



# Linear trends in temperature extremes in China, with an emphasis on non-Gaussian and serially dependent characteristics

Cheng Qian<sup>1,2</sup> · Xuebin Zhang<sup>3</sup> · Zhen Li<sup>1</sup>

Received: 3 July 2018 / Accepted: 19 December 2018 / Published online: 26 December 2018  
© The Author(s) 2018

## Abstract

Record-breaking hot and cold extremes have occurred in China in recent years and, therefore, it is compelling to investigate the long-term trend in temperature extremes at individual stations to see whether they have become more frequent. Many previous studies on the linear trend analysis of temperature extremes in China have used ordinary least squares (OLS) regression, without consideration of non-Gaussian and/or serially dependent characteristics, or nonparametric methods, again not considering the latter, thus leaving some uncertainty in the significance testing. The present study examines in detail these characteristics in eight commonly used extreme temperature indices, on the basis of both station data and gridded data across China. The results show that 71–100% of the stations or grids cannot directly use standard OLS regression to analyze the statistical significance of the linear trend, because of either non-Gaussian or Gaussian but serially dependent characteristics in the regression residuals. Also, more than 43% of the stations and more than 54% of the grid boxes for annual indices cannot directly use the original Sen's slope estimator and Mann–Kendall test because of serial dependence. Based on a nonparametric method that takes into account serial dependence, the spatial patterns of the linear trend on an annual basis, as well as in hot and cold extremes, are examined for the period 1960–2017. The results show that hot extremes at most stations have increased, more than 57% of which are statistically significant; whereas, cold extremes at almost all stations have decreased, more than 32% (85%) of which are statistically significant during daytime (at night).

**Keywords** Temperature extremes · Linear trend · Statistical significance · Non-Gaussian · Serial dependence

## 1 Introduction

In recent years, China has frequently witnessed record-breaking temperature extremes. For example, in 2013, extreme summer heat occurred in East China (Sun et al. 2014; Zhou et al. 2014; Qian 2016). In 2017, high extreme temperature in East China broke the record again; for instance, the maximum temperature reached 40.9 °C at Xuji-ahui station in Shanghai on 21 July—the highest temperature recorded in 145 years of observations. In contrast, extreme

low temperatures have also occurred. In January 2016, a record-breaking cold event occurred in eastern China, and Guanzhou in South China witnessed its first snowfall since 1951 (Qian et al. 2018). In January 2018, a cold extreme event devastated China again, with 108 counties and cities reaching the standard of cold extremes in terms of minimum temperature, four of which broke their low-temperature record. Some researchers have suggested a link between more cold extremes in mid-latitude Eurasia and recent Arctic warming amplification (e.g., Mori et al. 2014; Kug et al. 2015), although this is still a matter for debate (e.g., Barnes 2013; Francis 2017). Given the relationship between extreme temperatures and human mortality, local economics and public services, and crop safety, it is compelling to investigate the long-term trend in temperature extremes at individual stations in China.

To obtain a reliable long-term trend, appropriate statistical techniques are needed. Trend analysis is common practice in climate change studies; however, the misuse of a statistical technique can render the analysis meaningless, and/

✉ Cheng Qian  
qianch@tea.ac.cn

<sup>1</sup> Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, P. O. Box 9804, Beijing 100029, China

<sup>2</sup> University of Chinese Academy of Science, Beijing, China

<sup>3</sup> Climate Research Division, Environment and Climate Change Canada, Toronto, ON, Canada

or result in wrong conclusions (Zhang and Zwiers 2004). Many commonly used statistical methods are based on certain assumptions, and so it is important to check whether these assumptions are met when applying these methods.

Ordinary least squares (OLS) regression is the most commonly used linear trend estimator (IPCC 2013). Many previous studies on the estimation of the linear trend in temperature extremes in China have used OLS regression to estimate the spatial pattern of linear trends at individual stations, and used the Student's *t*-test or *F*-test to assess the corresponding statistical significances (e.g., Ding et al. 2010; Huang et al. 2010, 2015; Wang et al. 2012, 2018; Du et al. 2013; Zhao et al. 2013; Ye et al. 2014; Ding and Ke 2015; Zhou et al. 2016; Liu et al. 2018). Some studies have even used OLS regression and the Student's *t*-test to estimate the spatial pattern of precipitation extreme indices in high-resolution grids (e.g., Zhou et al. 2016) or at individual stations (e.g., Du et al. 2013; Zhao et al. 2013; Liu et al. 2018). A prerequisite of using the Student's *t*-test is that the data being tested follow a Gaussian distribution and, under this circumstance, the test statistic follows Student's *t*-distribution (Wilks 2011, p. 141). The statistical inference of OLS regression with the Student's *t*-test assumes that the regression residuals (errors) are independent Gaussian random variables with a zero mean and constant variance (referred to as standard OLS). In cases where this assumption is not met—for example, if the regression residuals have long-tailed distributions, to which the confidence interval is sensitive (e.g., Hogg 1979)—the inference is unlikely to be reliable, and thus the confidence intervals as well as the associated statistical significance of the OLS trend will not be appropriate (von Storch and Zwiers 1999; Wilks 2011). Likewise, if the Gaussian assumption is met but the independent assumption is not met, the statistical significance will again not be appropriate.

Some studies have combined OLS regression with the nonparametric Mann–Kendall test (Mann 1945; Kendall 1955) to analyze the spatial pattern of the linear trend and the corresponding statistical significance for temperature extremes in China (e.g., Qian and Lin 2004; Zhou and Ren 2011; Qian et al. 2011b; Jiang et al. 2012; Chen et al. 2018; Shi et al. 2018). Although the Mann–Kendall test does not make assumptions about the underlying distribution of the data being tested, it does assume the target data are serially independent (Kendall 1955), which is not always the case. In addition, although OLS provides an unbiased and consistent estimate for the regression coefficient as long as the data have finite variance, it is sensitive to outliers—especially those at the ends of the data series, which can have a big influence on the trend estimate, since by definition it minimizes the square errors (von Storch and Zwiers 1999; Wilks 2011). Thus, the linear trend best-estimate may lack robustness when using OLS regression to analyze data with outliers (Wilks 2011, Fig. 3.16d); for example, some of the

record-breaking hot summer extremes in East China in 2013 (Qian 2016).

Some studies have used nonparametric Kendall's tau-based Sen–Theil estimator, also known as Sen's (1968) slope estimator, to estimate the spatial pattern of the linear trend in climate extremes in China, along with the nonparametric Mann–Kendall test to assess the corresponding statistical significances for station data (e.g., Zhai and Pan 2003; You et al. 2013; Chen and Zhai 2017; Lin et al. 2017) or gridded data (e.g., Yin et al. 2015). Both methods do not make assumptions about the underlying distribution of the climate indices. Sen's slope estimator is the median of all possible slopes, so it is a robust tool. However, both Sen's slope estimator and Mann–Kendall test assume the target data are serially independent (Sen 1968; Kendall 1955).

Although the probability density function of daily temperature tends to be approximately Gaussian (Cubasch et al. 2013, Fig. 1.8), indices that are used to describe extreme temperatures are theoretically unlikely to follow a Gaussian distribution. For example, percentile-based indices such as the annual total number of days that daily maximum temperature is above its 90th percentile (TX90p) will follow a binomial distribution  $B(n, p)$ , with  $n = 365$  and  $p = 0.1$ , if independence among the days holds true. This binomial distribution can be approximated well by a Gaussian distribution. However, daily temperatures are highly persistent over time, and thus it is unclear if a Gaussian distribution can be used to approximate a temperature percentile index. This is especially the case when one is interested in the seasonal values of the indices, for which  $n = 90$ . The annual maximum (or minimum) values of daily maximum (or minimum) temperatures (TXx or TNn) are also used to characterize temperatures. They are unlikely to follow a Gaussian distribution. According to extreme value theory, these values should converge to a generalized extreme value (GEV) distribution if they are sampled from a sufficiently large data block. However, as daily temperatures often follow a Gaussian distribution (e.g., Cubasch et al. 2013, Fig. 1.8), extremes sampled from a Gaussian distribution converge to a GEV quite slowly. As the extreme values only occur within a short seasonal window (for example, annual minimum daily temperature only occurs in the cold season at mid-to-high latitudes), the proper distributional forms for these annual extremes are not easy to determine. At small spatial scales, such as a station or a grid, or for short data lengths—in China, typically 50–60 years of observations—the central limit theorem may not work either, because the sample size is small. In addition, the indices of extreme temperature may also be serially dependent. As a result, the estimate of the confidence interval of a trend may be too narrow if serial dependence is not properly addressed, resulting in possible false detection of a significant trend (von Storch and Zwiers 1999). The studies reviewed above on the linear trend

analysis of temperature extremes failed to consider either the non-Gaussian or serial dependent characteristics, thus leaving some uncertainties.

This study has two main parts. Firstly, we examine the distribution and serial independence throughout China of the linear trend residuals of the eight commonly used extreme temperature indices as defined by the World Meteorological Organization Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al. 2011). To the best of our knowledge, this is the first time that this has been done, and the findings help us to determine the appropriate method for estimating the linear trends, as well as testing the statistical significance of the trends. Accordingly, in the second part of the study, we then compute the spatial patterns of the linear trends in temperature extremes using this method for the data updated to 2017.

## 2 Data and methods

### 2.1 Station data

Homogenized data are important for climate change analysis, especially in China, where station relocations are frequent (Xu et al. 2013; Ren and Zhou 2014; Yan et al. 2014). The data used in this study are the daily maximum temperature ( $T_{max}$ ) and daily minimum temperature ( $T_{min}$ ) for the period 1960–2017 updated from the CHTM4.0 dataset, which is the next version of CHTM3.0 (Li et al. 2016). The dataset was homogenized using the Multiple Analysis of Series for Homogenization method (Szentimrey 1999). There are 758 national Reference Climatic and Basic Meteorological Stations used in this study, not including stations Shapingba (57516) and Changshou (57520), who have missing values for the entire month of April 2017.

