Homogenization of the Daily Land Skin Temperature (LST) over China from 1960 to 2017

Dan Wang[1, 2], Aihui Wang[1], and Xianghui Kong[1]

[1]Nansen-Zhu International Research Center, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China.

[2]University of Chinese Academy of Sciences, Beijing 100029, China.

Corresponding author: A. Wang (wangaihui@mail.iap.ac.cn)

Key Points:

- The warming biases of the daily LST in cold months in regions north of 40ºN have been corrected;
- The magnitude, interannual variability, and linear trend of LST have been reduced after homogenization;
- The annual mean LST after homogenization shows a significantly warming trend (0.22 ºC $decadal^{-1}$) during 1960-2017.

Abstract

Land skin temperature (LST) is one of the most important factors in the land-atmosphere interaction process. Raw measured LSTs may contain biases due to instrument replacement, changes in recording procedures, and other nonclimatic factors. This study attempts to reduce the above biases in raw daily measurements and achieves a homogenized daily LST dataset over China using 2360 stations from 1960 to 2017. The high-quality land surface air temperature (LSAT) dataset is used to correct the LST warming biases in cold months in regions north of 40ºN due to the replacement of observation instruments around 2004. Subsequently, the Multiple Analysis of Series for Homogenization (MASH) method is adopted to detect and then adjust the daily observed LST records. In total, 3.68×10^3 significant breakpoints in 1.65×10^6 monthly records are detected. A large number of these significant breakpoints are located over large parts of the Sichuan Basin and southern China. After MASH procedure, LSTs at more than 80% of the breakpoints are adjusted within +/- 0.5 ºC, and 10% of the breakpoints are adjusted over 1.5 ºC. Compared to the raw LST dataset over the whole domain, the homogenization significantly reduces the mean LST magnitude and its interannual variability as well as its linear trend at most stations. Finally, we preliminarily analyze the homogenized LST and find that the annual mean LST averaged across China shows a significant warming trend (0.22 ºC decadal^{-1}). The homogenized LST dataset can be further adopted for a variety of applications (e.g., model evaluation and extreme event characterization).

Key words: Land Skin Temperature (LST), daily observation, homogenization, the MASH method, variability

1 Introduction

Land skin temperature (LST) is a pivotal factor in the land-atmosphere interaction process. LST directly influences the sensible and latent heat fluxes from the land surface to the planetary boundary layer (Zeng et al., 2012; Zhou & Wen, 2016). Changes in LST and soil temperature can also be used as indicators of climate change

(McMichael & Burke, 1998; Qian et al., 2011). Xue et al. (2018) found that soil temperature has a relationship with subsequent downstream climate variables and can be used as a predictor of extreme hydrological events. Furthermore, several researchers have considered the difference between LST and land surface air temperature (LSAT) as a teleconnection factor to predict changes in precipitation in downstream regions (Xue et al., 2012; zhou & Huang, 2006). Therefore, through its local and teleconnection effects, LST plays a significant role in modulating climate and climatic variability.

Long-term instrumental LST data have been collected and stored for many decades in China. As one of the conventional observations, LST is liable to be influenced by changes in nonclimatic factors (e.g., station migration, equipment replacement, and changes in the environment around the station location) (Li, 2016). Of these factors, equipment replacement is always conducted at multiple stations simultaneously (e.g., see Figure 1 of Xu et al., 2019). Ren et al. (2013) compared the soil temperature observed from both manual and automatic techniques at the same stations during the same period in China and found that significant differences existed. Furthermore, the sites of approximately 80% of observation stations have been relocated at least once since 1950 due to the growth and expansion of cities (Cao et al., 2013). If we keep such inhomogeneities and directly analyze the raw data, which may inaccurately describe actual climate variations, then we will potentially reach incorrect conclusions, particularly in the representation of long-term trends and extremes (Peterson et al., 2002; Xue et al., 2012). Therefore, it is necessary to homogenize the raw observed LST before it can be used as the “ground truth” in various applications.

To homogenize the LST dataset, several methods have been developed and applied in different countries (e.g., Hu & Feng, 2003; Xu et al., 2019; Zhou et al., 2017). Hu and Feng (2003) applied a quality-control method to the in situ soil temperature in the United States and then reproduced a high-quality soil temperature dataset at multiple soil layers. Zhou et al. (2017) applied the RHtest software package to homogenize a monthly observed LST dataset from approximately 2200 stations from 1979 to 2003 in China and then used the homogenized LST to assess eight reanalysis products. Xu et al. (2019) incorporated additional ancillary and metadata into a raw LST dataset and then constructed a homogenized monthly LST dataset from 686 meteorological stations in China. Previous studies have mainly focused on homogenizing LST datasets on a monthly time scale. The daily LST can describe temporal variations at synoptic scales and is much needed in weather forecasts as well as in modeling applications (Wu & Zhang, 2014; Xue et al., 2012). For instance, the daily LST from atmospheric reanalysis products has been widely used in climate research, but most of these projects underestimate the LST in China (e.g., Xu et al., 2019; Zhou & Wang, 2016; Zhou et al., 2017). Therefore, a publicly accessible homogenized and high-quality in situ daily LST dataset in China is urgently needed.

Since the 1980s, various methods and techniques have been developed and used to eliminate data inhomogeneities (Aguilar et al., 2003; Cao & Yan, 2012; Peterson et al., 1998). However, there is no universal homogenization method for all climatic elements across different spatial and temporal scales. Each homogenizing method has its advantages and limitations, and the choice and applicability of one method versus another is often dictated by a few influencing factors, such as the station density, availability of metadata, and type of variable (Peterson et al., 1998). Globally, two methods are popularly used to homogenize in situ observations. One is the RHtest method (Wang & Feng, 2013; Wang et al., 2010) and the other is the Multiple Analysis of Series for Homogenization (MASH, Szentimrey, 1999, 2013, 2014). The RHtest method applies a two-phase regression model to calculate linear trends of a target time series, detect their statistically significant breakpoints, and finally, adjust them using the quantile-matching adjustment method. The MASH method uses a point-to-point comparison to detect breakpoints through an intercomparison of station observations within the same climatic area. Then, the method homogenizes the breakpoints relied on its neighbors (a detailed description is provided in section 2.2). Because the MASH method is not very restricted by metadata, it is widely used to homogenize meteorological variables. Li and Yan (2010) applied the MASH method to the daily air temperature of Beijing with additional reference information (with or without metadata) and found that the MASH method could successfully detect the majority of inhomogeneities in the raw dataset. For a long-term historical station observation dataset, it is usually difficult to preserve and obtain metadata (Tao et al., 2004), and this is the case for the in situ measured LST dataset in China. Nevertheless, the MASH method allows users to perform data homogenization in the case of no metadata. As a result, we select the MASH method in the present work.

