normal distribution. We adopt the three parameters U_{Skewness} , U_{Kurtosis} , and P values to assess the normality of time series data, in order to ensure the performance of the assessing results. It is shown that there are two or three test parameters proving normality of time series data at each station; if there is only parameter showing the normality, we think that the time series data do not conform to normal distribution. There are six stations which do not conform to normal distribution, if the OLS method is conducted, the results of fitting or calculating change slope may be wrong. We also used Q-Q plots and histogram to test the normality of time series data. Figures 9 and 10 show consistent results with Table 2; this indicates that the test results of the three methods including parameters, Q-Q plots, and histogram are consistent and effective, and there are Shijiazhuang, Xingtai, Qinhuangdao, Langfang, Raoyang, and Huanghua which do not conform to the normal distribution.

Table 3 presents change trends by using the OLS method; the values of P indicate determination of the linear trends at the significance level of 0.05. There is no station with P less than 0.05 which keep pace with the generalization trend by the Mann-Kendall trend test. However, for the amplitude of change slope, there is only Weichang station with increasing trend (0.33 mm/year) based on the OLS method, which shows

Table 3 The result of linear regression at each station

Station name	Slope(change/ b)	Intercept(a)	T values	P
Zhangbei	− 0.16	391.97	− 0.26	0.79 > 0.05
Weixian	− 0.11	409.71	− 0.14	0.89 > 0.05
Shijiazhuang	− 0.78	550.99	− 0.52	0.61 > 0.05
Xingtai	− 1.64	562.88	− 1.09	0.28 > 0.05
Fengning	− 1.11	489.35	− 1.46	0.15 > 0.05
Weichang	0.33	430.42	0.44	0.66 > 0.05
Zhangjiakou	− 0.37	411.92	− 0.45	0.65 > 0.05
Huailai	− 0.49	394.88	− 0.69	0.49 > 0.05
Chenlai	− 0.87	546.75	− 0.96	0.34 > 0.05
Zunhua	− 2.92	798.01	− 1.66	0.10 > 0.05
Qinglong	− 2.40	761.98	− 1.41	0.16 > 0.05
Qinhuangdao	− 2.23	705.09	− 1.32	0.19 > 0.05
Langfang	− 1.05	536.34	− 0.67	0.51 > 0.05
Tangshan	− 1.95	666.01	− 1.43	0.16 > 0.05
Leting	− 2.36	671.03	− 1.46	0.15 > 0.05
Baoding	− 2.15	579.67	− 1.30	0.20 > 0.05
Raoyang	− 2.22	577.80	− 1.72	0.09 > 0.05
Huanghua	− 3.03	668.35	− 1.83	0.07 > 0.05
Nangong	− 0.47	493.76	− 0.39	0.70 > 0.05

consistency with the result of Sen's trend (0.32 mm/year) analysis, Zunhua (− 2.92 mm/year), Leting (− 2.36 mm/year), and Qinglong (− 2.40 mm/year) have approximate values comparing to the trend slopes calculated by Sen's trend analysis, respectively. The absolute change was less than 3 mm/year at all stations whose results are the same as those in Fig. 7 and the slopes in the two kinds of methods are inconsistent in most stations. The Mann-Kendall Sen trend and the linear regression trend are summarized in Table 4; there no one station with a significant trend ($p < 0.05$). The Sen slope method does not need the data to be a normal distribution and trends of OLS failure to pass the significance level test; therefore, we can draw the conclusion that the result of MK trend tests have a better confidence degree which indicates a further application.