### 2.2 HadEX2 gridded data

To illustrate the potential non-Gaussian and serially dependent characteristics in gridded data, the commonly used HadEX2 (the gridded land-based dataset of indices

of temperature and precipitation extremes) covering the period 1901–2010 (Donat et al. 2013) is adopted. This dataset includes a set of temperature and precipitation indices calculated based on high-quality in situ station observations across the globe using a consistent approach recommended by the ETCCDI (Zhang et al. 2011). These index data are on  $3.75^\circ \times 2.5^\circ$  grids.

### 2.3 Calculation of the extreme temperature indices for station data

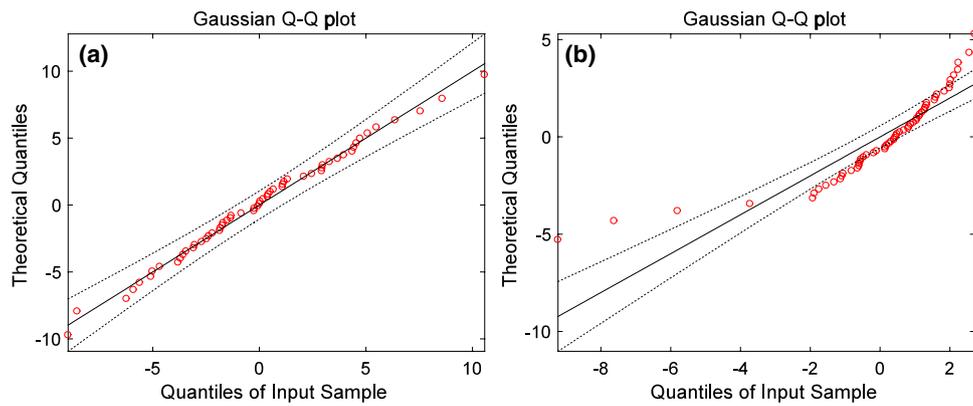
A set of eight extreme temperate indices (Table 1) are analyzed. All these indices are adopted directly from the ETC-CDI (Zhang et al. 2011; also see [http://etccdi.pacificclimate.org/list\\_27\\_indices.shtml](http://etccdi.pacificclimate.org/list_27_indices.shtml)), and have been widely used in the literature (e.g., Alexander et al. 2006; Zhang et al. 2011; Donat et al. 2013). To ensure consistency in the calculation of the indices with other regions, the RClimDex version 1.1 software packages (Zhang and Yang 2004) are used. The percentiles, required for some of the temperature indices, are calculated from the base period of 1961–1990 using a bootstrapping method to avoid possible inhomogeneities (Zhang et al. 2005). The same base period of 1961–1990 is used, as recommended by the ETCCDI, because using different base periods would result in different mean annual cycles and anomalies, thus make the results difficult to compare with others (Qian et al. 2011a).

### 2.4 Methods for linear trend estimation and significance testing

The most commonly used method for linear trend estimation is OLS regression. The statistical inference of the confidence interval of the standard OLS trend assumes that the regression residuals are independent, identically Gaussian distributed random variables. We therefore test the normality of the residuals first. Gaussian quantile–quantile (Q–Q) plotting with 95% confidence intervals (Fig. 1) is used to test whether the OLS regression residuals of the extreme temperature index at each station or HadEX2 grid box is Gaussian distributed. This testing method does not assume serial independence. If all the

**Table 1** Definitions of the eight extreme temperature indices analyzed

Abbreviation	Index name	Definition	Units
TXx	Hottest day	Monthly maximum value of daily $T_{max}$	°C
TNx	Warmest night	Monthly maximum value of daily $T_{min}$	°C
TXn	Coldest day	Monthly minimum value of daily $T_{max}$	°C
TNn	Coldest night	Monthly minimum value of daily $T_{min}$	°C
TX90p	Warm days	Percentage of days when daily $T_{max} > 90$ th percentile	%
TN90p	Warm nights	Percentage of days when daily $T_{min} > 90$ th percentile	%
TX10p	Cold days	Percentage of days when daily $T_{max} < 10$ th percentile	%
TN10p	Cold nights	Percentage of days when daily $T_{min} < 10$ th percentile	%



**Fig. 1** Examples of normality testing by Gaussian Q–Q plots: **a** Gaussian case for the OLS regression residual of the annual TN90p index at Beijing (54511) station; **b** non-Gaussian case for the OLS regression residual of the annual TNn index at Fangxian (57259) station. Red circles indicate the distribution of the target index, and

black solid lines represent the Gaussian distribution, with the 95% confidence interval shown as black dashed lines. The approximate linearity of the circles (all within the confidence interval) suggests that the target data are normally distributed

points of the testing data fall within the 95% confidence intervals, we consider the data as Gaussian distributed (Fig. 1a). Otherwise, it is non-Gaussian (Fig. 1b). It should be mentioned here that this type of Gaussian distribution we classified does not necessarily mean it is perfectly Gaussian; rather, it can be regarded as quasi-Gaussian. If the result is non-Gaussian, the standard OLS method is not appropriate. For each station or grid box having Gaussian distributed residuals, the first-order autocorrelation [hereafter AR(1)] of the OLS regression residuals for the extreme temperature index is further estimated to see whether these residuals are independent. This is because the statistical significance of a standard OLS trend is estimated using the Student’s *t*-test with  $N - 2$  degrees of freedom under the assumption of an independent regression residual. When the AR(1) of the OLS regression residual (hereafter  $r_1$ ) is larger than zero, this assumption is violated and the effective degrees of freedom is reduced to  $N_e - 2$  (Santer et al. 2008), where  $N_e$  is the effective sample size for data and is expressed as (Hartmann et al. 2013):

$$N_e = \begin{cases} N \frac{1-r_1}{1+r_1}, & r_1 > 0 \\ N, & r_1 \leq 0 \end{cases} \quad (1)$$

The significance testing method is then modified to allow AR(1) in the regression residual  $\hat{\epsilon}(t)$  (Santer et al. 2008; Hartmann et al. 2013):

$$\hat{\sigma}_b = \left[ \frac{\sum_{t=1}^N \hat{\epsilon}(t)^2}{N_e - 2} \bigg/ \sum_{t=1}^N (t - \bar{t})^2 \right]^{\frac{1}{2}}, \quad t = 1, \dots, N; \quad (2)$$

$$b = \hat{b} \pm q \hat{\sigma}_b, \quad (3)$$

Here,  $\hat{\sigma}_b$  is the variance of the trend slope estimator;  $b$  is the regression coefficient, with a probability level  $p$  (for example, 95%) confidence interval;  $\hat{b}$  is the best estimate of the trend slope; and  $q$  is the  $(1 + p)/2$  quantile of the Student’s *t*-distribution with  $N_e - 2$  degrees of freedom. If  $b$  does not contain zero, then the OLS trend is considered as statistically significant at the  $(1 - p)$  level. This modified method is referred to as OLS-M. Formula (3) indicates that if there is serial correlation—namely,  $r_1$  is larger than zero—the OLS confidence intervals will be narrower than those of OLS-M. Therefore, an actually not-significant trend would potentially be mistaken as significant when using OLS.

The nonparametric Kendall’s tau-based Sen’s slope estimator (Sen 1968) is an alternative to OLS regression in estimating the linear trend. It is the median of the set of slopes  $\frac{Y_j - Y_i}{j - i}$ . It does not assume a distribution for the residuals and is much less sensitive to outliers in the time series. However, Sen’s (1968) slope estimator assumes the sample data to be serially independent. The nonparametric Mann–Kendall test (Mann 1945; Kendall 1955) for statistical significance testing of the linear trend also assumes the sample data to be serially independent. If there is a positive AR(1) in the time series, the test rejects the null hypothesis more often than specified by the significance level, and thus the testing result is unreliable (von Storch and Navarra 1995; Yue et al. 2002; Zhang and Zwiers 2004). Taking into account the fact that the trend and autocorrelation often concur in a time series, we adopt an iterative method, proposed by Zhang et al. (2000) and later refined by Wang and Swail (2001, Appendix A), to properly estimate the AR(1) of a time series and eliminate this effect of autocorrelation in using Sen’s slope estimator and the Mann–Kendall test. This method to compute the trend slopes and to test their statistical significance is

referred to as WS2001. In case there are ties (repeated values in the extreme index) in the sample data, the variance of the Mann–Kendall test statistic  $S$  is calculated by:

$$VAR(S) = \frac{n(n-1)(2n+5) - \sum_{j=1}^g u_j(u_j-1)(2u_j+5)}{18}, \tag{4}$$

where  $g$  is the number of tied groups and  $u_j$  is the number of repeated values in the  $j$ th group (Kendall 1955). In this study, the linear trend is regarded as statistically significant if it is significant at the 5% level.

### 3 Results

#### 3.1 Normality testing and serial correlation in terms of the OLS regression

##### 3.1.1 Station data

Theoretically, indices based on absolute values (TXx etc.) should not be Gaussian if the block size is large enough, and percentile-based indices may be approximated by a Gaussian distribution if the daily temperature data are sufficiently independent. However, a Gaussian approximation is not appropriate for many cases of percentile-based indices because of dependence in the daily data. This is more problematic for seasonal values because of smaller number sizes, for which there are a higher percentage of stations failing the Gaussian test. This is supported by Table 2 and Fig. 2.

More specifically, Fig. 2 shows that more than half of the stations are non-Gaussian for each of the eight annual extreme temperature indices. For example, most of the stations around the middle and lower reaches of the Yangtze River and the Huai River for TNn (Fig. 2d), many of the

stations in southwestern China for TX90p (Fig. 2e), and many of the stations in northeastern China for TN10p (Fig. 2h), are non-Gaussian. The number of non-Gaussian stations varies among the indices. The largest percentage of non-Gaussian stations accounting for the overall 758 stations is 74.4% for TNn, whereas the smallest one is 58.2% for TNx (Table 2).