In this study, we aim to develop a high-quality LST dataset that can be used in future applications for long-term climatic and soil-related research. Using the MASH method, we examine and identify the inhomogeneities in the daily in situ collected LST from 2360 stations in the Chinese mainland. The manuscript is organized as follows. Section 2 describes the dataset and methods used in this study. Section 3 presents the detailed homogenization procedure. Section 4 demonstrates and discusses the changes in LST due to homogenization. The conclusions are presented in section 5.

2 Data and Methods

2.1 In situ observed LST

The raw daily LST dataset from meteorological stations in China was obtained from the National Meteorological Information Center of the China Meteorological Administration (NMIC/CMA) from 1960 to 2017. Fundamental quality control (e.g., screening for unreasonable extreme values) was performed on this dataset before it was released to the public (http://data.cma.cn/data). In China, a large number of meteorological stations were established in the early 1950s, and the station density and records were maintained with relatively high continuity and stability after 1960 (Cao et al., 2016; Shi et al., 2014). Currently, there exist more than 2400 national meteorological stations in mainland China. LST is one of the factors regularly measured and collected every day at the meteorological stations. Before 2000, LST was recorded and stored manually four times a day (at 02:00, 08:00, 14:00, and 20:00 Beijing time) with a surface and bent stem earth mercurial thermometer. After 2000, the observation system was updated gradually to employ a platinum resistance temperature sensor, which is an automatic instrument and is more effective than mercurial thermometers. After the update, the observation frequency became hourly (twenty-four times a day, Ren et al., 2013). The renewal period is concentrated during 2000-2007 across mainland China, and the observation system north of 40ºN was updated in 2004 (see Figure 1 of Xu et al., 2019). The arithmetic average of the observations collected at multiple times within a day is regarded as the daily LST in the data collection.

In the present work, daily LST measurements from up to 2360 meteorological stations are used. The majority of stations are located in relatively low altitudinal regions, which are usually densely populated (Figure 1a). For example, there is an obviously higher density of stations in the Yellow and Yangtze River Basins, Southeast China and the Central China Plain than in other regions. The station distribution is relatively sparse in high altitudinal and sparsely populated areas, such as the Northwest China and Qinghai-Tibet Plateau regions. In addition, the number of effective stations (i.e., stations with measurements) was 1476 in 1960, after which it increased generally over time (stable during 1980-2010). In 2017 the number of effective stations increased to 2628 (Figure 1b).

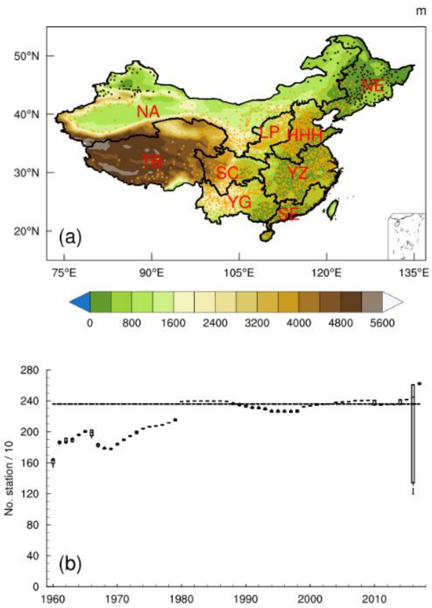

Figure 1. (a) Location of the meteorological stations (dots) and the division of the nine subregions (black curves). The black dots (a total of 197 dots) denote the specific stations described in Section 3.1; **(b)** the number of the daily nonmissing stations in

each year. The top (bottom) of the whisker plot represents the 90th (10th) percentile of the station number, the top (bottom) of the box represents the 75th (25th) percentile, and the middle line represents the 50th percentile. The straight line indicates the final 2360 stations used in the current study. (Additional information for the stations is shown in **Table 1.**)

2.2 Multiple Analysis of Series for Homogenization (MASH)

The MASH method was developed by Szentimrey (1999, 2014) and is based on the hypothesis test method to detect possible breakpoints at a given significance level. Through comparisons of measurements at different stations within the same climate region, the MASH method does not require prior assumptions of a homogeneous time series. This method has been widely applied to detect and adjust the inhomogeneity of raw meteorological observations at ground stations (e.g., Li et al., 2018, 2020). Guijarro et al. (2017) tested nine commonly used homogenization methods (including the RHtest and MASH methods) and compared their homogenized performances in terms of the root mean square errors and trends of air temperature time series. They found that the MASH method produced more reliable results than other methods (Guijarro et al., 2017).

To homogenize daily records, the MASH method must first homogenize the corresponding monthly data. Therefore, the daily LST at each station is first aggregated to monthly values, and then, the breakpoints of the monthly LST are detected and adjusted. Finally, the MASH method is again applied to homogenize the daily LST with the incorporation of the homogenized monthly values. Detailed information of the MASH method, including the mathematics it uses and its technique, is provided in its online manual (https://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/). The latest version of MASH, v3.03, is used in this study. According to the description in the MASH manual, two different models can be selected and used. One is the additive model, which requires that the targeted dataset has a normal distribution (e.g., temperature). The other is the multiplicative model and can be used for quasi lognormally distributed data series (e.g., precipitation). In this study, we choose the additive model and define a Monte Carlo threshold of 0.05 as the significance level.

In the default MASH procedure, the monthly/annual value is set as missing if any missing day/month exists within that month/year. Following this criterion, many LST stations would be discarded, which would lead to the loss of a great deal of useful data because most stations only have a few days missing in a specific month. Furthermore, the loss of useful data is not the intention of data homogenization. Here, we use a lenient threshold condition for the missing value judgment to retain as many useful observation stations as possible. At each station, the monthly LST is set as a missing value when the daily LST is available for less than 9 days of the current month. Meanwhile, the annual LST is set as a missing value when more than 3 continuous monthly values are missing. Finally, we remove stations with available annual values totaling less than 30 years of the full 58 years. Eventually, 2360 stations remained and were homogenized by the MASH method (shown as the straight solid line in Figure 1b). According to the principle of MASH as well as the spatial variability of LST, we perform MASH in each individual climate region. Based on the natural conditions for agricultural production and also in consideration of climate characteristics, we divide the mainland China areas into nine subregions (Figure 1a): Huang-Huai-Hai Plain (HHH), Loess Plateau (LP), Middle-lower Yangtze Plain (YZ), Northeast China Plain (NE), Northern arid and semiarid region (NA), Qinghai Tibet Plateau (TP), Sichuan Basin and surrounding regions (SC), South China (SE), and Yunnan-Guizhou Plateau (YG). The shapefile of each subregion is available from the Institute of Geographical Sciences and Resources, Chinese Academy of Sciences (http://www.resdc.cn/data.aspx?DATAID=275), and the number of effective stations in each subregion is shown in Table 1.

