ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 26, NO. 1, 2009, 143–153

Detection and Adjustment of Undocumented Discontinuities in Chinese Temperature Series Using a Composite Approach

LI Qingxiang*[1,2] (李庆祥) and DONG Wenjie[1,3] (董文杰)

[1] *Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100049

[2] *National Meteorological Information Center, China Meteorological Administration, Beijing* 100081

[3] *State Key Laboratory of Earth Surface Processes and Ecology Resource, Beijing Normal University, Beijing* 100875

(Received 10 December 2007; revised 15 May 2008)

ABSTRACT

Annually averaged daily maximum and minimum surface temperatures from southeastern China were evaluated for artificial discontinuities using three different tests for undocumented changepoints. Changepoints in the time series were identified by comparing each target series to a reference calculated from values observed at a number of nearby stations. Under the assumption that no trend was present in the sequence of target-reference temperature differences, a changepoint was assigned to the target series when at least two of the three tests rejected the null hypothesis of no changepoint at approximately the same position in the difference series. Each target series then was adjusted using a procedure that accounts for discontinuities in average temperature values from nearby stations that otherwise could bias estimates of the magnitude of the target series step change. A spatial comparison of linear temperature trends in the adjusted annual temperature series suggests that major relative discontinuities were removed in the homogenization process. A greater number of relative change points were detected in annual average minimum than in average maximum temperature series. Some evidence is presented which suggests that minimum surface temperature fields may be more sensitive to changes in measurement practice than maximum temperature fields. In addition, given previous evidence of urban heat island (i.e., local) trends in this region, the assumption of no slope in a target-reference difference series is likely to be violated more frequently in minimum than in maximum temperature series. Consequently, there may be greater potential to confound trend and step changes in minimum temperature series.

Key words: temperature series, China, urban heat island, changepoints

Citation: Li, Q. X., and W. J. Dong, 2009: Detection and adjustment of undocumented discontinuities in Chinese temperature series using a composite approach. *Adv. Atmos. Sci.*, **26**(1), 143–153, doi: 10.1007/s00376-009-0143-8.

1. Introduction

Homogeneous climate series are essential to the study of climate variability and change, including assessments of possible anthropogenic impacts. Extensive and rapid urbanization in parts of Southeastern China, for example, has been implicated in the rate of regional temperature trends (Li et al., 2004b; Zhou et al., 2004). Unfortunately, China's historical temperature archives, like many long-term climate series, contain an unknown number of artificial change points that are caused by meteorological station moves or other changes to observation practice (Liu and Li, 2003; Li et al., 2004a; Yan et al., 2001; Zhai and Eskridge, 1996). Consequently, tests for artificial and undocumented change points are necessary to reduce uncertainty in the quantification and possible attribution of observed climate change and variability.

Many approaches to the homogenization of climatic time series have been documented in the literature (Potter, 1981; Alexandersson, 1986; Jones et al., 1986; Karl and Williams, 1987; Solow, 1987; Easterling and Peterson, 1995; Vincent, 1998). Some methods require the use of supporting metadata (i.e., station histories) to identify artificial change points while other methods can be used to detect undocumented

*Corresponding author: LI Qingxiang, liqx@cma.gov.cn

changes. Karl and Williams (1987), for example, used a t-test to identify discontinuities introduced into the U. S. temperature record at the known instants of observation practice changes. Values from surrounding stations were used to build a composite reference series that minimizes the confidence limits of the estimated changepoint magnitude. Jones et al. (1986) examined the differences between pairs of temperature series from nearby stations for abrupt jumps in global surface temperature series and used the metadata record as guidance wherever possible. Easterling and Peterson (1995) used a two-phase regression model to identify changepoints in difference series between a target and its composite reference series (Peterson and Easterling, 1994) without the use of metadata records.

Undocumented changepoint detection methods are particularly relevant to climate studies in China where the metadata archives are far from complete. Li et al. (2004a), therefore, applied the Easterling and Peterson (1995) two-phase regression approach to detect and remove relative changepoints in a collection of Chinese temperature records. The metadata archive was used to verify identified discontinuities wherever possible. However, subjective decisions regarding which undocumented changepoints to adjust were required because some apparent changepoints were errors caused by the use of an incorrect set of F-statistic percentiles (see Lund and Reeves, 2002, for details). Additional errors may have been artifacts of changepoints in the composite reference series that can be erroneously attributed to the target series (Menne and Williams, 2005).

There are two types of error that may occur in artificial changepoint hypothesis testing. Type I errors occur when the null hypothesis of homogeneity is rejected when it is true. Type II errors occur when the null hypothesis of homogeneity is not rejected when it is false. Although the impact of making Type I or Type II errors may be small in the calculation of large area averages (provided that artificial changes are random), changepoint testing errors will have a bigger impact on regional or local studies. Thus, the consequences of incomplete metadata (Karl and Williams, 1987), inappropriate test statistic percentiles in undocumented changepoint analysis (Lund and Reeves, 2002), or inhomogeneities in composite reference series (Menne and Williams, 2005) are of greater concern on smaller spatial scales of interest.

The goal of this analysis was to evaluate an alternative procedure for homogenizing temperature series from Southeastern China. In addition to using test statistic percentiles appropriate for undocumented changepoint detection, this procedure applies multiple hypothesis tests on each series of target-reference differences in an attempt to improve overall detection skill. A list of probable discontinuities was generated for all temperature series before an estimate of step-change size (magnitude) was made for any particular series. The magnitude of step change then was estimated using a procedure developed to avoid incorporating artificial shifts from surrounding stations into the sample of observations. The procedure was used to calculate the relative magnitude of the target changepoint (Karl and Williams, 1987). In addition, because of the likelihood of urban warming during recent decades in parts of Southeastern China (Li et al., 2004b), the possibility of confounding local trends with step changes is discussed. The changepoint detection method is described in section 2. The temperature dataset from Southeastern China is described in section 3. Results are discussed in section 4 and some discussion is offered in section 5. Concluding remarks are provided in section 6.