Among the Gaussian distributed stations, many are serially dependent, with the  $r_1$  value larger than zero for each extreme temperature index (Fig. 2). These serial correlations will potentially introduce incorrect significance test results that suggest significant trends when actually they are not, if using the standard Student’s  $t$ -test with  $N-2$  degrees of freedom. The largest percentage of serially dependent stations accounting for the overall 758 stations is 31.4% for TN90p, whereas the smallest one is 11.1% for TNn (Table 2). For TX90p (Fig. 2e) and TN90p (Fig. 2f), there are 15.7% (6.1%) and 19.4% (10.8%), respectively, of the 758 stations whose  $r_1$  is larger than 0.2 (0.3), which indicates  $N_e$  is no more than 2/3 (half) of the data length. The maximum  $r_1$  values for the eight indices are 0.46, 0.46, 0.44, 0.33, 0.56, 0.63, 0.52 and 0.55, respectively, which indicates the  $N_e$  values at these stations are only 37.0%, 37.0%, 38.9%, 50.4%, 28.2%, 22.7%, 31.6% and 29.0%, respectively, of the data length.

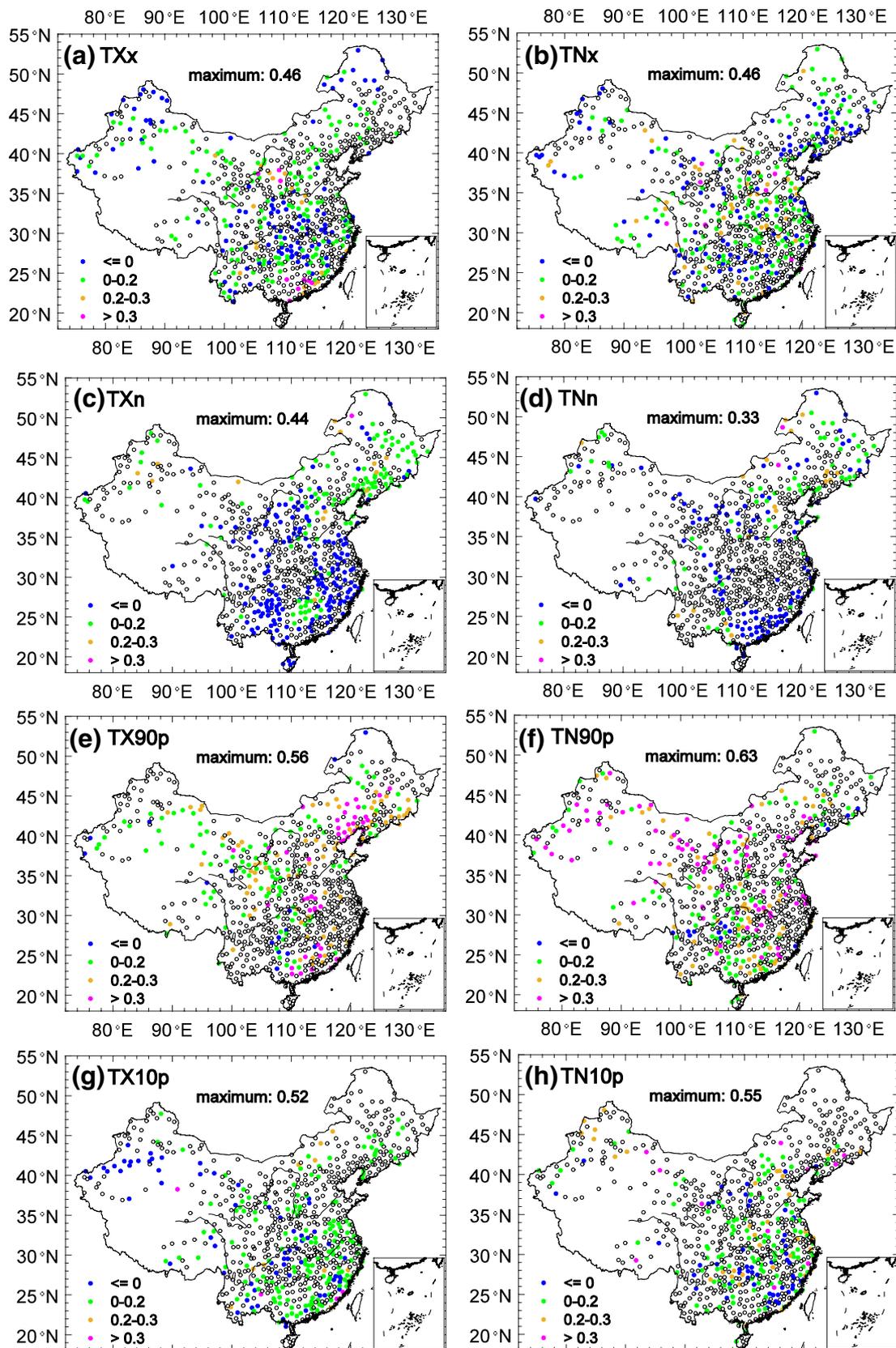
If the numbers of non-Gaussian stations and Gaussian but serially dependent stations are added up, more than 2/3 of the stations cannot use standard OLS regression to estimate their confidence intervals as well as the statistical significance of the linear trend in the eight indices, especially for TX90p and TN90p (Table 2). For these two indices, this is the case for more than 98% of the stations.

In terms of summer (June–July–August, JJA) indices (Table 2), the number of non-Gaussian stations increases substantially for the latter four percentile-based indices relative to the annual cases. The largest amount of non-Gaussian

**Table 2** Percentage of stations whose extreme temperature indices are non-Gaussian distributed or Gaussian distributed but serially dependent ( $AR1 > 0$ ) in terms of the OLS regression residual

	TXx	TNx	TXn	TNn	TX90p	TN90p	TX10p	TN10p
Annual								
Non-Gaussian	63.3	58.2	59.5	74.4	69.4	66.8	68.6	72.2
Dependent	22.3	26.7	16.4	11.1	28.6	31.4	22.7	20.8
Sum	85.6	84.9	75.9	85.5	98.0	98.2	91.3	93.0
Summer								
Non-Gaussian	61.2	57.5	65.3	74.0	94.6	85.4	93.4	92.4
Dependent	24.1	26.3	15.2	16.1	4.6	9.5	3.2	4.5
Sum	85.3	83.8	80.5	90.1	99.2	94.9	96.6	96.9
Winter								
Non-Gaussian	61.5	64.5	65.3	74.7	90.4	74.8	99.7	96.0
Dependent	16.1	14.4	6.1	5.2	3.2	13.7	0.0	1.9
Sum	77.6	78.9	71.4	79.9	93.6	88.5	99.7	97.9

The sum of the non-Gaussian and Gaussian but serially dependent categories represents the percentage of stations that cannot use OLS and the Student’s  $t$ -test with  $N-2$  degrees of freedom in significance testing (units: %)



**Fig. 2** Normality test results according to Q–Q plots and the AR(1) values of OLS regression residuals for eight annual extreme temperature indices at 758 stations during 1960–2017. Black circles indicate non-Gaussian stations; colored dots indicate Gaussian stations. The AR(1) values are calculated only for Gaussian stations. The AR(1) value of 0.2 (0.3) in the legend indicates that the  $N_e$  in formula (1) reduces by approximately 1/3 (50%) of the original sample size  $N$ . The maximum AR(1) values are also listed for each index. Only blue dots indicate that the regression residual at a station is serially independent and can use the original OLS regression and Student's  $t$ -test with  $N - 2$  degrees of freedom

stations is from TX90p. For this index, 94.6% of the stations are non-Gaussian. For Gaussian but serially dependent stations, there are 15–26% for the former four indices and 3–10% for the latter four indices. Altogether, approximately 80–99% of the stations cannot use standard OLS regression to estimate their confidence intervals as well as the statistical significance of the linear trend in the eight indices, with the smallest amount of stations for TXn and the largest for TX90p.

In terms of the winter (December–January–February, DJF) indices (Table 2), the number of non-Gaussian stations substantially increases for TX90p, TX10p and TN10p, relative to the corresponding annual indices. The largest amount of non-Gaussian stations is from TX10p, which amounts to 99.7%. For the former four indices, approximately 62–75% of the stations are non-Gaussian, and approximately 5–16% of the stations are Gaussian but serially dependent. For the latter four indices, about 75–100% of the stations are non-Gaussian, and about 0–14% of the stations are Gaussian but serially dependent. Overall, 71–100% of the stations cannot use standard OLS regression to estimate their confidence intervals as well as the statistical significance of the linear trend in the eight indices, with the smallest amount of stations for TXn and the largest for TX10p.

It should be mentioned here that TXx (TNx) can be in May or September and TXn (TNn) can be in the previous January or February (in terms of winter) for some of the stations. So, there are slight differences between the annual TXx (TNx) and summer TXx (TNx), and between the annual TXn (TNn) and winter TXn (TNn), shown in Table 2.