Table 1. Number of stations in each subregion

Full name	Abbreviation	Number of stations

Huang-Huai-Hai Plain	HHH	414
Loess Plateau	LP	202
Middle-lower Yangtze Plain	YZ	488
Northeast China Plain	NE	183
Northern arid and semiarid region	NA	328
Qinghai Tibet Plateau	TP	90
Sichuan Basin and surrounding regions	SC	200
South China	SE	164
Yunnan-Guizhou Plateau	YG	291
Mainland China	China	2360

In addition, other statistical methods, including linear regression and standard deviation, are used to comparatively analyze the raw and homogenized LST dataset. Two-tailed Student's test is used to test the significance of these statistics.

3 Inhomogeneity in the raw climate time series

3.1 Preliminary adjustment of LSTs in northern China

As mentioned in the introduction section, automatic instruments for LST measurements began to replace manual ones in 2004 in northern China (mainly north of 40ºN, including Xinjiang Province and NE and NA subregions, shown as blue dots in Figure 1a), which resulted in LSTs increasing abruptly in cold months since 2005. To understand this result, we select one cold month (December) to analyze the time series of LST over northern China. Figure 2 shows the daily LST time series averaged across 197 stations in northern China in December from 1960 to 2017. There is a distinct jump in 2005, after which the magnitude of LST increases remarkably. The mean value is -8.1 ºC for 2005-2017, which is 8 ºC higher than that for 1960-2005 (-16.1 ºC). This phenomenon is prevalent at all 197 stations. Xu et al. (2019) found a similar problem in the same regions (the NE and NA subregions) when they compared the differences of LST measurements from manual and automatic instruments during parallel observation periods. The warm shift phenomenon is principally caused by the change of the observation system from a manual to an automatic one around 2004. In winter, when the ground surface is covered by snow, the automatic instrument sensor measures the temperature at the soil surface under snow, whereas the manual instrument measures the snow surface. Because snow provides strong insulation of heat, it can absorb both longwave radiation from the ground surface transmitted upward and shortwave radiation from the atmosphere transmitted downward. Therefore, the measured LST from the automatic instrument would change much more slowly under snow and cannot represent real LST change features (Liu et al., 2008; Ren et al., 2013). When the manual sensor was buried in snow, it was removed to measure the snow surface temperature instead of the soil surface temperature (Administration, 2003).

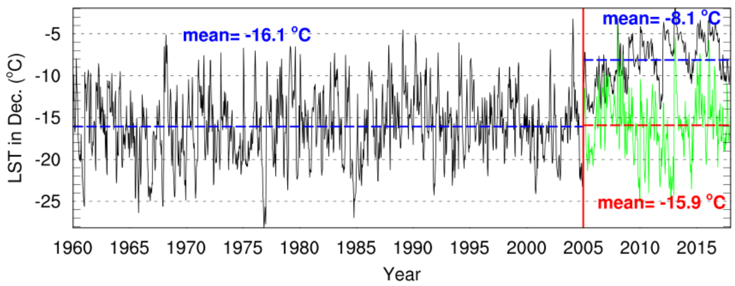

Figure 2. Time series of the raw daily LST averaged across 197 stations (the black dots in **Figure 1a**) in December for the period of 1960-2017. The red solid line is a reference line to separate the years before and after 2005; the two blue dotted lines are the mean LSTs derived from the raw daily dataset during 1960-2005 and 2005-2017. The green solid curve represents the daily LST time series after LSAT adjustments (Section 3.1), and the red dotted line is the mean LST during 2005-2017. The corresponding mean LST values averaged for different periods are also indicated.

More specifically, we take a station (ID 50862, 46.98ºE, 128.05ºN) at the NE subregion as an illustrative example. Figure 3 shows the abrupt warming shift in cold months (from October to the following April) since 2005. A similar phenomenon appears at all 197 stations (black dots in Figure 1a) in the NE and NA subregions. In the MASH procedure, possible breakpoints are determined by comparing the values at the candidate station with those at the nine nearest reference stations. If those surrounding reference stations show similarly abrupt changes or shifts as the candidate station, then the MASH method cannot detect them (Li, 2016). Due to how MASH works, it will fail to deal with the above problematic phenomenon (compared to the blue and orange lines in Figure 3b). Thus, before applying the MASH method, we perform a preprocessing procedure for the LSTs from these 197 stations.

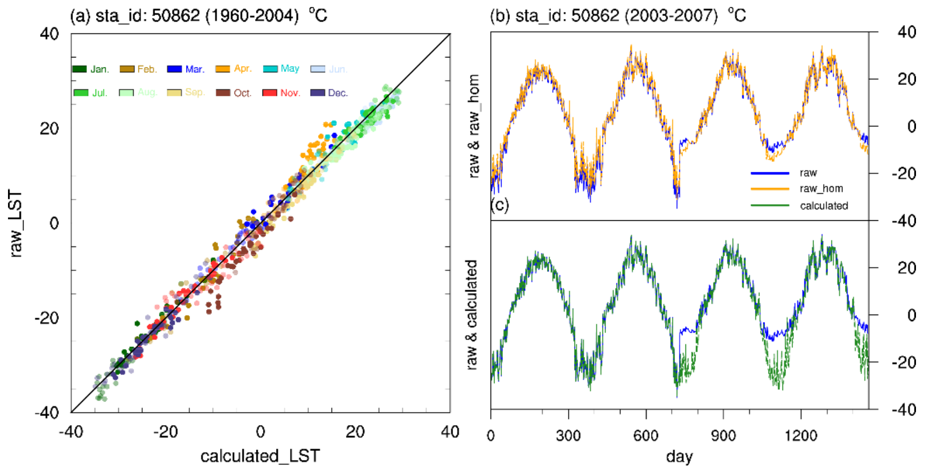

Figure 3. Case station (ID 50862, 46.98ºE, 128.05ºN) for the raw and calculated LST, and the homogenized results (unit: ºC). **(a)** Raw daily LST plotted against the calculated daily LST in each month during 1960-2004, where the black solid line is the reference (x=y) line; **(b)** raw (blue) and MASH homogenized (orange) daily LST, and **(c)** raw and calculated (green line) LST during 2003-2007. The calculated daily LST is the LST adjusted by the LSAT (described in section 3.1).