2. Method

Although numerous changepoint tests have been used to detect artificial discontinuities in climate series (see Peterson, 1998 for a review), not all are appropriate for the identification of undocumented changepoints. Likewise, a number of approaches to building a reference series have been discussed (e.g., Peterson and Easterling, 1994; Alexandersson and Moberg, 1997; Vincent, 1998). When a regional collection of climate series contains an unknown number of artificial changepoints, or when the data record is incomplete, the usual assumption of composite reference series homogeneity may be violated (e.g., Szentimrey, 1999; Caussinus and Mestre, 2004). In that case, there is a risk of falsely attributing changepoints in the reference series to the target series when the two are evaluated for relative breakpoints.

Menne and Williams (2005) evaluated the changepoint detection skill of three test statistics and three formulations of reference series in groups of simulated series with one to multiple step changes. Their goal was to quantify the impact that step changes in the reference component series may have on undocumented changepoint detection skill, particularly when a (regional) composite reference series is used in automated detection procedures. The degree of agreement between changepoint tests also was evaluated. The composite reference series formulations that were evaluated are described by Peterson and Easterling (1994), Alexandersson and Moberg (1997) and Vincent (1998). The three calculated test methods included a two-phase regression F-test (Lund and Reeves, 2002; referred to here as TPR), the Standard Normal Homogeneity or Likelihood Ratio test (e.g., Alexandersson, 1986; referred to here as SNHT) and a multiple linear regression-based test for serial (lag 1) correlation in the residuals (Vincent, 1998; referred to here as Lag 1). The form of SNHT that was used in the evaluation is essentially a zero-slope version of the TPR model. In their evaluation of simulated time series, the number of Type I errors (false alarms) was reduced by the multiple hypothesis test (Menne and Williams, 2005). In fact, a requirement of agreement between any two of the three changepoint tests reduced the number of false alarms more than test sensitivity under a wide variety of Monte Carlo simulated series with step changes in the target and/or reference series components. When step changes were present in the reference series components, many changepoints in the target-reference difference series were caused by changepoints in the component series that were used to calculate the composite reference. Requiring a consensus of hypothesis tests particularly improved overall detection skill in such circumstances.

Here, we also calculated the three separate test statistics and required the null hypothesis of homogeneity to be rejected by at least two of three tests to reject the null hypothesis of target minus reference series homogeneity. A 0.05 significance level was used for each test. Since there is some independence between the tests, even between the related Standard Normal Homogeneity Test (SNHT) and Two Phase Regression (TPR) tests, requiring an agreement between tests alters the significance level and therefore the Type I error rate. Successive hypothesis testing, which is commonly used when multiple undocumented changepoints may be present, also alters the Type I error rate (Menne and Williams, 2005). Essentially, forcing an agreement between tests reduces the Type I error rate, while successive hypothesis testing increases the Type I rate since sub-sequences of the series are evaluated recursively. Thus a semi hierarchic splitting algorithm used by Menne and Williams (2005) was applied to compare hypothesis testing to optimal solutions in this paper. As reported by Menne and Williams (2005), the use of this splitting and merging process (Hawkins, 1976) in successive hypothesis testing combined with the requirement of agreement between tests reduced the Type I error rate to less than 0.05 for simulated series in which the null hypothesis was true. The definition of agreement used here means that the probable positions of detected discontinuities coincided within a span of no more than ± 2 time steps (in this case, year) between the various tests. All difference series formed between the target and its reference were evaluated using the three hypothesis tests. When the null hypothesis was rejected by 2 of the 3 tests, a changepoint was assigned to the target series.

In many climate data homogenization studies, an emphasis is placed on detection of artificial change points while their adjustment is viewed as a relatively straightforward task. When artificial change points occur in most series, however, estimating the magnitude of a relative change point in any one series may be complicated by changes that occur at nearby sites, especially when they are close to the time of a target series changepoint whose magnitude is in question. The procedure described by Karl and Williams (1987) is one way to help prevent non-climatic changes that occur at sites near the target from biasing the estimates of the target changepoint magnitude. The original method required the use of metadata archives (station histories), but is used here by substituting a list of dates of probable undocumented change points for the metadata records. The main steps are as follows:

(1) Compile a list of the target (candidate, referred as CAN) station's neighbors. Within this list, there should be, in general, at least several stations that experienced no undocumented changes during the several years before and after a potential discontinuity in the target series. Given the density of stations in Southeastern China, the 20 nearest stations (the neighbors, referred as NEIGH) were identified for each CAN station.

(2) For the periods of simultaneous operation, calculate the correlation between the CAN and each NEIGH series. Correlation is used to rank the similarity of stations.

(3) Using the list of detected change points at the target site, obtained from the undocumented changepoint detection procedure (consensus by two of three tests) described above, proceed backward in time and identify the first potential discontinuity since the most discontinuity or the most recent year of data (2001 in this paper).

(4) Form a difference series between the target (CAN) and each NEIGH series. The usable interval for the difference series will be limited by other discontinuities in either the CAN or the NEIGH series. Consequently, this number will differ from one NEIGH to the next. A minimum of five values before and after the target breakpoint is required for a NEIGH series to be used.

(5) Calculate the confidence interval of a two sample student's t-test using a preset significance level (in this case 0.05) in which the sequences of differences (CAN-NEIGH), d, "before" the discontinuity (d_{b}) are compared to those "after" the potential discontinuity (d_{a}). This adjustment procedure assumes that the SNHT step-change model is appropriate for each breakpoint (see discussion below). It is important to note that the t-test is not used to identify the position

of the changepoint since the changepoint was identified using undocumented changepoint tests. Rather, it is used at this stage only to estimate confidence limits for the step-change magnitude.