### 3.1.2 HadEX2 gridded data

For HadEX2, each of the eight annual indices has many non-Gaussian grid boxes within the China domain (Fig. 3; Table 3), although a grid box may be from the average of several stations and thus have a larger sample size than a single station to meet the central limit theorem. Non-Gaussian grid boxes account for approximately 21–67% of the entire 102 grid boxes within the China domain, with the least for TN90p and the most for TXn (Table 3). Non-Gaussian grid boxes exist mainly in western China for TXn (Fig. 3c);

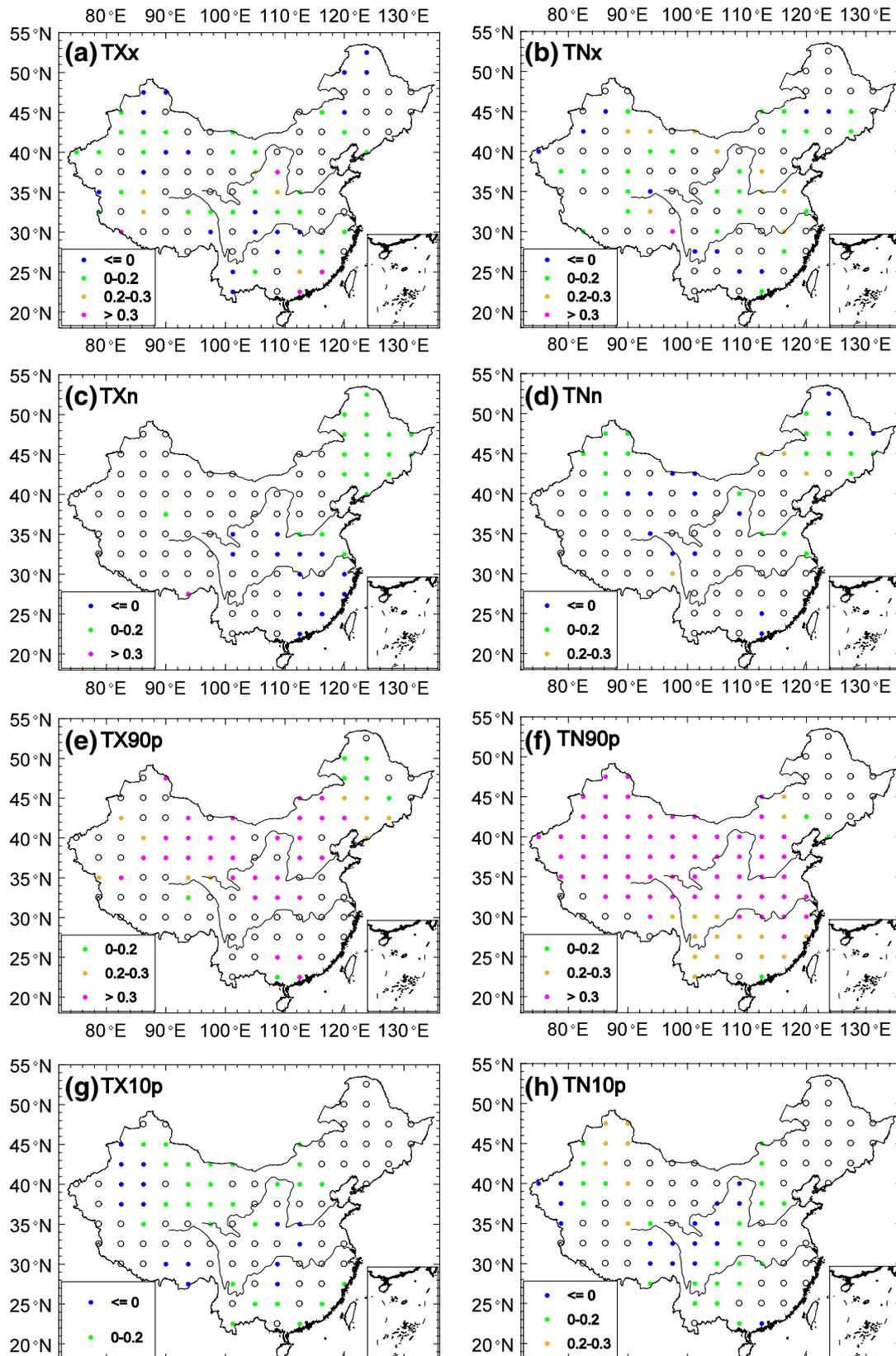
southern China for TNn (Fig. 3d) and TX90p (Fig. 3e); and northeastern China for TN90p (Fig. 3f), TX10p (Fig. 3g) and TN10p (Fig. 3h). For Gaussian grid boxes, serially dependent grid boxes account for 20–79%, with the least for TXn and the most for TN90p (Table 3). Most of the grid boxes in China have an  $N_e$  of no more than 2/3 of the data length for TN90p (Fig. 3f). Altogether, approximately 81–100% of the grid boxes in China cannot use standard OLS regression to estimate the confidence intervals as well as the statistical significance of the linear trend in the eight annual indices, with the smallest amount of stations for TXx and the largest for TX90p and TN90p (Table 3).

Table 3 also shows that, for summer indices, approximately 28–88% of the grid boxes in China are non-Gaussian, with the least for TNn and the most for TX10p. Approximately 0–51% of the grid boxes are Gaussian but serially dependent. Altogether, approximately 79–94% of the grid boxes cannot use standard OLS regression to carry out their significance testing. For winter indices, approximately 48–100% of the grid boxes are non-Gaussian, with the least for TN90p and the most for TX10p; and approximately 0–27% of the grid boxes are Gaussian but serially dependent. Altogether, approximately 72–100% of the grid boxes cannot use standard OLS regression to carry out their significance testing. In short, non-Gaussian and/or serial dependent characteristics should also be considered for gridded indices if one wants to use standard OLS to carry out the significance testing.

## 3.2 Serial correlation in terms of Sen's slope estimator and the Mann–Kendall test

### 3.2.1 Station data

Figure 4 shows that many of the stations are serially dependent for the eight annual indices, especially for TX90p (Fig. 4e) and TN90p (Fig. 4f). The maximum AR(1) values calculated from the WS2001 method are 0.50, 0.47, 0.43, 0.46, 0.58, 0.68, 0.53 and 0.71, for the eight indices. Table 4 shows that, for each of the eight annual indices, more than 43% of the stations have positive AR(1) values and thus cannot directly use the original Mann–Kendall test to test the statistical significance of the linear trend. Nor can they directly use Sen's slope estimator to calculate the linear trend slope. This is because  $Y_j$  and  $Y_i$ , which are input in all possible slopes ( $\frac{Y_j - Y_i}{j - i}$ ) to estimate the median value, are assumed in Sen's slope estimator to be independent (Sen 1968). The differences for whether or not to take into account serial correlation are illustrated later, in Sect. 3.3. Special attention should be paid to the annual TX90p and TN90p indices, in which more than 88% of the stations have positive serial correlation (Table 4). For summer indices, the



**Fig. 3** As in Fig. 2, but for the OLS regression residual of HadEX2 gridded data during 1961–2010. The results are not shown for some of the grid boxes with 1 year of missing values in **a** and **b**



**Table 3** As in Table 2, but for HadEX2 data for the period 1961–2010

	TXx	TNx	TXn	TNn	TX90p	TN90p	TX10p	TN10p
Annual								
Non-Gaussian	47.1	54.9	66.7	62.8	52.9	20.6	58.8	53.9
Dependent	34.3	32.4	19.6	22.6	47.1	79.4	26.5	29.4
Sum	81.4	87.3	86.3	85.4	100	100	85.3	83.3
Summer								
Non-Gaussian	43.1	58.8	69.6	28.4	83.3	73.5	88.2	75.5
Dependent	36.3	33.3	13.7	51.0	10.8	19.6	0	13.7
Sum	79.4	92.1	83.3	79.4	94.1	93.1	88.2	89.2
Winter								
Non-Gaussian	55.9	65.7	64.7	62.8	97.1	48.0	100	88.2
Dependent	26.5	15.7	10.8	8.8	2.0	35.3	0	3.9
Sum	82.4	81.4	75.5	71.6	99.1	83.3	100	92.1

There are 102 grid boxes within the China domain

percentage of stations having positive serial correlation falls within 43–59%, with the smallest for TXn and the largest for TXx. For winter indices, this range is 27–68%, with the smallest for TXn and the largest for TX10p. Therefore, serial correlation should be considered when using Sen’s slope estimator and the Mann–Kendall test to analyze the station-based trend in temperature extremes in China.