There is a very close relationship between the LSAT and LST, and their differences determine surface heat fluxes (Zeng et al., 2012). The observed LSATs have undergone a strict quality control process and do not show such remarkable warming shifts. Therefore, the LSAT is used as a reference to correct the LST warming bias at each station. Simplistically and practically, it is assumed that the characteristics of the changes in LSATs are consistent with those in LSTs at the same station. This assumption is reasonable because both LSAT and LST show similar variabilities. The following procedure is used to adjust the LST via the LSAT. First, we construct a regression equation between the LST (dependent variable) and LSAT (independent variable) for each cold month (from October to the following April) at each of the 197 stations from 1960 to 2004. Second, the regression coefficient and the raw daily LSAT in the current month are used to compute the corresponding daily LST from 2005 to 2017. The computed LSTs are regarded as the “calculated LSTs” at the 197 stations. To verify the reasonableness and applicability of the above hypothetical relationship, we compare the raw and calculated LSTs in Figure 3a. All the dots spread around the reference y-x line, and the monthly absolute mean differences between the calculated and raw LSTs are generally less than 1 ºC. In cold months, the absolute differences are less than 0.5 ºC, except in April (October), when there is a relatively large error in the mean differences of approximately -2 ºC (2 ºC). However, the standard deviation of the raw LST during 1960-2005 in April (October) is 3 ºC (6 ºC) and is higher than the mean difference. Therefore, it still can be concluded that the calculated LSTs are consistent with the raw LSTs during the whole period. Similar results are also shown at other stations. After the above process, the distinct warm shift of the LST in cold months is removed (Figure 3c). The calculated daily LST across all 197 stations averaged in December from 2005 to 2017 (green line in Figure 2) also shows consistency with the raw daily LST from 1960 to 2005, and the mean value for 2005-2017 (-15.9 ºC) is close to that for 1961-2005 (-16.1 ºC). The above calculated LSTs at 197 stations for 2005-2017 are added to the raw dataset and replace the values at the same time and station.

3.2 Breakpoint detection by MASH

After the above preliminary adjustments, we apply the MASH method at all 2360 stations in China. The first step of MASH is to detect the temporal breakpoints of the LST time series at each station. The breakpoint (also called the change-point) is where the time series displays significant differences before and after that point. The time series show a distinct leap at the breakpoint. To help understand this process, we take one station (ID 57710 in the YG subregion) as an illustrative example (Figure 4). The LST at the candidate station (the red dot in Figure 4a) is first compared with the values from the nine closest stations (the blue dots in Figure 4a). Compared with the other nine reference time series (Figure 4b), a breakpoint appears in 1970 in the candidate time series, of which the mean LST during 1960-1970 is 21.1 ºC, 3 ºC higher than that during 1970-2017 (18.1 ºC). The variability of the homogenized time series (the blue line in Figure 4c) is akin to the nine neighbor reference time series. The time series has a warming trend (0.09 ºC decadal^{-1}), showing the opposite tendency compared to the raw time series (-0.26 ºC decadal^{-1}).

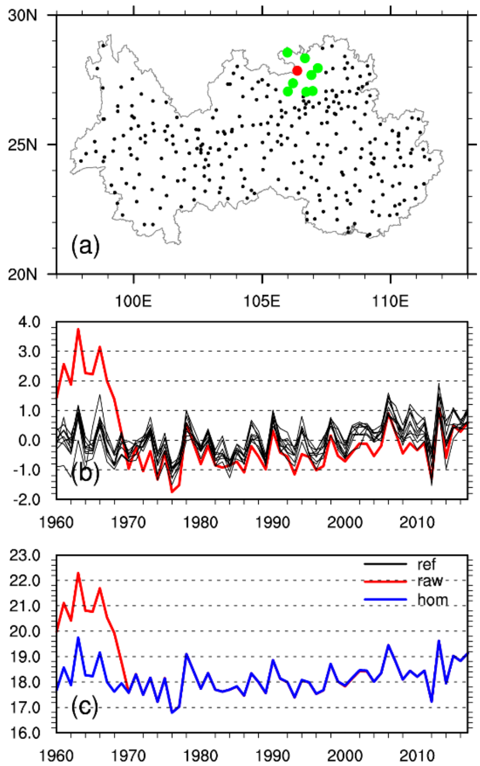

Figure 4. Annual LST series at station ID 57710 (27.85ºE, 106.37ºN, a candidate station) in the YG subregion (unit: ºC). **(a)** Distribution of stations, including the candidate station (red dot), the nine nearest reference stations (green dots), and other stations (black dots) in the YG subregion; **(b)** the raw annual LST anomalies of the candidate station (red curve) and the nine reference stations (black curves); **(c)** annual LST from the raw (red curve) and homogenized time series at the candidate station.

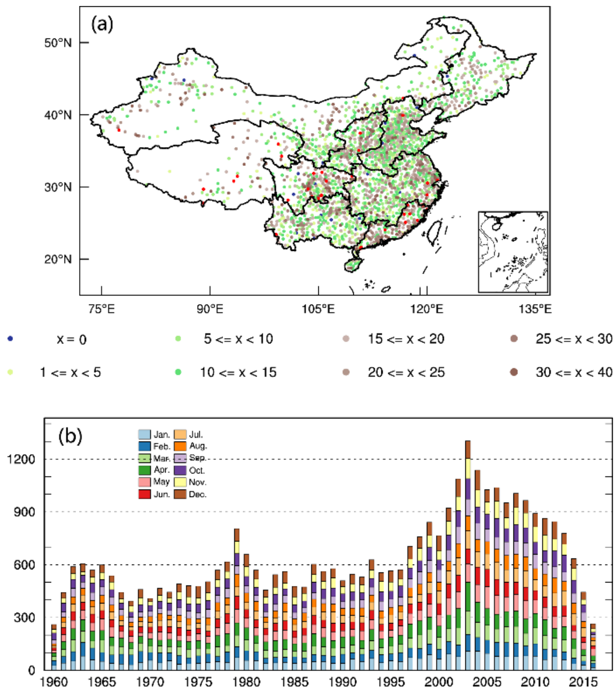

Figure 5. Number of breakpoints for the monthly LST records at the effective stations during 1960-2017. **(a)** Spatial distributions at all stations and **(b)** accumulation in each month and in each year across all stations.

Repeating the above procedure for all the stations in each of the nine subregions, we identify the breakpoints at all 2360 stations from the monthly LST time series for the period of 1960-2017. Note that only breakpoints with an absolute value higher than 0.5 ºC are treated as significant breakpoints, and there are a total of 3.68×10^3 significant breakpoints counted in Figure 5. The figure clearly shows that the monthly LSTs at the majority of stations contain 5-25 significant breakpoints (Figure 5). Of all 2360 stations, 35 stations contain over 40 significant breakpoints, of which most are located in the SC and SE subregions (the red dots in Figure 5a), while 21 stations scattered across China contain no significant breakpoints (the dark blue dots in Figure 5a). The breakpoints are relatively concentrated in the SE, HHH, and SC subregions, where there are at least 20 significant breakpoints identified in most stations. Having aggregated the numbers of significant breakpoints at all stations (Figure 5b), nearly 500 significant breakpoints occur each year on average. However, the number of significant breakpoints shows distinct interannual variations and evidently increases over time, in particular, since 2000. In 2003, when the replacement of the LST measurement instruments occurred at many stations throughout China, more than 1200 breakpoints are identified, and this number reaches the maximum value. For the monthly LST time series, the number of breakpoints in cold months (January to April, and November to December) is higher than in warm months (May to September) in every year. Previous documents and reports have also mentioned that the LST measurement instruments in in situ observation systems have been gradually changed from manual to automatic ones since 2000 (Liu et al., 2008; Xu et al., 2019). Furthermore, it is well known that urbanization has developed rapidly during recent decades in China. The fact that some stations suffered from changes in the surrounding environment may also induce abrupt changes in the LST records, which is another reasons for the increase of the number of breakpoints.