(6) The confidence intervals are ranked from narrowest to widest for each usable NEIGH series. A correlation weighted composite reference series then is generated from the two different series, incorporating the narrowest confidence limits and calculating a confidence interval for the target minus the composite reference series. Additional "*d*" series are added to the composite reference series recursively until a minimum in the composite confidence interval is reached.

(7) If the minimum confidence interval (CI) does not include zero, then the average difference before and after the changepoint ($d_{\text{b, CAN-Neigh}} - d_{\text{a, CAN-Neigh}}$) is the estimate of the step-change magnitude and is used to adjust the target series (see Karl and Williams, 1987, for details).

A t-test at the last step of the procedure will never reject the statistical significance of a change point, found using an undocumented changepoint test, unless the discontinuity was falsely attributed to the target series because of non-homogeneities in the composite reference series. Since, a minimum of 10 values, that is, at least five before and five after the detected discontinuity, are required to estimate the confidence limits, apparent change points that occur near the ends of the series, or in close succession, may not meet the minimum number of observations criterion for magnitude estimates and thus will remain unadjusted.

3. Data

Annual average daily maximum and minimum temperature series were calculated using observations from stations located throughout Southeastern China, essentially in the region east of 105°E and south of 35°N. Daily maximum and minimum temperatures were obtained using standard observation rules for China (Chinese Meteorological Administration, 1979). Temperature values were quality controlled by procedures at the meteorological data center in the China National Meteorological Information Center. Only those series for which serially complete annual values (i.e., none missing) could be calculated during the period from 1960 to 2001 were evaluated. The result is a group of maximum and minimum temperature series with a reasonably dense and uniform distribution in Southeastern China, as shown in Fig. 1. Consequ-

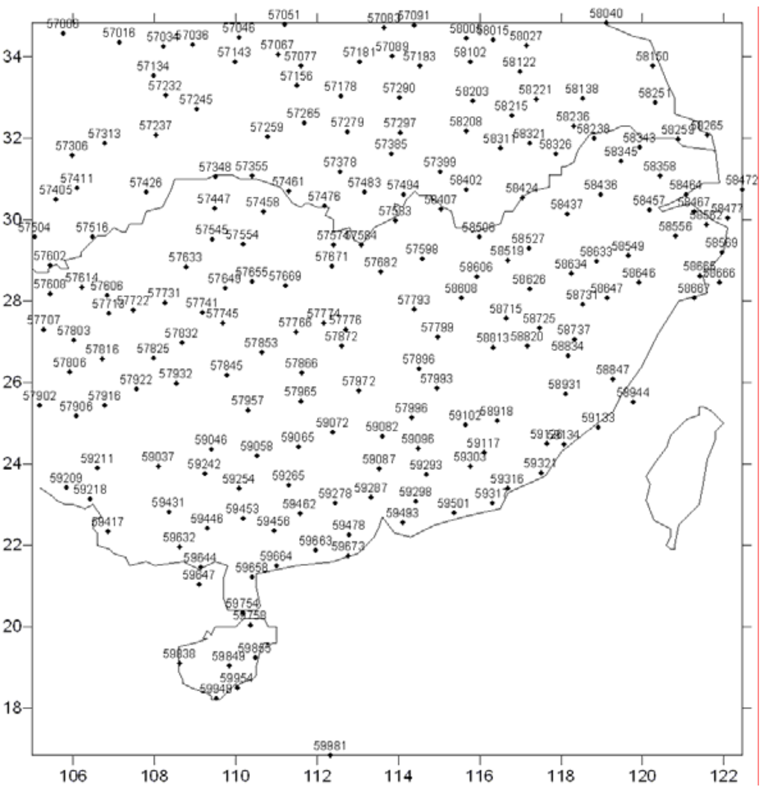

Fig. 1. Distribution of temperature stations in Southeastern China. The station identification number is provided above each location symbol. (x-axis: Longitude °E; y-axis: Latitude °N)

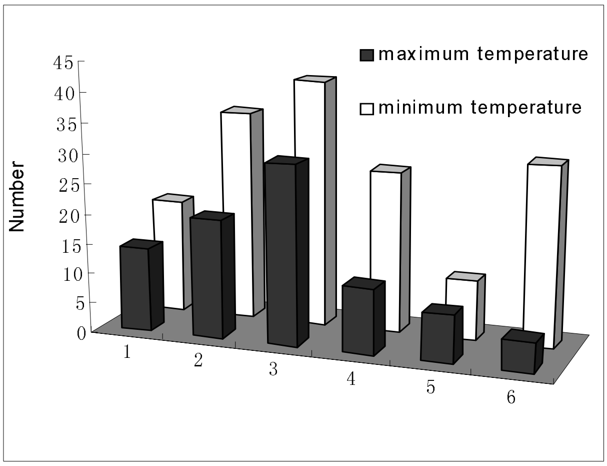

Fig. 2. Distribution of changepoints by magnitude (in °C): the numbers 1 through 6 refer to magnitude (a) estimates of $|a| < 0.1, 0.1 \leqslant |a| < 0.2, 0.2 \leqslant |a| < 0.3, 0.3 \leqslant |a| < 0.4, 0.4 \leqslant |a| < 0.5, |a| \geqslant 0.5$, respectively.

ently, as discussed in section 5, the area seems well suited for a simple sensitivity analysis whereby the impact of the number of series used to build the composite reference can also be evaluated.

4. Results

4.1 *Annual average maximum temperature*

By the consensus of the three test methods, a total of 106 discontinuities (61.3% of them were detected by SNHT, 76.4% by TPR, and 92.4% by Lag1) were detected in 92 of the annual average maximum temperature time series, or in 45.5% of the series evaluated. Thus, even the best subset of temperature series from Southeastern China is characterized by a large number of relative discontinuities. Confidence limits for change point magnitude were calculated for only 88 of the 106 maximum temperature change points. As shown in Fig. 2, the estimated magnitude of relative change points can be rather small. In fact, 53 of the 88 change points were estimated at below 0.2°C, while 29 were between 0.2°C and 0.5°C, and 5 were between 0.5°C and 1.0°C. Only one maximum temperature discontinuity was estimated at greater than 1.0°C.