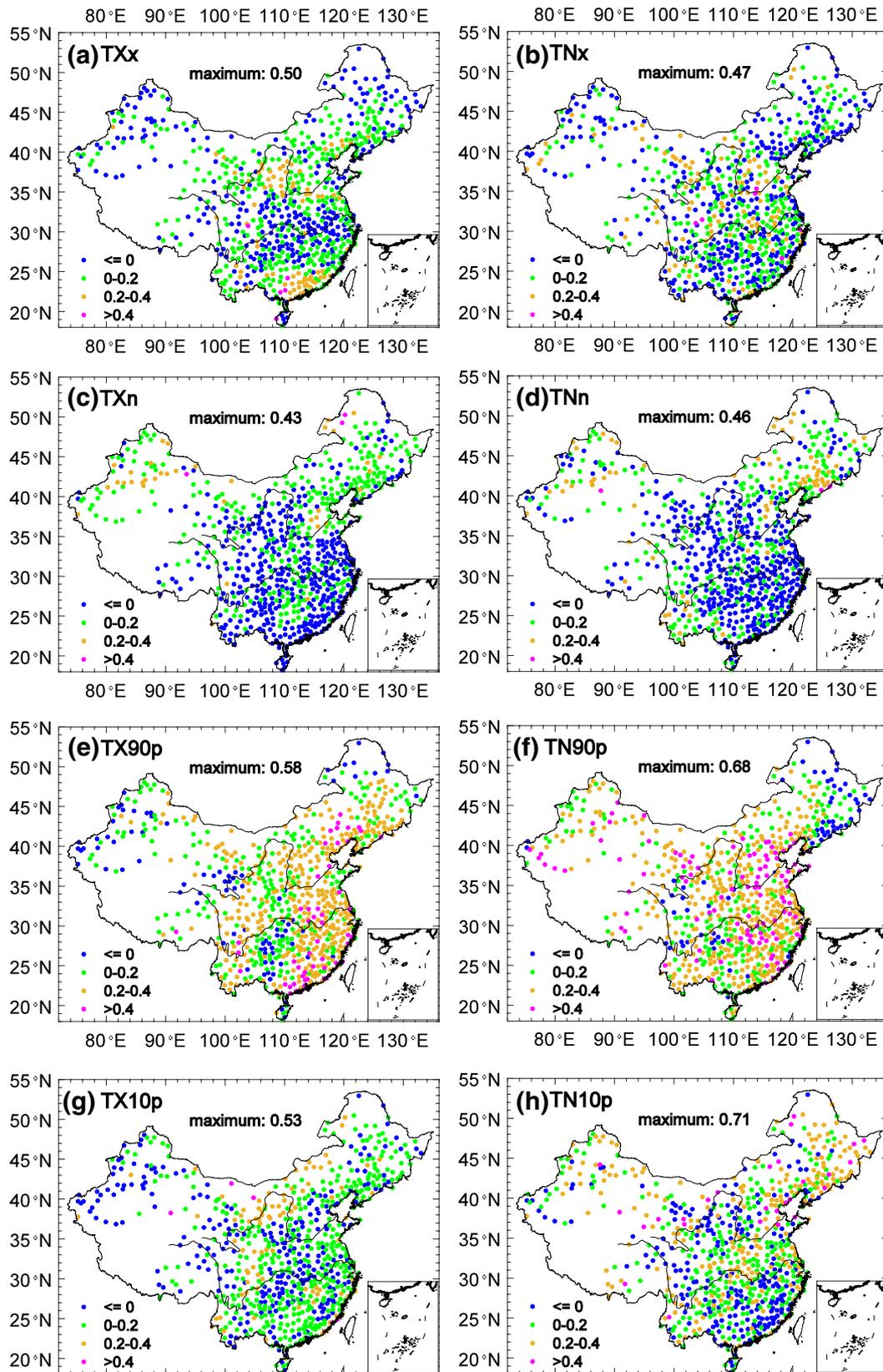
### 3.2.2 HadEX2 gridded data

For HadEX2, the numbers of serially independent grid boxes within the China domain are relatively small for each of the eight annual indices (Fig. 5). These grid boxes exist mostly in northeastern China and Xinjiang Autonomous Region for TXx (Fig. 5a), the upper reaches of the Yellow River and Yangtze River for both TXn (Fig. 5c) and TNn (Fig. 5d), Heilongjiang Province for TN90p (Fig. 5f), and the lower reaches of the Yellow River and Yangtze River for both TX10p (Fig. 5g) and TN10p (Fig. 5h). Approximately 54–96% of the grid boxes are serially dependent, with the least for TN10p and the most for TX90p (Table 5). For TX90p (Fig. 5g) and TN90p (Fig. 5f), most of the grid boxes within the China domain have an AR(1) larger than 0.2. For summer (winter) indices, approximately 41–66% (23–79%) of the grid boxes are serially dependent, with the least for TN10p (TX90p) and the most for TX90p (TN10p). Therefore, serial correlation should also be considered when using Sen’s slope estimator and the Mann–Kendall test to analyze the grid-based trend in temperature extremes in China.

### 3.3 Comparison of the spatial pattern of linear trends using different methods

The annual TX90p index is taken as an example to illustrate the impact of non-Gaussian and/or serial dependent characteristics on the estimation of the linear trend slope

and the corresponding statistical significance of the linear trend (Fig. 6). In order to see the results clearly, only part of China is shown. Figure 6a compares the two parametric methods and shows that the linear trend slope best-estimates are the same using the OLS and OLS-M method, but the statistical significances for the linear trends are not necessarily the same. For example, the trends at many Gaussian stations (Fig. 2e) in northeastern China are statistically significant using the OLS method, but not significant using the OLS-M method (Fig. 6a, with a typical example in its top-left corner), due to the presence of serial dependence at these stations (Fig. 2e). Figure 6b compares the two nonparametric methods and shows that both the linear trend slope best-estimates and the statistical significances for the linear trends obtained using the original Sen’s slope estimator and the Mann–Kendall test are not necessarily the same as those obtained using the WS2001 method, due to the presence of AR(1) at these stations (Fig. 4e). For example, most of the stations in the lower reaches of the Yellow River Basin have different trend slope magnitudes (Fig. 6b), and several stations there even have different trend signs; the trends at several stations in northeastern China are statistically significant using the original Mann–Kendall test, but are not significant using the WS2001 method (Fig. 6b). The reason for different slopes has been explained earlier, in Sect. 3.2.1. Figure 6c compares the refined parametric method with the refined nonparametric method and shows that the statistical significances are different from each other for some of the non-Gaussian stations—for example, those in northeastern China (Figs. 2e, 6d). Some stations show a statistically significant trend using WS2001 but not using OLS-M; whereas some stations are not significant using WS2001 but significant using OLS-M (Fig. 6c, d). In particular, all the trend slopes are different using OLS-M and WS2001 (Fig. 6c). It should be mentioned here that, even if many non-Gaussian stations show the same significance-test results between



**Fig. 4** AR(1) values according to the WS2001 method for eight annual extreme temperature indices at 758 stations during 1960–2017. The maximum AR(1) values are also listed for each index. Only stations colored blue can use the original Mann–Kendall test

**Table 4** Percentage of stations whose extreme temperature indices are serially dependent (WS2001  $AR1 > 0$ ) when using the Sen's slope estimator and Mann–Kendall test (units: %)

	TXx	TNx	TXn	TNn	TX90p	TN90p	TX10p	TN10p
Annual	59.8	53.2	44.6	43.1	88.5	90.6	65.3	69
Summer	58.8	53.3	42.9	44.5	55.8	51.2	43.8	48.3
Winter	46.6	35.4	27.2	32.1	30.2	42.2	67.9	54.1

OLS-M and WS2001 in this case (Fig. 6c), it is not always the case for every extreme index. Those from OLS-M just happened to be right for the wrong reason, because the pre-requirements of the methods were not met.

In summary, the differences described above suggest that the non-Gaussian and/or serial dependent characteristics should be considered when analyzing the trend of temperature extremes in China, especially in the assessment of the statistical significance of the linear trend. Thus, the spatial patterns of the linear trend in temperature extremes, as estimated using the WS2001 method, are reported in the following.

### 3.4 Spatial pattern of linear trends in temperature extremes

#### 3.4.1 Annual temperature extremes

Figure 7 shows that, for the majority of stations, the temperatures on the hottest day, warmest night, coldest day, and coldest night in a year have increased (Fig. 7a–d); the annual occurrences of warm days and warm nights have increased (Fig. 7e, f), whereas those of cold days and cold nights have decreased (Fig. 7g, h). These characteristics reflect an overall warming tendency. This tendency is seen more spatially coherent across China in the  $T_{\min}$ -related indices (Fig. 7b, d, f, h) than in the  $T_{\max}$ -related indices (Fig. 7a, c, e, g). All the stations have increasing trends, and 99% of them are statistically significant, for TN90p (Fig. 7f). Almost all (99%) the stations have significant decreasing trends, and no station has an increasing trend, for TN10p (Fig. 7h).

However, there are also regional differences. For TXx (Fig. 7a) and TXn (Fig. 7c), increasing trends and decreasing trends are scattered across China; fewer than one-third of stations have significant increasing trends (33% for TXx and 28% for TXn), mostly in the upper reaches of the Yellow River Basin and in the middle and lower reaches of the Yangtze River Basin, although hardly any station has a significant decreasing trend. For TNx (Fig. 7b) and TNn (Fig. 7d), more than two-thirds of the stations have significant increasing trends, particularly in semi-arid zones and East China, whereas a few stations in central China have slightly decreasing trends. For TX90p (Fig. 7e), the majority of the stations (77%) have significant increasing trends,

whereas a few stations in central-eastern China and southwestern China have slight decreasing trends. For TX10p (Fig. 7g), the majority of stations have decreasing trends, and 64% of all stations are statistically significant, most prominently in northern China and the Tibetan Plateau, but three stations in southern China have slight increasing trends.

#### 3.4.2 Hot and cold temperature extremes

Two hot extreme (summer high temperature) indices, i.e., JJA TX90p and JJA TN90p, and two cold extreme (winter low temperature) indices, i.e., DJF TX10p and DJF TN10p, are analyzed in the following (Fig. 8), because they are commonly related to human illness or even death. As reported in Sect. 3.1.1, these indices have a large amount of non-Gaussian stations. For the majority of stations, the occurrences of hot extremes have increased (Fig. 8a, b), whereas those of cold extremes have decreased (Fig. 8c, d). Like in the annual cases, the signs of the trends are more spatially coherent across China in the  $T_{\min}$ -related indices (Fig. 8b, d) than in the  $T_{\max}$ -related indices (Fig. 8a, c).

In more detail, for JJA TX90p (Fig. 8a), 57% of the stations have significant increasing trends, most prominently in western China and East China, whereas there are also a few stations in central-eastern China, parts of northeastern China, and the western end of Xinjiang Autonomous Region that have slight decreasing trends. A high proportion (92%) of the stations have significant increasing trends for JJA TN90p (Fig. 8b), whereas nine stations in central China have slight decreasing trends and two are statistically significant. For DJF TX10p (Fig. 8c), the majority of stations have decreasing trends, and 32% of the stations are statistically significant, mostly along the Yellow River and the Yangtze River; however, five stations in northeastern China and a few stations in southern China have a slight increasing trend. All except one station have decreasing trends, and 85% are statistically significant, for DJF TN10p (Fig. 8d).