3.3 Frequency distribution of biases in MASH

To explore the LST biases identified in the MASH procedure, we count the number of breakpoints with an absolute bias over 0.5 ºC (hereafter called a significant adjustment) at all stations in each subregion and throughout China (Figure 6). There are approximate 1.6×10^{6} monthly records, of which 3.28×10^{5} records (approximately 20% of total records) show significant adjustments in MASH. In each subregion, the majority of significant adjustments vary between 0.5 and 1.5 ºC and between -0.5 and -1.5 ºC (2.95×10^{5} or 90% of all significant adjustments), within which two peaks are found around 0.5 ºC to 1.0 ºC and -0.5 ºC to -1.0 ºC. Except for the TP subregion, stations in other subregions show that the number of breakpoints with positive adjustments is greater than the number with negative ones. Moreover, absolute biases less than 0.5 ºC (hereafter called a minor adjustment) are also counted in Figure 7. We find that the minor adjustments account for 80% of the total adjustments in most subregions and in all of China, which is in accordance with the above results (records with significant adjustments account for approximately 20% of the total records). The YZ (NE) subregion in the cold season shows the highest (lowest) percentage of minor adjustments, above 90% (below 80%). The percentage of minor adjustments in the SE subregion from June to December is less than 60%. Approximately 80% of the total records in the HHH subregion have experienced minor adjustments. We also display the frequency distribution of daily adjusted records (figures not shown), and it exhibits similar characteristics compared to the monthly results (Figure 6).

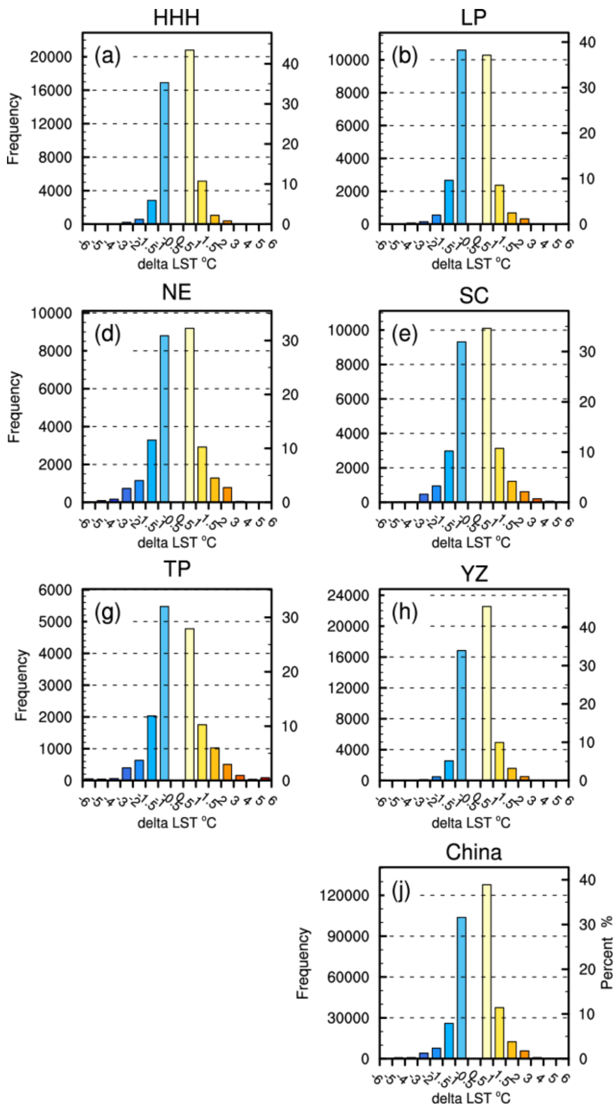

Figure 6. Frequency distribution of the significant breakpoints of LST in the nine subregions and mainland China. Biases at these breakpoints are adjusted with MASH.

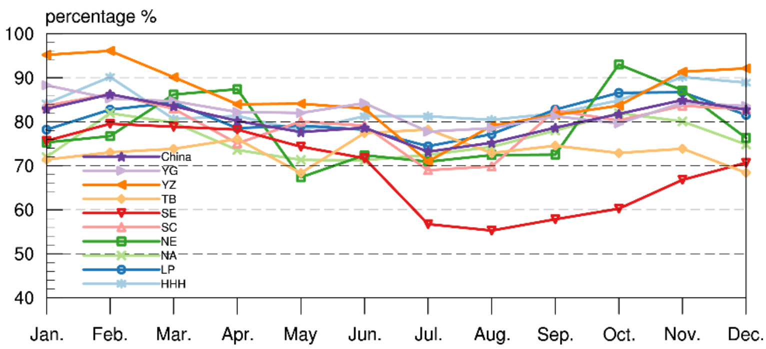

Figure 7. Percentage of records with adjusted LST biases between -0.5 ºC and 0.5 ºC in different months over the nine subregions and the Chinese mainland.

4 Characteristics of raw and homogenized LSTs during 1960-2017

To comprehensively understand the effects of inhomogeneities on the LST, we first compute three statistical metrics (mean, standard deviation, and linear trend) from both homogenized and raw LST at all stations, and we then compare their differences. To illustrate the results in different seasons, we use the statistical metrics in January and July of multiple years to represent the results in summer and winter, respectively (Figure 8).