Summary and conclusion

Multiple change-point detection and segmentation-based trend analysis are crucial techniques for the recognition of climate abnormalities and the monitoring and evaluation of environmental impact within the context of global climate change, as well as in research on the spatial-temporal characteristics of other subjects. Hebei Province, one of the major grain-producing regions in China, was served as the focus of the present study. The annual precipitation data at 19 meteorological stations in the study area during the period 1961–2014 were assessed. The nonparametric maximum likelihood

Table 4 A summary of the results of the Mann-Kendall test and linear regression trend test

19 record series	Mann-Kendall Sen trend			Linear regression trend of OLS		
	No.	Increasing	Decreasing	No.	Increasing	Decreasing
Number	0	3	16	0	1	18
Percentage (%)	0	15.8	84.2	0	5.3	94.7

No. is the number or percentage of station with significance $P < 0.05$

approach was applied and the number of change points at each station was optimized based on BIC. The lengthy time series was segmented based on the identified change points. MK trend tests and Sen's amplitude analysis were conducted for segmentations at each station and validated by comparing with basic ordinary least squares (OLS) regression analysis based on generalization trends. The proposed approach represents an effective combination of nonparametric approaches including the maximum likelihood, MK trend tests, Sen's trend analysis, and basic ordinary least squares (OLS) regression analysis. This approach enabled flexible and efficient multiple change-point detection and research on multi-segment trend characteristics. As reported above, the characteristics of the temporal-spatial pattern were revealed using the number of change points, standard deviation, timing of change points, minimum and maximum trends for segmentations, and generalization trends over the entire study period.

The standard deviation of annual precipitation can represent the variability of precipitation across several years, and the number of change points can represent the variability across several time segments. Change points in the dataset were generally obvious around 1974, 1981, and 1998, and at few stations in the study area before 1974 and after 2000; precipitation changes were relatively stable during the periods 1961–1974 and 2000–2014. The minimum trend for segmentations decreased at all stations but most significantly at Zunhua (− 211.60 mm/year), Weichang (− 99.80 mm/year), Qinhuangdao (− 70.38 mm/yea), and Leting (− 55.85 mm/year). The maximum trend for segmentations increased at all stations other than Leting, and most significantly at Qinglong (43.26 mm/year), Nangong (− 37.00 mm/year), and Raoyang (27.23 mm/year). The generalization trends were characterized by absolute change below 3 mm/year at all stations, and fell between the minimum and maximum trends on multiple segmentations at all but three stations. It also covered the trends characterized by abrupt changes in specific time sections among the segmented series.

The result derived from the proposed approach can better characterize the trend of climate variables over certain parts. The proposed approach represents a reference tool for regional water resource management and novel strategies to combat the effects of climate change. This approach can yield even more meaningful results if applied to the time series analysis on larger spatial scales and across more disciplines, such as

quantitative evaluations of economic management practices, forest protection initiatives, and environmental and social development.

Acknowledgments We would like to thank Guangyu Li for providing helpful support.

Funding information This research received financial support from the National Natural Science Foundation of China (No. 41601562), also sponsored by China Scholarship Council, the Research Project for Young Teachers of Fujian Province (No. JAT160085), and the Scientific Research Foundation of Fuzhou University (No. XRC-1536).