Given this distribution of change point magnitude, the impact of adjustments on a regional scale is fairly small. The average linear trend for all series combined using raw and adjusted values is 0.042°C (10 yr)$^{-1}$ and 0.049°C (10 yr)$^{-1}$, respectively. Nevertheless, at smaller spatial scales and especially for a single series, the difference may be much greater. The maximum temperature series from Wufeng (57458), which had the largest estimated discontinuity (−2.35°C), is an example. In Fig. 3, both adjusted and unadjusted annual maximum temperatures from this site are shown. The linear trend in average maximum temperatures from this station based on unadjusted values is 0.583°C (10 yr)$^{-1}$ and is a spatial outlier. The trend based on the adjusted series is only 0.065°C (10 yr)$^{-1}$, which is consistent with trends at other locations in the vicinity. A search of the available metadata archives indicated that the station was relocated in 1994 to an elevation about 290 m lower than the former location.

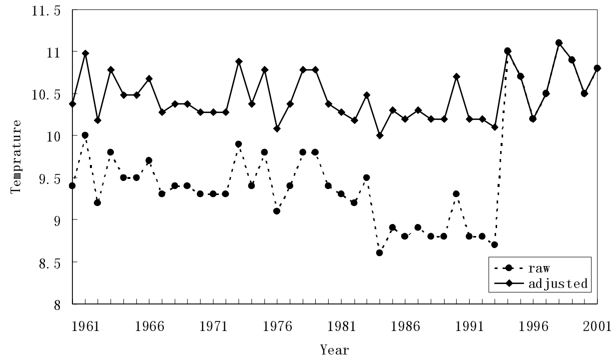

Fig. 3. Mean annual daily maximum temperature time series at Wufeng (57458) before (dashed) and after (solid) change point adjustments.

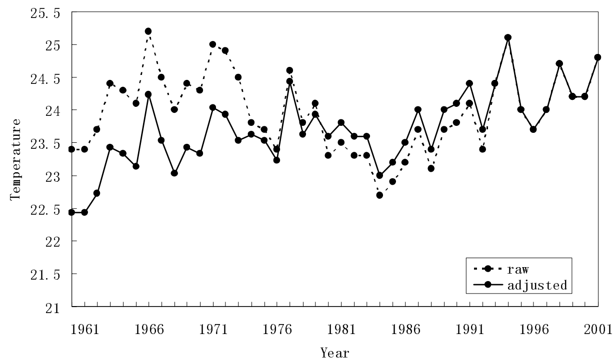

Fig. 4. Mean annual daily maximum temperature time series at Dongshan (59321) before (dashed) and after (solid) changepoint adjustment.

The maximum temperature series from Dongshan (59321) is an example of a series having multiple relative change points, in this case in the years 1973, 1979, and 1992. Breakpoint magnitudes were estimated at 0.97°C, 0.17°C and −0.30°C, respectively. There is no corresponding relocation information in the metadata archives, so the causes of these changes are unknown. Figure 4 similarly shows a comparison between adjusted and unadjusted values. In this example, the linear trends were calculated at −0.028°C (10 yr)$^{-1}$ and 0.322°C (10 yr)$^{-1}$ using unadjusted and adjusted

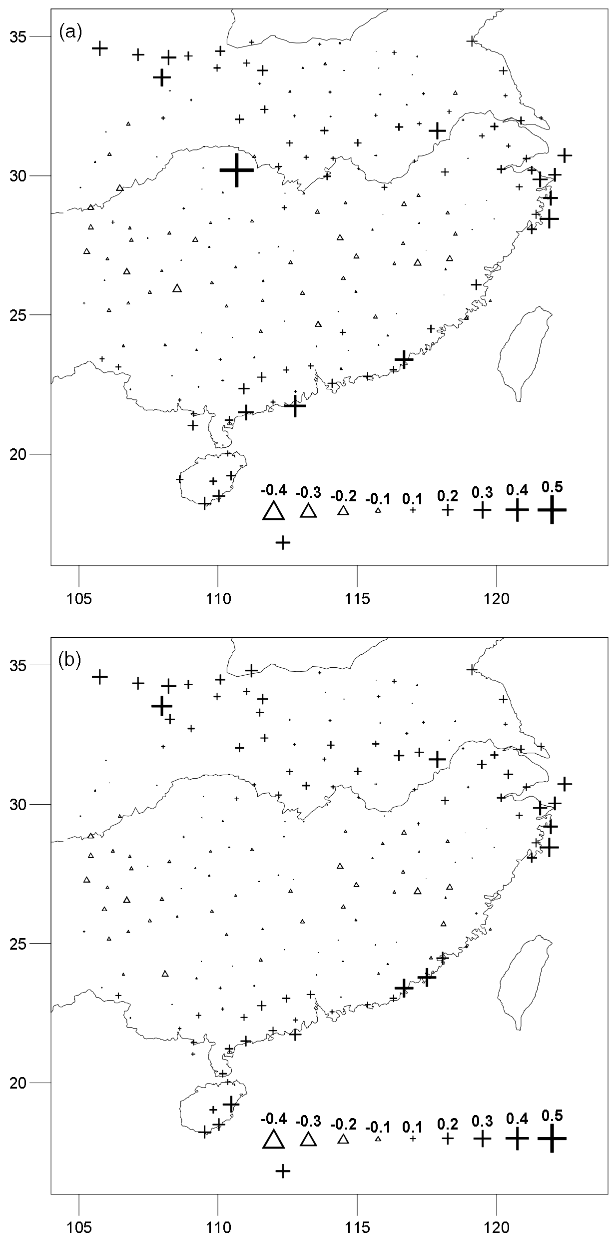

Fig. 5. Geographic distribution of the linear trend in 1960 to 2001 in average annual daily maximum temperature (a) before and (b) after adjustment [units: °C (10 yr)$^{-1}$]. (x-axis: Longitude °E; y-axis: Latitude °N)

values, respectively. As in the previous example, the trend in adjusted values agrees much more closely with series from nearby stations including Shantou (59316), which has no detected change points during the period and a linear trend in annual maximum temperatures of 0.303°C (10 yr)$^{-1}$ during the period.