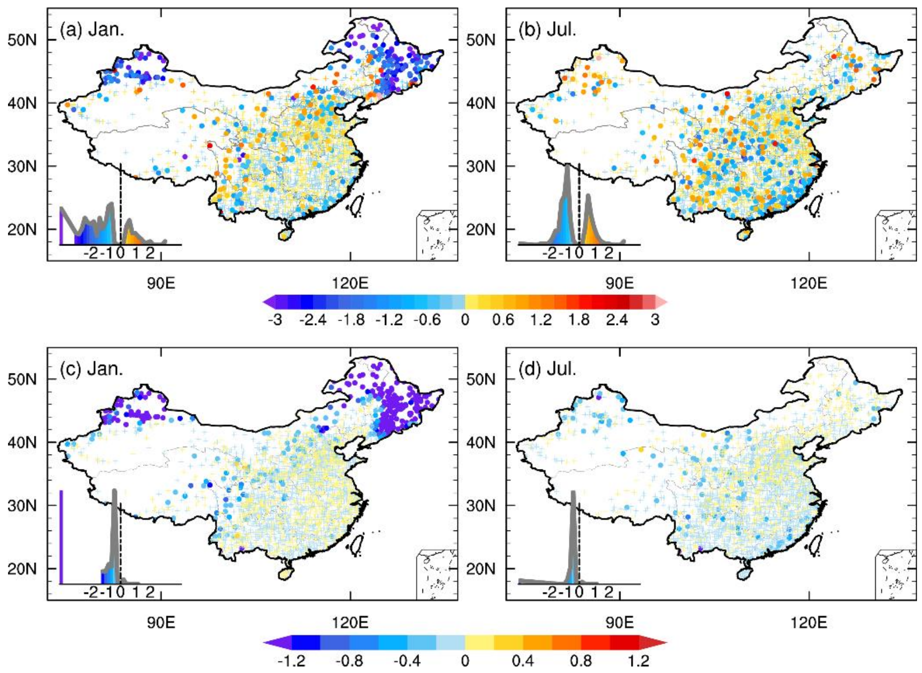

Figure 8. Spatial distribution of the differences of the monthly means (**(a)**, **(b)**) and standard deviations (**(c)**, **(d)**) between the homogenized and raw LSTs in January (**(a)**, **(c)**) and July (**(b)**, **(d)**). The filled dot and the '+' symbol indicate the mean LST differences passing or not passing the significance test (P=0.05), respectively. The histogram plot in the lower left corner shows the frequency distribution of significant differences (unit: ºC).

In winter, there are 81 (279) stations with significant positive (negative) differences between the homogenized and raw monthly mean LSTs (Figure 8a). Among these stations, 163 (58% of 279) stations with significant negative differences are located north of 40ºN, while some are in the SE subregion. In summer, significant differences between stations are intense in the YG, SC, SE, and NA subregions. Three hundred twenty-five stations have significant LST differences, of which 117 (208) stations have significant positive (negative) differences (Figure 8b). In both winter and summer, the significant negative differences mainly vary between -2 ºC and -1 ºC, while the significant positive differences are within 1 ºC (histogram plots in Figures. 8a and 8b).

We use the standard deviation to represent the interannual variability, and its difference between homogenized and raw data can be regarded as the changes of the interannual variability of LST. In winter, the interannual variability of the homogenized monthly LST is much smaller than that of the raw values (Figure 8c). All 197 stations north of 40ºN (Figure 1a) show negative differences. Stations with negative differences are also prevalent in the TP subregion. In summer (Figure 8d), the interannual variability of the homogenized mean exhibits a remarkable difference from that of the raw mean at 102 stations, of which 100 stations display negative differences, which indicates that our homogenization process reduces the interannual variability remarkably at most stations whether in winter or in summer.

To explore the changes in the long-term tendency of LST due to the MASH process, we also computed the linear trend for the homogenized and raw LSTs at all 2360 stations in winter and summer during 1960-2017. In winter (Figures 9a and 9b), 144 stations (6% of all 2360 stations) have negative linear trends in the raw dataset, of which only the trends at three stations are significant (p=0.05). Most stations show positive trends (94% of all 2360 stations), of which the LST trend at 57% of stations is significant (p=0.05). Stations with significant positive LST trends are concentrated in the NA, NE, LP, HHH, and TP subregions, and stations with negative trends are mainly located in the YG and YZ subregions. After homogenization, the number of stations with negative trends decreases from 144 to 45, and the LST trends at all these stations are not significant. The percentage of stations with significant positive LST trends

(57%) does not change, but the distribution differs greatly. Compared to the linear trend of raw LSTs, stations with significant positive trends are reduced north of 40ºN (the NE subregion and Xinjiang Province) but increase in the SE and TP subregions. In summer (Figures 9c and 9d), stations with negative trends increase and occupy almost the entire southern part of the Yellow River Basin. There are in total 648 stations with negative trends, and 12% of these are significant for the raw LSTs. This number increases to 659, while only 2% of them are significant for the homogenized LSTs, indicating that the homogenization process reduces the LST trends overall in mainland China.

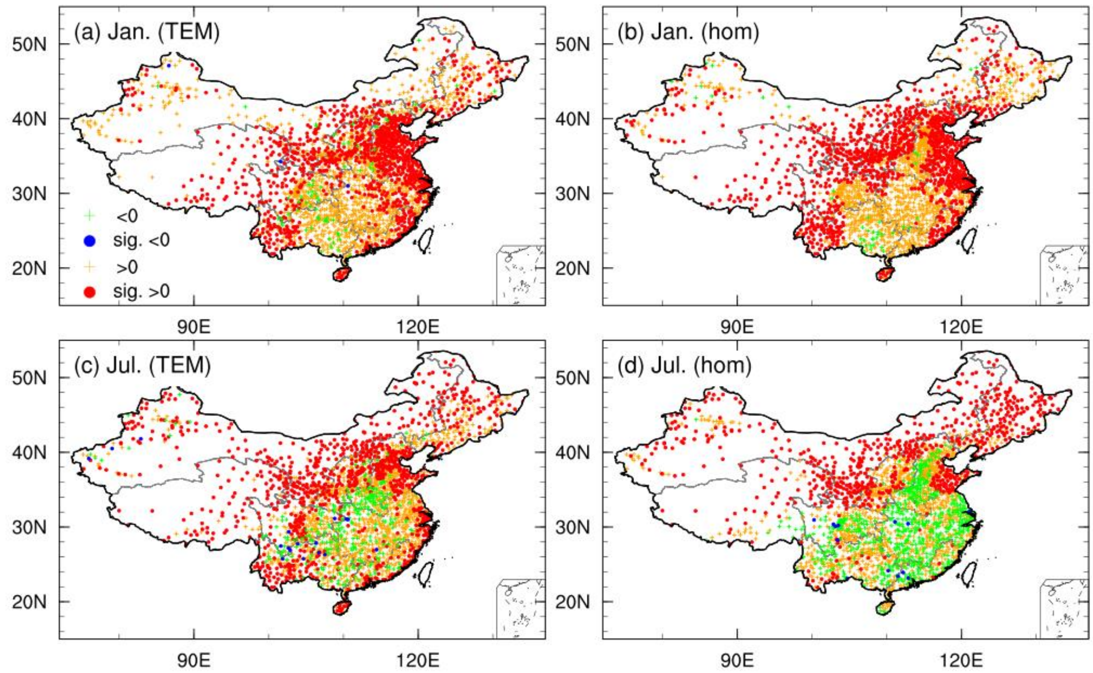

Figure 9. Distribution of linear trends of the monthly homogenized (**(b)** and **(d)**) and raw (**(a)** and **(c)**) LST in January (**(a)** and **(b)**) and July (**(c)** and **(d)**). The filled dot ('+') indicates stations with trends passing (not passing) the significance test (P=0.05).