It should be noted that the above results are based on the national Reference Climatic and Basic Meteorological Stations available for ordinary users. Due to rapid urban development in China, the trends of extreme temperature indices at these stations may have been affected to some extent by urbanization, as reported in previous studies

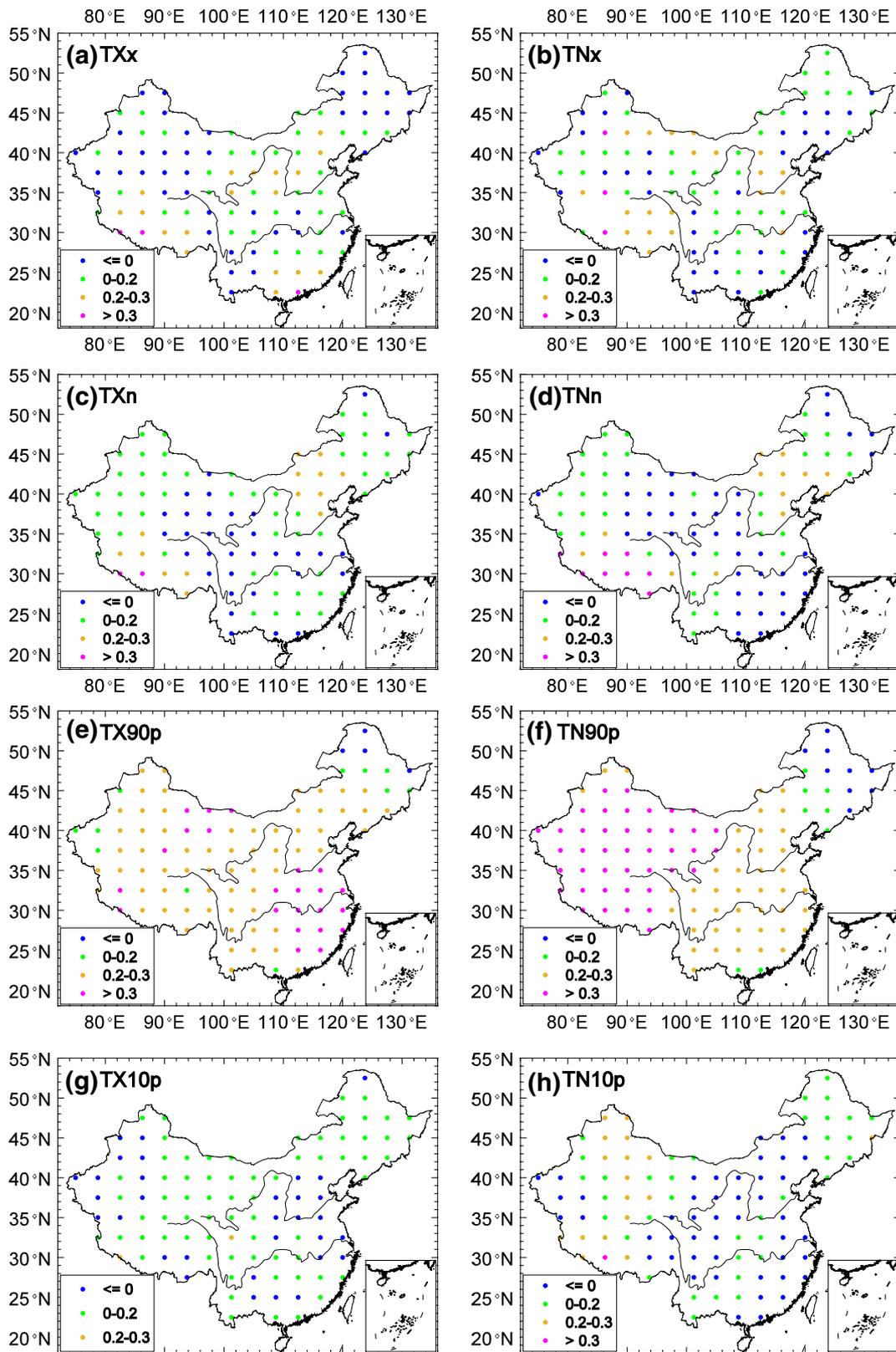
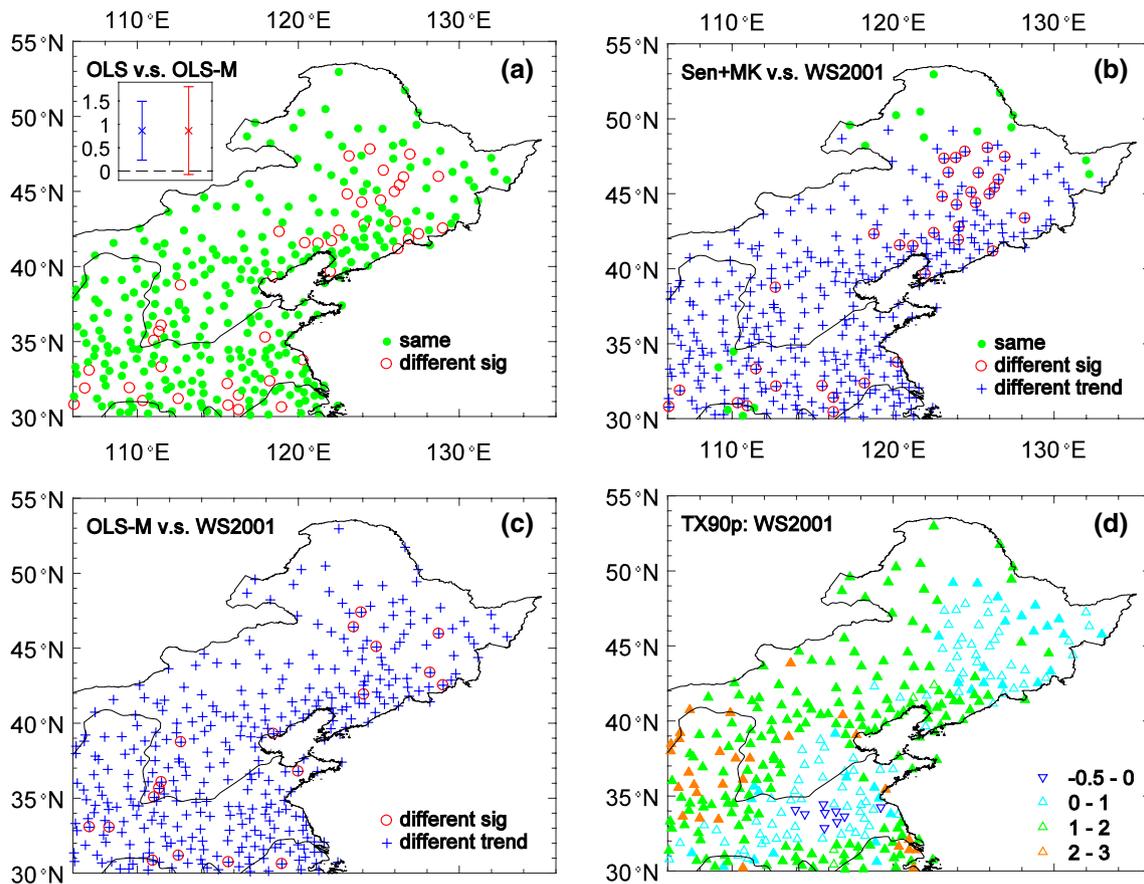


Fig. 5 As in Fig. 4, but for HadEX2 gridded data during 1961–2010

**Table 5** As in Table 4, but for HadEX2 data for the period 1961–2010 within the China domain

	TXx	TNx	TXn	TNn	TX90p	TN90p	TX10p	TN10p
Annual	56.9	64.7	67.7	55.9	96.1	91.2	68.6	53.9
Summer	59.8	68.6	35.3	46.1	65.7	61.8	48.0	41.2
Winter	56.9	42.2	30.4	31.4	22.6	38.2	67.7	79.4



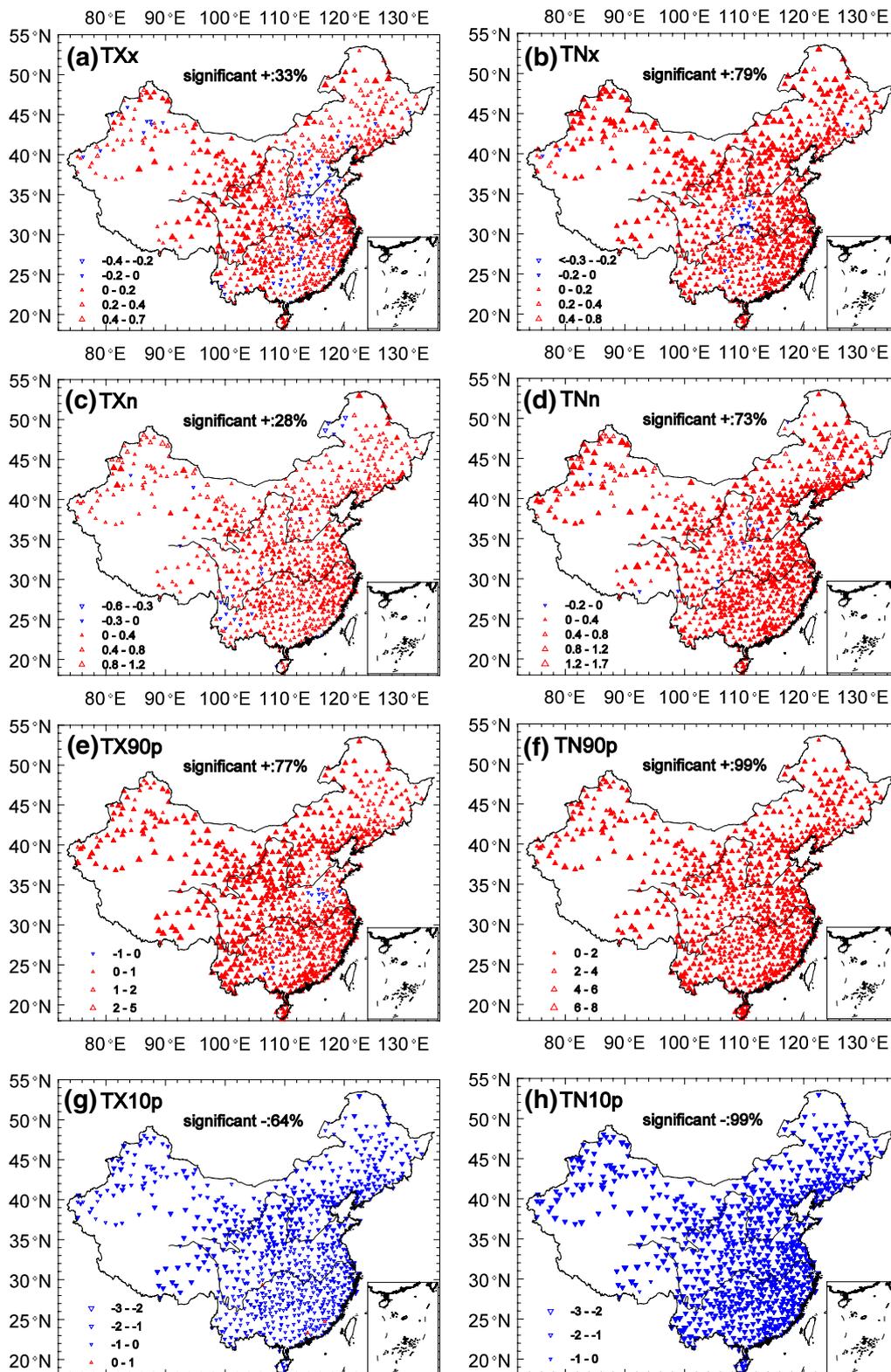
**Fig. 6** Comparison between the linear trends of the annual TX90p index for the period 1960–2017 using different methods: **a** between standard OLS regression and refined OLS regression, with consideration of AR(1) in the Student’s *t*-test; **b** between the original Sen’s slope estimator combined with the Mann–Kendall test and Sen’s slope estimator combined with the Mann–Kendall test, both considering AR(1); **c** between refined OLS regression and WS2001; **d** results based on WS2001, in which the units are %/decade and solid triangles indicate the linear trends at these stations are statistically significant

at the 5% level, whereas hollow triangles indicate the linear trends are not statistically significant. In **a–c**, green dots indicate stations with the same trend slope best-estimate and significance-test results (significant or not); red circles indicate stations with different significance-test results; blue crosses indicate stations with different trend slope best-estimates. In **a**, the confidence intervals of OLS (in blue) and that of OLS-M (in red) at Haerbin (50953) station are shown in the top-left corner to represent cases with different significances