A clear pattern in the geographic distribution of maximum temperature trends is evident even when raw values are used, as shown in Fig. 5a. Positive trends are apparent in the southern regions, the eastern coastal areas and in the Yangtze River Basin, and areas immediately to the north. In other areas, linear trends over the period of record are small or in some cases, negative. Nevertheless, some geographic outliers are evident, which are likely caused by abrupt shifts such as those in the series from Wufeng and Dongshan. A similar trend pattern also is evident in Fig. 5b, in this case based on adjusted values, but the obvious spatial outliers have been eliminated by the homogenization procedures.

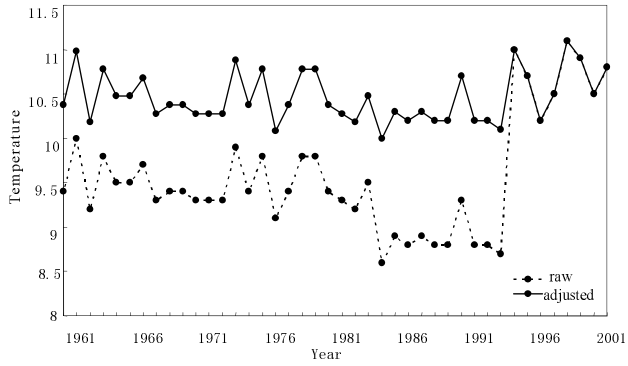

Fig. 6. Mean annual daily minimum temperature (°C) time series at Wufeng (57458) before (dashed) and after (solid) changepoint adjustments.

4.2 *Annual average minimum temperature*

A total of 196 relative change points (similar with the percentage from maximum temperature, 65.3% of them were detected by SNHT, 70.4% by TPR, and 89.2% by Lag1) were revealed in the sequence of annual minimum temperature values from 138 different stations, or, in about 68.3% of all minimum temperature series. Confidence limits for change point magnitude were estimated for 162 of the 196 discontinuities. The mean change point magnitude is somewhat larger than in the case of maximum temperature series adjustments (Fig. 2). Specifically, 63 of the 162 minimum temperature change points were estimated to be less than 0.2°C, 74 were between 0.2°C and 0.5°C, 20°C were between 0.5°C and 1.0°C, and 3 were greater than 1.0°C. As in the maximum temperature time series, a large discontinuity (1994) also occurs in minimum temperature observations at Wufeng (57458), as shown in Fig. 6. There is another minor discontinuity found in 1984, but it is a little different compared with the results in the maximum temperature. We checked the metadata of the Wufeng station and found that the station was moved on Jan 1st 1994 from a mountainside (altitude: 908.4 m) to a nearby mountaintop (altitude: 619.9 m), but no metadata record support the 1984 discontinuity. Likewise, the geographical distribution of average annual minimum temperature trends

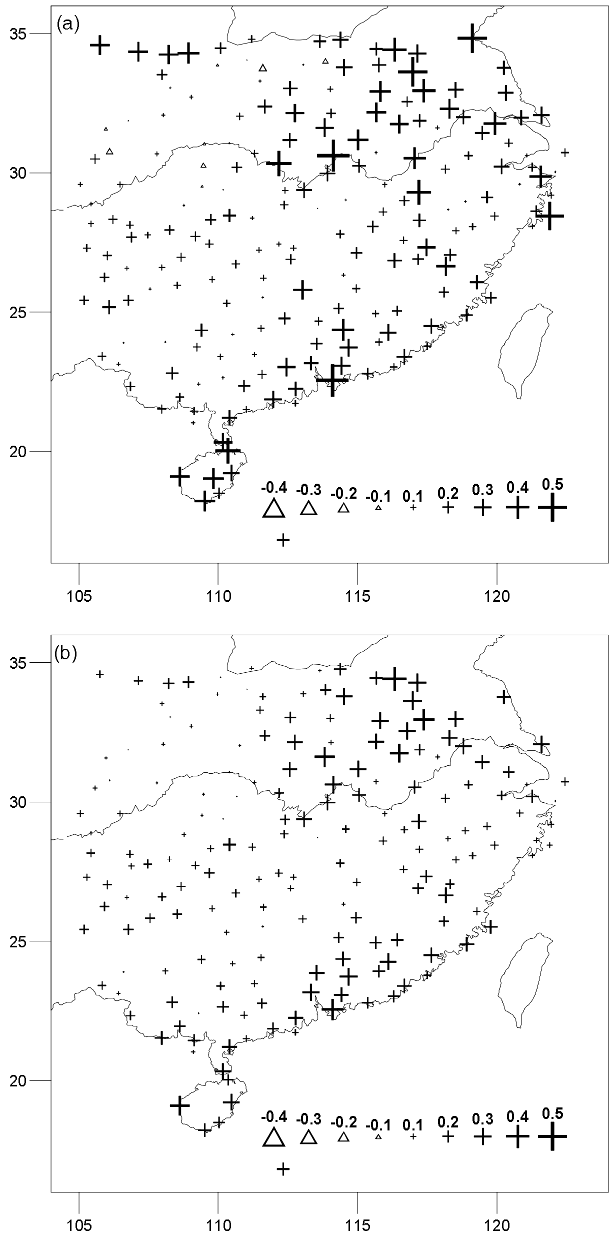

Fig. 7. Geographic distribution of the linear trend in 1960 to 2001 in average annual daily minimum temperature (a) before and (b) after adjustment [units: °C (10 yr)$^{-1}$]. (x-axis: Longitude °E; y-axis: Latitude °N)

before adjustments is shown in Fig. 7. As in the case of maximum temperature trends (Fig. 4), a number of the 1960 to 2001 linear trends appear to be spatial outliers. After the homogenization procedure, however, most of the linear trends are in agreement with surrounding series. As shown in Fig. 2, a larger number of change points were revealed in annual average minimum temperature series than in annual average maximum temperature series. Two possible causes are discussed below.