Finally, we perform a preliminary analysis of the homogenized LST dataset in terms of its temporal and spatial variations. Figure 10 displays the distribution of the annual mean homogenized LST and its time series averaged in China. Notably, the annual mean LST has negative values at very few stations in the NE subregion, while most stations have positive values. Moreover, for the entire period of 1960-2017, the annual mean LST shows a significantly warming trend of approximately 0.22 ºC decadal^{-1}, which is consistent with previous studies (Xu et al., 2019; Zhou et al., 2017). The trend of the annual mean LSAT shows a similar magnitude (0.21 ºC decadal^{-1}), and the annual mean LST is generally 2 ºC higher than the annual mean LSAT.

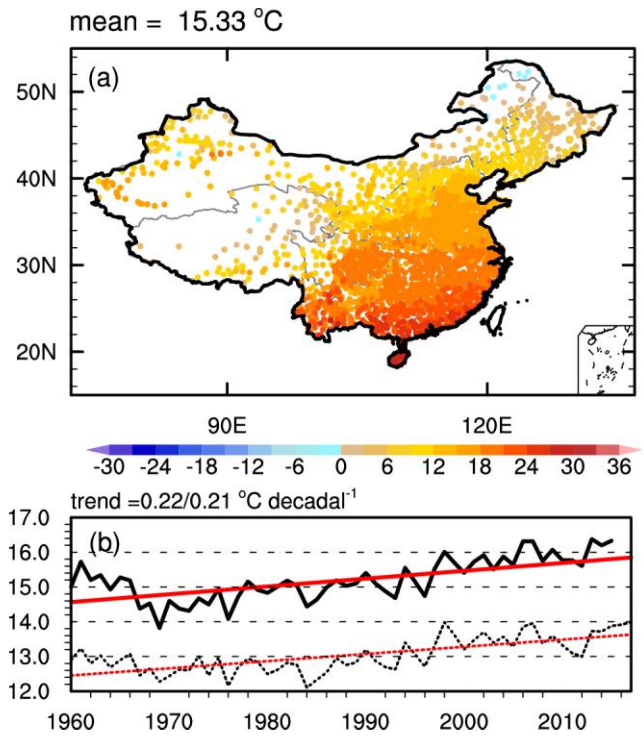

Figure 10. (a) Distribution of the multiyear mean homogenized LST and **(b)** its time series and the linear trend averaged over all stations in China (LST/LSAT with the solid line/dotted line) for the period of 1960-2017 (unit: ºC).

5 Summary and discussion

Long-term station observed datasets are the fundamental basis of climate research. A homogenized long-term LST record will greatly help us to thoroughly understand land-atmosphere interaction processes. This study is devoted to developing a homogenized long-term LST dataset from 2360 stations in the Chinese mainland for the period of 1960-2017 based on the NMIC/CMA in situ measured LST at meteorological stations. The main results are summarized as follows:

Due to the replacement of instruments from manual to automatic ones around 2004 north of 40ºN, the time series of the LSTs at 197 stations display remarkable warming shifts during cold months. This warming bias fails to be homogenized by the MASH method. Therefore, the raw LST records at the above stations are first adjusted using the high-quality LSAT under the assumption of the same monthly variability of both the LST and LSAT at the same station.

Then, the MASH method is applied to homogenize the LST dataset in China. During the period of 1960-2017, 5-25 significant breakpoints are detected in the LST monthly time series at most stations. Stations in southern China and most parts of the Sichuan Basin exhibited more intensive breakpoints than those in other subregions. The number of breakpoints has increased since 2000. Notably, in 2004, breakpoints are detected at more than 1200 stations. For most breakpoints, the absolute adjusted biases due to MASH are between 0.5 ºC and 1.5 ºC, with peaks around 0.5 ºC and -0.5 ºC. Moreover, comparing the difference between the homogenized and raw LSTs, the MASH process generally reduces the magnitude, interannual variability, and linear trends of LST. The interannual variabilities and linear trends of the homogenized LSTs are also reduced at the majority of stations, especially in the regions north of 40ºN in winter.

It should be noted that there are some limitations of the new homogenized LST dataset north of 40ºN. We preliminarily adjusted warm shifts of LSTs in cold seasons from 2005 onwards by briefly using a linear relationship between LST and LSAT, but many other factors can also directly affect changes in LST, including precipitation frequency, solar radiation, vegetation, and so on (Zhou et al., 2017). In a recently published article, Du et al. (2020) corrected LSTs through LSATs while considering the influences of snow depth and solar radiation. They found that the stable linear relationship between LST and LSAT is sensitive to the snow depth and solar radiation to some extent. Therefore, the above data uncertainty should be considered when this dataset is applied in research.

In summary, we provide a 58-year (1960-2017) homogenized daily LST dataset with abundant stations (2360) in China. The current paper only presents the homogenization process of the LST dataset and its preliminary analysis, but our work extends beyond this. We apply the MASH method on soil temperature at nine soil depths (0 cm, 5 cm, 10 cm, 15 cm, 20 cm, 40 cm, 80 cm, 160 cm and 320 cm). The results for other soil depths are similar to those for LST, so we do not present them in this paper. The long-term homogenized LST and soil temperature in different soil depths contain a relatively dense number of stations and can have a variety of applications, such as for model evaluation and climate and soil related research.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 41875106) and the National Science Fund for Distinguished Young Scholars (Grant No. 41925021). The data used are available from the National Meteorological Information Center of the China Meteorological Administration (NMIC/CMA) (http://data.cma.cn/data).

References

Administration, C. M. (2003), Specifications for surface meteorological observation. *Beijing: Meteorology Press*, 85-89

Aguilar, E., Auer, I., Boehm, R., Peterson, T. C., & Wieringa, J. (2003), Guidelines on climate metadata and homogenization. *WMO-TD No. 1186(1186)*

Cao, L. J., & Yan, Z. W. (2012), Progresses in Research of Homogenization of Climate Data. *Advances in Climate Change Research*, 3(2), 59-67. doi:10.3724/SP.J.1248.2012.00059

Cao, L. J., Zhu, Y. N., Tang, G., Yuan, F., & Yan, Z. (2016), Climatic warming in China according to a homogenized data set from 2419 stations. *International Journal of Climatology*, 36(13), 4384-4392. doi:10.1002/joc.4639

Cao, L. J., Zhao, P., Yan, Z. W., Jones, P., Zhu, Y. N., Yu, Y., & Tang, G. L. (2013), Instrumental temperature series in eastern and central China back to the nineteenth century. *Journal of Geophysical Research-Atmospheres*, 118(15), 8197-8207. doi:10.1002/jgrd.50615

Du, J. Z., Wang, K. C., Cui, B. S., & Jiang, S. J. (2020), Correction of inhomogeneities in observed land surface temperatures over China. *Journal of Climate*, 1-52. doi:10.1175/jcli-d-19-0521.1