(e.g. Zhou and Ren 2011; Ren et al. 2014; Qian 2016). It would be helpful to further analyze trends in extreme indices at individual stations based on homogenized data from 2419 stations (Cao et al. 2016), which include more rural stations. Nevertheless, the large-scale pattern of observed changes in temperature extremes is similar over Asia (Dong et al. 2018).

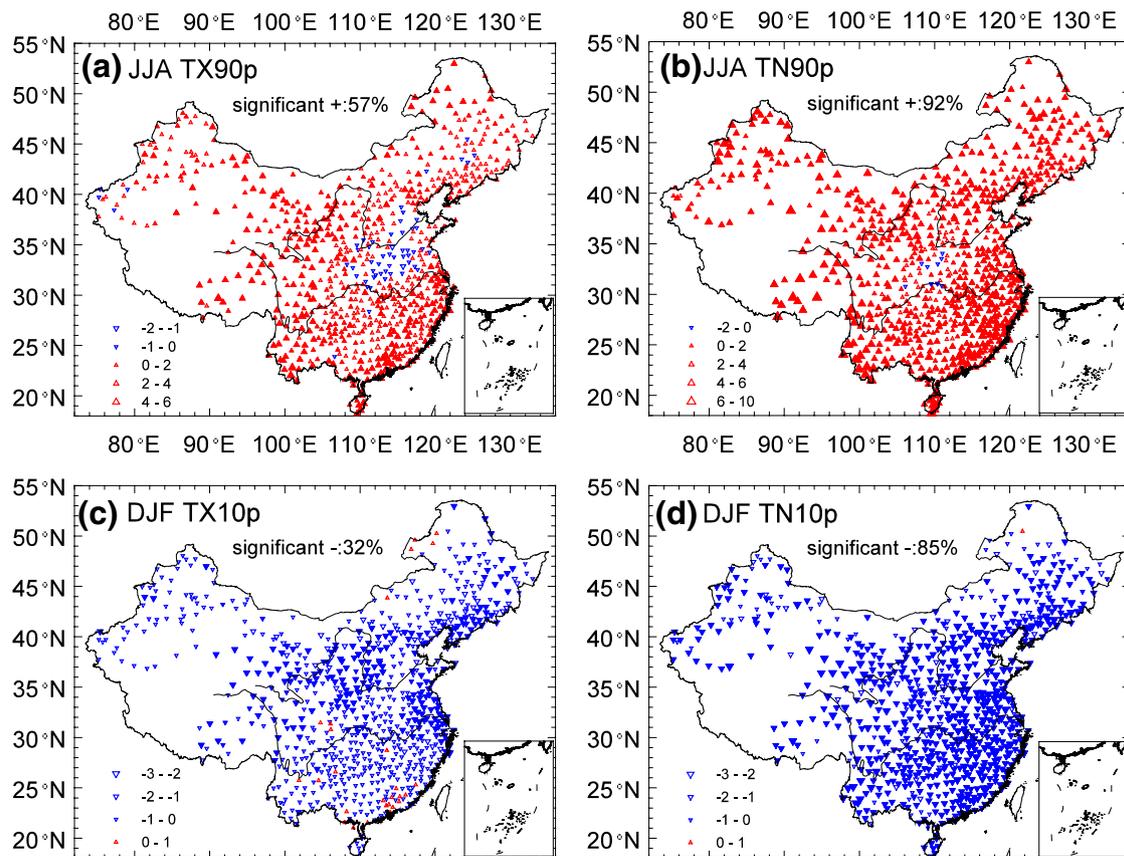
### 4 Conclusions and implications

In this paper, whether the linear trend residuals of eight commonly used extreme temperature indices at each station or each HadEX2 grid box across China are Gaussian and/or serial independent, is examined for the determination of appropriate linear trend analysis method for



**Fig. 7** Linear trends in eight annual extreme temperature indices at 758 stations for the period 1960–2017 as estimated by the WS2001 method (units: °C/decade for a–d and %/decade for e–h). Solid triangles indicate the linear trends at these stations are statistically sig-

nificant at the 5% level, whereas hollow triangles indicate the linear trends are not statistically significant. The percentages of stations whose trends are dominant in sign and statistically significant are also listed, for each index. The denominator is 758



**Fig. 8** As in Fig. 7, but for two hot extreme indices (a summer TX90p and b summer TN90p) and two cold extreme indices (c winter TX10p and d winter TN10p)

temperature extremes. The findings provide important insights for other researchers working on similar or related problems. The spatial patterns of the linear trend in annual temperature extremes, as well as those in hot extremes and cold extremes, are further analyzed, by taking into account the non-Gaussian and/or serially dependent characteristics, on the basis of updated homogenized station data for the period 1960–2017. The major findings can be summarized as follows:

1. Among the 758 stations analyzed, at least 57.5–99.7% are non-Gaussian and 71.4–99.7% cannot directly use standard OLS regression to analyze the confidence intervals and corresponding statistical significance of the linear trend in the eight annual/summer/winter extreme temperature indices, because of either non-Gaussian or Gaussian but serial dependent characteristics.
2. The proportion of stations unable to directly use the original Sen’s slope estimator and Man–Kendall test to analyze annual extreme temperature indices, because of serial dependence at these stations, ranges from 43 to

91%. For summer (winter) indices, this proportion is 43–59% (27–68%).

3. Non-Gaussian and/or serially dependent characteristics are also widespread in the HadEX2 gridded data. The percentages obtained from HadEX2 are similar to those obtained from the station data.
4. If using the original Sen’s slope estimator and Man–Kendall test, both the trend slope and statistical significance of temperature extremes will be potentially incorrect for those stations with serial dependence; plus, if using the refined OLS method that takes into account serial dependence, the statistical significance of temperature extremes will potentially be wrong for stations having non-Gaussian residuals.
5. For the majority of stations during 1960–2017, the temperatures on the hottest day, the warmest night, the coldest day, and the coldest night in a year have increased; the annual occurrences of warm days and warm nights have increased, whereas those of cold days and cold nights have decreased. Among them, 28–99% of the stations are statistically significant. The occurrences

of hot extremes have increased; whereas, those of cold extremes have decreased despite record-breaking cold extremes having occurred at some stations in recent years. The trends in the  $T_{\min}$ -related indices at the majority of stations are statistically significant, whereas those in the  $T_{\max}$ -related indices are much less spatially coherent.

The above results further highlight the importance of trend estimation and significance testing methods in the linear trend analysis of temperature extremes in China, as previously noted by Qian (2016). Many stations or grid boxes throughout China are found to be non-Gaussian and/or serially dependent. These characteristics should also be considered in the trend estimation of other climate extremes. For those indices whose trends are less prominent than temperature, the serial dependences will likely introduce larger differences in the significance testing results when considering these characteristics than when not. Some studies have discussed  $P$  values not being as reliable as many scientists assume (e.g. Nuzzo 2014), and have called for a stricter significance level in significance testing. For example, Benjamin et al. (2018) propose changing the default  $P$  value threshold for statistical significance from 0.05 to 0.005 by Bayes' rule for claims of new discoveries to improve reproducibility.