5. Discussion

5.1 *Homogenization and the nature of temperature fields*

Although discontinuities in temperature time series can be caused by any number of changes in, for example, sensor type (Guttman and Baker, 1996), the type of radiation shied (Hubbard et al., 2001) and /or rate of aspiration (e.g., Gall et al., 1992), and even the observation schedule (e.g., Baker, 1975), station relocations are the likely cause of the majority of abrupt shifts identified in the temperature series evaluated here. This is because the instrumentation system and other observing practices have remained basically unchanged in the Chinese surface observing network during the study period of record (Liu and Li, 2003). When a station is relocated, differences in the physical characteristics of the new and former instrument sites (e.g., topography, soil type, vegetation, surrounding man-made structures) will alter the exposure environment of the sensor. Moreover, minimum temperatures generally occur near sunrise when calm and stable atmospheric boundary layer conditions are prevalent. Under these conditions, near surface temperature fields are strongly coupled to the local surface characteristics (Oke, 1987). On the other hand, during daylight hours, the boundary layer is commonly well mixed, and microclimate differences between nearby sites may be less evident.

5.2 *Homogenization and local temperature trends*

There is evidence of an asymmetry in the climatologic trends of average daily maximum and average daily minimum temperatures across much of the world's land surface (Karl et al., 1991; Skinner and Gullett, 1993; Easterling et al., 1997; Jones et al., 2008). In general, recent decadal tendencies in maximum temperatures appear to be weakly positive or neutral whereas trends in minimum temperature are more strongly positive. Figures 4 and 6 suggest that this asymmetry is also present in surface temperatures throughout Southeastern China. The Urban Heat Island (UHI) effect may play some role in the asymmetric temperature trends because the UHI signal is generally strongest at night. While the UHI impact on globally averaged temperature trends may be minimal in recent decades (Parker, 2004), the UHI effect nevertheless may have a role in recent trends in Southeastern China (Zhou et al., 2004; Li et al., 2004b; Huang et al., 2004). It has been estimated that the UHI signal has increased the trends in annual mean temperatures in our study domain by about 0.011°C (10 yr)$^{-1}$ from the 1950s to present. They also concluded that about one-third of temperature series from stations in this region have a discernible UHI signal. Such sites are located in city centers having populations more than 50 000 or in or near cities with populations of more than 50 0000.

In many homogenization studies, including this one, a difference series between a target and its reference series is examined for abrupt changes only, even when a test method includes a slope parameter such as the TPR used here. No matter what test methods are used, however, step changes and trends may be easily confounded (Lund and Reeves, 2002). Local trends, if present, may be identified under a "zero-slope" assumption of relative change points if the trend is large enough to force a statistically significant change in mean level across some interval of the difference series. If a local trend is induced by a gradual UHI signal at the target site, the signal could be removed, at least in part, as an (albeit erroneous) abrupt change somewhere in the interval that encompasses the trend.

The station at Shenzhen (59493) has been identified as having a large UHI signal (Zhou et al., 2004; Li et al., 2004b). The city of Shenzhen has experienced the fastest population growth in China, growing from about 0.1 million inhabitants in 1982 to about 7 million in 2000 (Zhou et al., 2004). We us the temperature records from this site as an example of how trend and step changes may be confounded and show that the UHI signal is probably greater in minimum temperatures at this location. In our analysis, one relative change point was detected in the annual maximum temperature series (in 1965) while three change points were detected in annual minimum temperature series (in 1974, 1991, and 1996). The magnitude was estimated in each case so an adjustment for all change points was possible. The linear trend in the maximum temperatures is rather small, whether calculated using adjusted or unadjusted values, and is consistent with neighboring trends. In contrast, the annual minimum temperature trend is 0.563°C (10 yr)$^{-1}$ before change point adjustment and 0.424°C (10 yr)$^{-1}$ after adjustment. Both calculated minimum temperature trends appear to be spatial outliers, probably because of a significant UHI signal. It is also likely that one or more of the minimum temperature series changepoints was forced by the presence of this local trend. As shown in Fig. 8, a trend in the difference series between minimum temperatures at Shenzhen and its reference is evident beginning around the year 1980, near the start of rapid urbanization. The step changes attributed to the years 1991 and 1996 are probably artifacts of the trend. Nevertheless, in this case, much of the possible UHI signal appears to have survived the homogenization process.