References

- Alexander L, Zhang X, Peterson T, Caesar J, Gleason B, Klein Tank A, Haylock M, Collins D, Trewin B, Rahimzadeh F (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res Atmos* 111. <https://doi.org/10.1029/2005JD006290>
- Ay M, Kisi O (2015) Investigation of trend analysis of monthly total precipitation by an innovative method. *Theor Appl Climatol* 120(3-4):617–629
- Burn DH, Elnur MAH (2002) Detection of hydrologic trends and variability. *J Hydrol* 255:107–122
- Chen S, Li Y, Kim J, Kim SW (2016) Bayesian change point analysis for extreme daily precipitation. *Int J Climatol* 37:3123–3137. <https://doi.org/10.1002/joc.4904>
- Crujeiras RM, Van Keilegom I (2010) Least squares estimation of non-linear spatial trends. *Comput Stat Data Anal* 54:452–465
- Djaman K, Balde AB, Rudnick DR, Ndiaye O, Irmak S (2016) Long-term trend analysis in climate variables and agricultural adaptation strategies to climate change in the Senegal River Basin. *Int J Climatol* 37:2873–2888. <https://doi.org/10.1002/joc.4885>
- Einmahl JH, McKeague IW (2003) Empirical likelihood based hypothesis testing. *Bernoulli* 9:267–290
- Fan L, Lu C, Yang B, Chen Z (2012) Long-term trends of precipitation in the North China Plain. *J Geogr Sci* 22:989–1001
- Fealy R, Sweeney J (2005) Detection of a possible change point in atmospheric variability in the North Atlantic and its effect on Scandinavian glacier mass balance. *Int J Climatol* 25:1819–1833
- Fischer A, Weigel A, Buser C, Knutti R, Künsch H, Liniger M, Schär C, Appenzeller C (2012) Climate change projections for Switzerland based on a Bayesian multi-model approach. *Int J Climatol* 32:2348–2371
- Foody GM (2003) Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI–rainfall relationship. *Remote Sens Environ* 88:283–293
- Frich P, Alexander L, Della-Marta P, Gleason B, Haylock M, Tank AK, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. *Clim Res* 19:193–212