**Acknowledgements** This study was sponsored by the National Key Research and Development Program of China (Grant 2018YFC1507701), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant XDA20020201), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2016075), and the Jiangsu Collaborative Innovation Center for Climate Change.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alexander LV, Zhang X, Peterson TC, Caesa J, Gleason B, Klein Tank AMG, Haylock M, Collins D, Trewin B, Rahimzadeh F, Tagipour A, Rupa Kumar K, Revadekar J, Griffiths G, Vincent L, Stephenson DB, Burn J, Aguilar E, Brunet M, Taylor M, New M, Zhai P, Rusticucci M, Vazquez-Aguirre JL (2006) Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res Atmos* 111:D05109. <https://doi.org/10.1029/2005JD006290>
- Barnes EA (2013) Revisiting the evidence linking Arctic amplification to extreme weather in midlatitudes. *Geophys Res Lett* 40:4734–4739. <https://doi.org/10.1002/grl.50880>
- Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges LV, Held L, Ho TH, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchner M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouders J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J, Johnson VE (2018) Redefine statistical significance. *Nat Human Behav* 2(1):6–10
- Cao LJ, Zhu YN, Tang GL, Yuan F, Yan Z (2016) Climatic warming in China according to a homogenized data set from 2419 stations. *Int J Climatol* 36:4384–4392. <https://doi.org/10.1002/joc.4639>
- Chen Y, Zhai P (2017) Revisiting summertime hot extremes in China during 1961–2015: overlooked compound extremes and significant changes. *Geophys Res Lett* 44:5096–5103. <https://doi.org/10.1002/2016GL072281>
- Chen A, He X, Guan H, Cai Y (2018) Trends and periodicity of daily temperature and precipitation extremes during 1960–2013 in Hunan Province, central south China. *Theor Appl Climatol* 132:71–88. <https://doi.org/10.1007/s00704-017-2069-x>
- Cubasch U, Wuebbles D, Chen D, Facchini MC, Frame D, Mahowald N, Wintner J-G (2013) Introduction. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge
- Ding T, Ke Z (2015) Characteristics and changes of regional wet and dry heat wave events in China during 1960–2013. *Theor Appl Climatol* 122:651–665. <https://doi.org/10.1007/s00704-014-1322-9>
- Ding T, Qian WH, Yan ZW (2010) Changes in hot days and heat waves in China during 1961–2007. *Int J Climatol* 30:1452–1462. <https://doi.org/10.1002/joc.1989>
- Donat MG et al (2013) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J Geophys Res Atmos*. <https://doi.org/10.1002/jgrd.50150>
- Dong S, Sun Y, Aguilar E, Zhang X, Peterson TC, Song L, Zhang Y (2018) Observed changes in temperature extremes over Asia and their attribution. *Clim Dyn* 51:339–353. <https://doi.org/10.1007/s00382-017-3927-z>
- Du Y-D, Ai H, Duan H-L et al (2013) Changes in climate factors and extreme climate events in South China during 1961–2010. *Adv Clim Change Res* 4(1):1–11. <https://doi.org/10.3724/SPJ.1248.2013.001>
- Francis JA (2017) Why are Arctic linkages to extreme weather still up in the air. *Bull Am Meteorol Soc* 98(12):2551–2557. <https://doi.org/10.1175/BAMS-D-17-0006.1>
- Hartmann DL, Klein Tank AMG, Rusticucci M, Alexander L, Brönnimann S, Charabi Y, Dentener F, Dlugokencky E, Easterling D, Kaplan A, Soden B, Thorne P, Wild M, Zhai PM (2013) Observations: atmosphere and surface supplementary material. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*
- Hogg RV (1979) An introduction to robust estimation. In: Launer RL, Wilkinson GN (eds) *Robustness in statistics*. Academic Press, New York, pp 1–17



- Huang DQ, Qian YF, Zhu J (2010) Trends of temperature extremes in China and their relationship with global temperature anomalies. *Adv Atmos Sci* 27(4):937–946. <https://doi.org/10.1007/s00376-009-9085-4>
- Huang L, Chen A-F, Zhu Y-H, Wang H-L, He B (2015) Trends of temperature extremes in summer and winter during 1971–2013 in China. *Atmos Ocean Sci Lett* 8(4):220–225
- IPCC (2013) Climate change 2013: the physical science basis. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, p 1535
- Jiang DJ, Li Z, Wang QX (2012) Trends in temperature and precipitation extremes over Circum-Bohai-Sea region, China. *Chin Geogra Sci* 22(1):75–87. <https://doi.org/10.1007/s11769-012-0515-3>
- Kendall MG (1955) Rank correlation methods, 2nd edn. Charles Griffin, London, p 196
- Kug J-S, Jeong J-H, Jang Y-S, Kim B-M, Folland CK, Min S-K, Son S-W (2015) Two distinct influences of Arctic warming on cold winters over North America and East Asia. *Nat Geosci* 8:759–762. <https://doi.org/10.1038/ngeo2517>
- Li Z, Cao L, Zhu Y et al (2016) Comparison of two homogenized datasets of daily maximum/mean/minimum temperature in China during 1960–2013. *J Meteorol Res* 30(1):53–66. <https://doi.org/10.1007/s13351-016-5054-x>
- Lin P, He Z, Du J, Chen L, Zhu X, Li J (2017) Recent changes in daily climate extremes in an arid mountain region, a case study in northwestern China's Qilian Mountains. *Sci Rep* 7:2245. <https://doi.org/10.1038/s41598-017-02345-4>
- Liu C, Deng H, Lu Y, Qiu X, Wang D (2018) Changes in 'hotter and wetter' events across China. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-017-2344-x>
- Mann HB (1945) Nonparametric tests against trend. *Econometrica* 13:245–259
- Mori M, Watanabe M, Shiogama H, Inoue J, Kimoto M (2014) Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nat Geosci* 7:869–873. <https://doi.org/10.1038/NGEO2277>
- Nuzzo R (2014) Scientific method: statistical errors. *Nature* 506(7487):150–152
- Qian C (2016) On trend estimation and significance testing for non-Gaussian and serially dependent data: quantifying the urbanization effect on trends in hot extremes in the megacity of Shanghai. *Clim Dyn* 47:329–344
- Qian W, Lin X (2004) Regional trends in recent temperature indices in China. *Clim Res* 27:119–134
- Qian C, Wu Z, Fu C, Wang D (2011a) On changing El Niño: a view from time-varying annual cycle, interannual variability and mean state. *J Clim* 24:6486–6500
- Qian C, Yan ZW, Wu Z, Fu CB, Tu K (2011b) Trends in temperature extremes in association with weather-intraseasonal fluctuations in eastern China. *Adv Atmos Sci* 28(2):297–309
- Qian C, Wang J, Dong S, Yin H, Burke C, Ciavarella A, Dong B, Freychet N, Lott FC, Tett SFB (2018) Human influence on the record-breaking cold event in January of 2016 in Eastern China [in "Explaining Extreme Events of 2016 from a Climate Perspective"]. *Bull Am Meteorol Soc* 99(1):S118–S122. <https://doi.org/10.1175/BAMS-D-17-0095.1>
- Ren G, Zhou Y (2014) Urbanization effect on trends of extreme temperature indices of national stations over Mainland China, 1961–2008. *J Clim* 27:2340–2360
- Santer BD et al (2008) Consistency of modelled and observed temperature trends in the tropical troposphere. *Int J Climatol* 28:1703–1722. <https://doi.org/10.1002/joc.1756>
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's Tau. *J Am Stat Assoc* 63:1379–1389
- Shi J, Cui L, Wen K et al (2018) Trends in the consecutive days of temperature and precipitation extremes in China during 1961–2015. *Environ Res* 161:381–391
- Sun Y, Zhang X, Zwiers FW, Song L, Wan H, Hu T, Yin H, Ren G (2014) Rapid increase in the risk of extreme summer heat in Eastern China. *Nat Clim Change* 4:1082–1085. <https://doi.org/10.1038/NCLIMATE2410>
- Szentimrey T (1999) Multiple analysis of series for homogenization (MASH). *Proc Sec Semin Homogen Surf Climatol Data Budapest* 41:27–46
- von Storch H, Navarra A (1995) Analysis of climate variability: applications of statistical techniques. Springer, Berlin, p 334
- von Storch H, Zwiers FW (1999) Statistical analysis in climate research. Cambridge University Press, Cambridge, p 484
- Wang XL, Swail VR (2001) Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. *J Clim* 14:2204–2222
- Wang ZY, Ding YH, Zhang Q, Song YF (2012) Changing trends of daily temperature extremes with different intensities in China. *Acta Meteorol Sin* 26(4):399–409. <https://doi.org/10.1007/s13351-012-0401-z>
- Wang J, Tett SFB, Yan Z, Feng J (2018) Have human activities changed the frequencies of absolute extreme temperatures in eastern China? *Environ Res Lett* 13:014012
- Wilks DS (2011) Statistical methods in the atmospheric sciences, 3rd edn. Academic Press, New York, p 676
- Xu W, Li Q, Wang XL, Yang S, Cao L, Feng Y (2013) Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J Geophys Res Atmos*. <https://doi.org/10.1002/jgrd.50791>
- Yan ZW, Li Z, Xia JJ (2014) Homogenization of climate series: the basis for assessing climate changes. *Sci China Earth Sci* 57:2891–2900
- Ye DX, Yin JF, Chen ZH et al (2014) Spatial and temporal variations of heat waves in China from 1961 to 2010. *Adv Clim Change Res* 5(2):66–73. <https://doi.org/10.3724/SPJ.1248.2014.066>
- Yin H, Donat MG, Alexander LV, Sun Y (2015) Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *Int J Climatol* 35:2809–2827
- You QL, Ren GY, Fraedrich K, Kang SC, Ren YY, Wang PL (2013) Winter temperature extremes in China and their possible causes. *Int J Climatol* 33:1444–1455
- Yue S, Pilon P, Phinney B, Cavadias G (2002) The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol Process* 16(9):1807–1829
- Zhai P, Pan X (2003) Trends in temperature extremes during 1951–1999 in China. *Geophys Res Lett* 30(17):1913. <https://doi.org/10.1029/2003GL018004>
- Zhang XB, Yang F (2004) RCLimDex (1.0) user manual. Climate Research Branch Environment, Canada
- Zhang X, Zwiers FW (2004) Comment on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test" by Sheng Yue and Chun Yuan Wang. *Water Resour Res* 40:W03805. <https://doi.org/10.1029/2003WR002073>
- Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmos Ocean* 38:395–429
- Zhang X, Hegerl G, Zwiers F, Kenyon J (2005) Avoiding inhomogeneity in percentile-based indices of temperature extremes. *J Clim* 18:1641–1651
- Zhang X, Alexander L, Hegerl GC, Jones P, Klein Tank A, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim Change* 2:851–870. <https://doi.org/10.1002/wcc.147>
- Zhao C-Y, Wang Y, Zhou X-Y et al (2013) Changes in climatic factors and extreme climate events in Northeast China during 1961–2010. *Adv Clim Change Res* 4(2):92–102. <https://doi.org/10.3724/SPJ.1248.2013.092>

- Zhou YQ, Ren GY (2011) Change in extreme temperature event frequency over mainland China, 1961–2008. *Clim Res* 50:125–139
- Zhou T, Ma S, Zou L (2014) Understanding a hot summer in central eastern China: summer 2013 in context of multi-model trend analysis [in “Explaining Extreme Events of 2013 from a Climate Perspective”]. *Bull Am Meteorol Soc* 95(9):S54–S57
- Zhou B, Xu Y, Wu J, Dong S, Shi Y (2016) Changes in temperature and precipitation extreme indices over China: analysis of a

high-resolution grid dataset. *Int J Climatol* 36:1051–1066. <https://doi.org/10.1002/joc.4400>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.