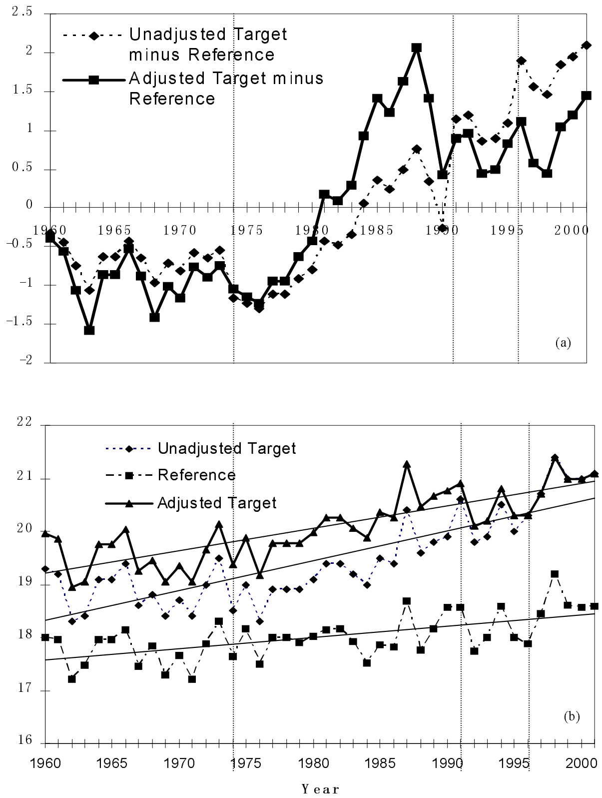

Fig. 8. (a) Standardized difference series (z-score) between the raw and adjusted annual minimum temperature series at Shenzhen and a weighted average reference series computed using temperature values from surrounding stations; (b) raw and adjusted minimum temperature series (in °C) at Shenzhen and its reference series. Dotted vertical lines designate the instants of apparent changepoints.

Whatever the cause, the presence of local trends in minimum temperature series from Southeast China or elsewhere may be misinterpreted as changes in mean level and may increase the number of apparent minimum temperature change points. On the other hand, if local and regional trends are approximately equal in magnitude (whether near zero or otherwise), little or no trend should be present in the series of differences between target and reference. In that case, the UHI signal (if non-zero) or other type of trend should pass relatively unaltered though the homogenization process. To help avoid confounding step changes with trend changes, the TPR described by Wang (2003) may be used as an additional test method. When changepoints are identified, the best method for a changepoint could then be determined using measures such as those described by Akaike (1974) or Schwarz (1978). Although beyond the scope of this paper, the example in Fig. 8 illustrates that identifying the correct method for a breakpoint is necessary even when a regional reference series is used.

5.3 *Homogenization and the number of reference series components*

Given the conclusions by Zhou et al. (2004) and the high station density present in Southeast China, 20 temperature series from nearby stations were used to construct the composite reference series. In other regions of the country, however, it is problematic to require a large number of component series to build the composite reference. A simple sensitivity analysis was therefore conducted to qualitatively assess the effect of varying the number of the reference stations on change point detection.

Using the same sample of temperature series in southeastern China, 5, 10 and 20 temperature series from stations nearest each target site were used to calculate the composite reference series. These three versions of the reference then were subtracted from each target temperature series and the sequences of differences were evaluated for changepoints using the three test methods. As shown in Table 1, the number of discontinuities in the difference series increases systematically with a decrease in number of component series used to build the reference. Not surprisingly, it appears that because the weight of any one-component series increases when fewer are used, the likelihood also increases that a changepoint in one of the component series will cause a changepoint in the reference. This effect was also observed in calculations using simulated temperature series generated according to the method described in Menne and Williams (2005).

6. Conclusions

An evaluation of temperature series from stations in Southeast China revealed that there are extensive heterogeneities even in the best subset of temperature series from the country. Some relative changepoints were estimated to be rather large (>2°C), and will impact, among other things, estimates of trends. A homogenization process is therefore critical in local and regional climate change studies. In addition, there are more relative changepoints in annual minimum temperature values than in maximum temperatures. It appears that the physical characteristics of minimum temperature fields make them more sensitive to relative changepoints than maximum temperature fields based on considerations of boundary layer mixing. During nighttime hours, the atmosphere is more stable and relatively minor station changes (e.g., relocation, instrumentation changes) can cause obvious changes to the field. On the other hand, the atmosphere is more fully mixed during daylight hours because of incoming short-wave radiation, which likely increases the spatial homogeneity of temperature fields. The greater spatial homogeneity of daytime temperature fields likely masks the consequences of minor station changes and possibly some of the impacts of urbanization.

Table 1. Number of discontinuities detected using different numbers of reference series components to calculate the composite.

Test method	Number reference series components		
	5	10	20
TPR	160	153	119
SNH	241	208	191
Lag1	281	242	201

While only undocumented changepoints were considered in this study, good metadata are invaluable to climate series homogenization (Aguilar et al., 2003). With metadata records, the dates of changepoint risks are known. This knowledge can improve the power of changepoint tests by eliminating the date of change as an unknown parameter (Lund and Reeves, 2002). Metadata records can also be useful in avoiding the attribution of an undocumented changepoint in the composite reference series to the target series. Unfortunately metadata records are generally incomplete and, for some locations, entirely unavailable. Consequently, undocumented changepoint evaluation is necessary in most circumstances. A multi-test approach may help to avoid the misattribution of some undocumented change points, however, a more reliable method is probably a pairwise comparison of series. In addition, some new techniques were brought forward in recent years (Wang et al., 2007; Wang, 2008), which gave an entirely new eye on the homogenization of the climatic series. More detailed comparisons of different methods and the identification of the appropriate model for a given breakpoint, are the subjects of ongoing investigations and will be discussed in future reports.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (40605021) and National Science and Technology Supporting Item project (2007BAC29B01). Partial support was also provided by the U.S. Department of Energy, Office of Biological and Environmental Research. Most of the work was conducted when the first author was at NOAA/NCDC as a research scholar. We would like to thank Drs. David R. Easterling and Russell S. Vose for reviewing an early version of this manuscript. We also thank Dr. Thomas C. Peterson for his support during the first author's visit to NCDC. Drs. Methew Menne, Claude N. Williams Jr. and Bomin Sun have given much support to the manuscript. And Dr. X. L. Wang from the Canada Meteorological Services has given many constructive comments and suggestions to this paper. Dr. Lucie Vincent and another anonymous reviewer presented much improvement on the paper both in content and language.

REFERENCES

Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa, 2003: Guidance on metadata and homogeneity. WMO Publication TD No. 1186, World Meteorological Organization, Geneva. 50pp.