Guijarro, J. A., López, J. A., Aguilar, E., Domonkos, P., Venema, V. K. C., Sigró, J., & Brunet, M. (2017), Comparison of homogenization packages applied to monthl series of temperature and precipitation: The multiltest project, Ninth Seminar for Homogenization and quality Control in Climatological Databases and Fourth Conference on Spatial Interpolation Techniques inClimatology and Meteorology. *Budapest, Hungary, WCDMP-No. 85*, 46-62

Hu, Q., & Feng, S. (2003), A daily soil temperature dataset and soil temperature climatology of the contiguous United States. *Journal of Applied Meteorology*, 42(8), 1139-1156. doi:10.1175/1520-0450(2003)042<1139:adstda>2.0.co;2

Li, Q. X. (2016), Climate data homogeneity studies in China: progresses and prospects. *Advances in meteorological science and technology*, 6(3), 67-74. doi: 10.3969/j.issn.2095-1973.2016.03.009

Li, Z., & Yan, Z. W. (2010), Application of multiple analysis of series for homogenization to Beijing daily temperature series (1960-2006). *Advances in Atmospheric Sciences*, 27(4), 777-787. doi:10.1007/s00376-009-9052-0

Li, Z., Yan, Z. W., Cao, L. J., & Jones, P. D. (2018), Further-adjusted long-term temperature series in China based on MASH. *Advances in Atmospheric Sciences*, 35(8), 909-917. doi:10.1007/s00376-018-7280-x

Li, Z., Yan, Z., Zhu, Y., Freychet, N., & Tett, S. (2020), Homogenized daily relative humidity series in China during 1960-2017. *Advances in Atmospheric Sciences*, 37(4), 318-327. doi:10.1007/s00376-020-9180-0

Liu, X. N., Ren, Z. H., & Wang, Y. (2008), Differences between automatic-observed and manual-observed surface temperature. *Journal of Applied Meteorolgical Science*, 19(5), 554-563

McMichael, B. L., & Burke, J. J. (1998), Soil temperature and root growth. *HortScience*, 33(6), 947-951. doi:10.21273/HORTSCI.33.6.947

Peterson, T. C., et al. (2002), Recent changes in climate extremes in the Caribbean region. *Journal of Geophysical Research-Atmospheres*, 107(D21). doi:10.1029/2002jd002251

Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology*, 18(13), 1493-1517. doi:10.1002/(sici)1097-0088(19981115)18:13<1493::aid-joc329>3.0.co;2-t

Qian, B., Gregorich, E. G., Gameda, S., Hopkins, D. W., & Wang, X. L. (2011), Observed soil temperature trends associated with climate change in Canada. *Journal of Geophysical Research-Atmospheres*, 116. doi:10.1029/2010jd015012

Ren, Z. H., Wang, G. A., & Zou, F. L. (2013), Difference between soil temperatures obtained through automatic observation and manual observation and analysis of its causes. *Acta Pedologica Sinica*, 50(4), 657-663

Shi, P. J., Sun, S., Wang, M., Li, N., Wang, J. A., Jin, Y. Y., Gu, X. T., & Yin, W. X. (2014), Climate change regionalization in China (1961-2010). *Science China-Earth Sciences*, 57(11), 2676-2689. doi: 10.1007/s11430-014-4889-1

Szentimrey, T. (1999), Multiple analysis of series for homogenization (MASH). *Proceedings of the second seminar for homogenization of surface climatological data, Budapest, Hungary; WMO, WCDMP-No. 41*, pp. 27-46

Szentimrey, T. (2013), Theoretical questions of daily data homogenization. *Idojaras*, 117(1), 113-122

Szentimrey, T. (2014), Manual of homogenization software MASHv3.03. *Hungarian Meteorological Service*, p.69

Tao, S. Y., Fu, C. B., Zeng, Z. M., Zhang, Q. Y., & P., K. D. (2004), Two long-term instrumental climatic data bases of the People's Republic of China. *Oak Ridge National Laboratory ORNL/CDIAC-47, Oak Ridge(TN)*. doi:10.1175/bams-85-4-587

Wang, X. L., & Feng, Y. (2013), RHtestsV4 user manual. http://etccdi.pacificclimate.org/software.shtml, 29 pp

Wang, X. L., Chen, H. F., Wu, Y. H., Feng, Y., & Pu, Q. (2010), New techniques for the detection and adjustment of shifts in daily precipitation data series. *Journal of Applied Meteorology and Climatology*, 49(12), 2416-2436. doi:10.1175/2010jamc2376.1

Wu, L. Y., & Zhang, J. Y. (2014), Strong subsurface soil temperature feedbacks on summer climate variability over the arid/semi-arid regions of East Asia. *Atmospheric Science Letters*, 15(4), 307-313. doi:10.1002/asl2.504

Xu, W., Sun, C., Zuo, J., Ma, Z., Li, W., & Yang, S. (2019), Homogenization of monthly mround surface temperature in China during 1961-2016 and performances of

GLDAS reanalysis products. *Journal of Climate*, 32(4), 1121-1135 doi:10.1175/jcli-d-18-0275.1

Xue, Y. K., Vasic, R., Janjic, Z., Liu, Y. M., & Chu, P. C. (2012), The impact of spring subsurface soil temperature anomaly in the western U.S. on North American summer precipitation: A case study using regional climate model downscaling. *Journal of Geophysical Research-Atmospheres*, 117, 11. doi:10.1029/2012jd017692

Xue, Y. K., et al. (2018), Spring land surface and subsurface temperature anomalies and subsequent downstream late spring-summer droughts/floods in North America and East Asia. *Journal of Geophysical Research-Atmospheres*, 123(10), 5001-5019. doi:10.1029/2017jd028246

Zeng, X. B., Wang, Z., & Wang, A. H. (2012), Surface skin temperature and the interplay between sensible and ground heat fluxes over arid regions. *Journal of Hydrometeorology*, 13(4), 1359-1370. doi:10.1175/jhm-d-11-0117.1

Zhou, C. L., & Wang, K. C. (2016), Land surface temperature over global deserts: Means, variability, and trends. *Journal of Geophysical Research-Atmospheres*, 121(24), 14344-14357. doi:10.1002/2016jd025410

Zhou, C. L., Wang, K. C., & Ma, Q. (2017), Evaluation of eight current reanalyses in simulating land surface temperature from 1979 to 2003 in China. *Journal of Climate*, 30(18), 7379-7398. doi:10.1175/jcli-d-16-0903.1

zhou, L. T., & Huang, R. H. (2006), Characteristics of interdecadal variability of the difference between surface temperature and surface air temperature in spring in arid and semi-arid region of Northwest China and its impact on summer precipitation in North China. *Climatic and Environmental Research*, 11(1), 1-13. doi:10.3878/j.issn.1006-9585.2006.01.01

Zhou, L. T., & Wen, L. M. (2016), Characteristics of temporal and spatial variations in land-air temperature difference in China and its association with summer rainfall. *Climatic and Environmental Research*, 21(6), 621-632