- Fu X, Kuo C, Gan T (2015) Change point analysis of precipitation indices of Western Canada. *Int J Climatol* 35:2592–2607
- Haktanir T, Citakoglu H (2014) Trend, Independence, Stationarity, and Homogeneity Tests on Maximum Rainfall Series of Standard Durations Recorded in Turkey. *J Hydrol Eng* 20(10):770–776. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000973](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000973) 05014009
- Haktanir T, Citakoglu H (2015) Closure to "Trend, Independence, Stationarity, and Homogeneity Tests on Maximum Rainfall Series of Standard Durations Recorded in Turkey" by Tefaruk Haktanir and Hatice Citakoglu. *J Hydrol Eng* 20(10):07015017. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001246](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001246).
- Hovmöller E (1949) The Trough-and-Ridge diagram. *Tellus*. 1:62–66
- Ip WC, Zhang L, Wong H, Xia J (2011) Multi-scale variability and trends of precipitation in North China. *Water Res* 38:18–28
- Jamali S, Jönsson P, Eklundh L, Ardö J, Seaquist J (2015) Detecting changes in vegetation trends using time series segmentation. *Remote Sens Environ* 156:182–195
- Kendall MG (1948) Rank correlation methods. Charles Griffin, London, p 202
- Khaliq M, Ouarda T, St-Hilaire A, Gachon P (2007) Bayesian change-point analysis of heat spell occurrences in Montreal, Canada. *Int J Climatol* 27:805–818
- Liu X, Zhang Q (2016) Analysis of the return period and correlation between the reservoir-induced seismic frequency and the water level based on a copula: A case study of the Three Gorges reservoir in China. *Phys Earth Planet Inter* 260:32–43
- Liu J, Liu M, Zhuang D, Zhang Z, Deng X (2003) Study on spatial pattern of land-use change in China during 1995–2000. *Sci China Ser D Earth Sci* 46:373–384
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13: 245–259
- Mihailović D, Lalić B, Drešković N, Mimić G, Djurdjević V, Jančić M (2015) Climate change effects on crop yields in Serbia and related shifts of Köppen climate zones under the SRES-A1B and SRES-A2. *Int J Climatol* 35:3320–3334
- Nouri M, Homae M, Bannayan M (2017) Quantitative Trend, Sensitivity and Contribution Analyses of Reference Evapotranspiration in some Arid Environments under Climate Change. *Water Resour Manag* 31:2207–2224
- Panda DK, Kumar A, Singandhupe R, Sahoo N (2013) Hydroclimatic changes in a climate-sensitive tropical region. *Int J Climatol* 33: 1633–1645
- Rahmani V, Hutchinson SL, Harrington JA Jr, Hutchinson J, Anandhi A (2015) Analysis of temporal and spatial distribution and change-points for annual precipitation in Kansas, USA. *Int J Climatol* 35: 3879–3887
- Razmi A, Golian S, Zahmatkesh Z (2017) Non-Stationary Frequency Analysis of Extreme Water Level: Application of Annual Maximum Series and Peak-over Threshold Approaches. *Water Resour Manag* 31:2065–2083
- Ribeiro S, Caineta J, Costa AC (2016) Review and discussion of homogenisation methods for climate data. *Phys Chem Earth* 94:167–179
- Roosbeh M, Arashi M (2016) Least-trimmed squares: asymptotic normality of robust estimator in semiparametric regression models. *J Stat Comput Simul* 87:1130–1147
- Ruggieri E (2013) A Bayesian approach to detecting change points in climatic records. *Int J Climatol* 33:520–528
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63:1379–1389
- Sharma S, Swayne DA, Obimbo C (2016) Trend analysis and change point techniques: a survey. *Energy Ecol Environ* 1:123–130
- Suhaila J, Yusop Z (2018) Trend analysis and change point detection of annual and seasonal temperature series in Peninsular Malaysia. *Meteorog Atmos Phys* 130:565–581
- Wang Q, Wu J, Lei T, He B, Wu Z, Liu M, Mo X, Geng G, Li X, Zhou H (2014) Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. *Quat Int* 349:10–21
- Wang Q, Shi P, Lei T, Geng G, Liu J, Mo X, Li X, Zhou H, Wu J (2015) The alleviating trend of drought in the Huang-Huai-Hai Plain of China based on the daily SPEI. *Int J Climatol* 35:3760–3769
- Wang Q, Wu J, Li X, Zhou H, Yang J, Geng G, An X, Liu L, Tang Z (2017) A comprehensively quantitative method of evaluating the impact of drought on crop yield using daily multi-scale SPEI and crop growth process model. *Int J Biometeorol* 349:10–21. <https://doi.org/10.1007/s00484-016-1246-2>
- Wang Q, Tang J, Zeng J, Qu Y, Zhang Q, Shui W, Wang W, Yi L, Leng S (2018) Spatial-temporal evolution of vegetation evapotranspiration in Hebei Province, China. *J Integr Agric* 17:60345–60347. [https://doi.org/10.1016/S2095-3119\(17\)61900-2](https://doi.org/10.1016/S2095-3119(17)61900-2)
- Warner K, Ranger N, Surminski S, Arnold M, Linnerooth-Bayer J, Michel-Kerjan E, Kovacs P, Herweijer C (2009) Adaptation to climate change: Linking disaster risk reduction and insurance. United Nations International Strategy for Disaster Reduction, Geneva
- Wu S, Yin Y, Zhao D, Huang M, Shao X, Dai E (2010) Impact of future climate change on terrestrial ecosystems in China. *Int J Climatol* 30: 866–873
- Xu Z, Chen Y, Li J (2004) Impact of climate change on water resources in the Tarim River basin. *Water Resour Manag* 18:439–458
- Yang X, Wang S, Zhang W, Zhan D, Li J (2016) The impact of anthropogenic emissions and meteorological conditions on the spatial variation of ambient SO₂ concentrations: A panel study of 113 Chinese cities. *Sci Total Environ* 584-585:318–328
- Yao YC, Au ST (1989) Least-squares estimation of a step function. *SankhyaSer:A* 51:370–381
- Zenklusen Mutter E, Blanchet J, Phillips M (2010) Analysis of ground temperature trends in Alpine permafrost using generalized least squares. *J Geophys Res Earth Surf* 115(F04009):79–93
- Zhang XC (2013) Verifying a temporal disaggregation method for generating daily precipitation of potentially non-stationary climate change for site-specific impact assessment. *Int J Climatol* 33:326–342
- Zou C, Yin G, Feng L, Wang Z (2014) Nonparametric maximum likelihood approach to multiple change-point problems. *Ann Stat* 42: 970–1002