Akaike, H., 1974: A new look at the statistical identification model. *IEEE Trans. Auto. Control*, **19**, 716–723.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *International Journal of Climatology*, **6**, 661–675.

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *International Journal of Climatology*, **17**, 25–34.

Baker, D. G., 1975: Effect of observation time on mean temperature estimation. *J. Appl. Meteor.*, **14**, 471–476.

Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Journal Royal Statistical Societies Series C*, **53**(3), 405–425.

Chinese Meteorological Administration, 1979: *The Rules of Surface Meteorological Observation*. The Meteorological Press, Beijing, 186pp. (in Chinese)

Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *International Journal of Climatology*, **15**, 369–377.

Easterling, D. R., and Coauthors, 1997: Maximum and minimum temperature trends for the globe. *Science*, **277**, 364–367.

Gall, R., K. Young, R. Schotland, and J. Schmitz, 1992: The recent maximum temperature anomalies in Tucson: Are they real or an instrumental problem? *J. Climate*, **5**, 657–665.

Guttman, N. B., and C. B. Baker, 1996: Exploratory analysis of the difference between temperature observations recorded by ASOS and conventional methods. *Bull. Amer. Meteor. Soc.*, **77**, 2865–2873.

Huang, J., X. Liu, and Q. Li, 2004: The study of relationship between heat island effect and population in cities over south of China. *Journal of Tropical Meteorology*, **20**, 713–722. (in Chinese)

Hawkins, D. M., 1976: Point estimation of the parameters of a piecewise regression model. *Appllied Statistics*, **25**, 51–57.

Hubbard, K. G., X. Lin, and E. A. Walter-Shea, 2001: The effectiveness of the ASOS, MMTS, Gill, and CRS air temperature radiation shields. *J. Atmos. Oceanic Technol.*, **18**, 851–864.

Jones, P. D., S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly, and T. M. L. Wigley, 1986: Northern Hemisphere surface air temperature variations: 1851–84. *J. Climate Appl. Meteor.*, **25**, 161–179.

Jones, P. D., D. H. Lister, and Q. X. Li, 2008: Urbanization effects in large-scale temperature records, with an emphasis on China. *J. Geophys. Res.*, **113**, D16122: 1–12.

Karl, T. R., and C. N. Williams, 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.*, **26**, 1744–1763.

Karl, T. R., G. Kukla, V. N. Razuvayev, M. J. Changery, R. G. Quayle, R. R. Heim, D. R. Easterliing, and C. B. Fu, 1991: Global warming: Evidence for asymmetric diurnal temperature change. *Geophys. Res. Lett.*, **18**, 2253–2256.

Li, Q., X. Liu, H. Zhang, T. C. Peterson, and D. R. Easterling, 2004a: Detecting and adjusting on temporal inhomogeneities in Chinese mean surface air temperature datasets. *Adv. Atmos. Sci.*, **21**, 260–268.

Li, Q., H. Zhang, X. Liu, and J. Huang, 2004b: UHI effect on annual mean temperature during recent 50 years in China. *Theoritical and Appllied Climatology*, **79**, 165–174.

Liu, X., and Q. Li, 2003: Research on inhomogeneity tests of climate data series in China. *Acta Meteorologica Sinica*, **17**, 492–502.

Lund, R., and J. Reeves, 2002: Detection of undocumented changepoints: A revision of the two-phase regression model. *J. Climate*, **15**, 2547–2554.

Menne, M. J., and C. N. Williams, 2005: Detection of undocumented change points using multiple test statistics and composite reference series. *J. Climate*, **18**, 4271–4286.

Oke, T. R., 1987: *Boundary Layer Climates*. 2nd ed., Routledge, London, 435pp.

Parker, D., 2004: Large-scale warming is not urban. *Nature*, **432**, 290–290.

Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *International Journal of Climatology*, **14**, 671–680.

Peterson, T. C., Coauthors 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, **18**, 1493–1517.

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review*, **109**, 2040–2045.

Schwarz, G., 1978: Estimating the dimension of a model. *Annals Statistics*, **6**, 461–464.

Skinner, W. R., and D. W. Gullett, 1993: Trends of daily maximum and minimum temperature in Canada during the past century. *Climatology Bulletin*, **27**, 63–77.

Szentimrey, T., 1999: Multiple analyses of series for homogenization (MASH). *Proc. Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary WMO-TD No. 962, 27–46.

Solow, A. R., 1987: Testing for climate change: An application of the two-phase regression model. *J. Climate Appl. Meteor.*, **26**, 1401–1405.

Vincent, L., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J.*

Climate, **11**, 1094–1104

Wang, X. L., 2003: Comments on "Detection of Undocumented Change points: A Revision of the Two-Phase Regression Model". *J. Climate*, **16**, 3383–3385.

Wang, X. L., 2008: Penalized maximal F-test for detecting undocumented mean-shift without trend-change. *J. Atmos. Oceanic Technol.*, **19**, 368–384

Wang, X. L., Q. H. Wen, and Y. Wu, 2007: Penalized maximal t-test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorological Climatology*, **46**, 916–931.

Yan, Z. W., C. Yang, and P. Jones, 2001: Influence of inhomogeneity on the estimation of mean and extreme temperature trends in Beijing and Shanghai. *Adv. Atmos. Sci.*, **18**, 309–321.

Zhai, P. M., and R. E. Eskridge, 1996: Analysis of inhomogeneities in radiosonde and humidity time series. *International Journal of Climatology*, **9**, 884–894.

Zhou, L., R. E. Dickinson, Y. Tian, J. Fang, Q. Li, R.K. Kaufmann, C. J. Tucker, and R. B. Myneni, 2004: Evidence for a significant urbanization effect on climate in China. *Proceedings of National Academy of Sciences of USA*, **101**, 9540–9544.