A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900

Wenhui Xu[1], Qingxiang Li[2*], Phil Jones[3], Xiaolan L Wang[4], Blair Trewin[5]，Su Yang[1], Chen Zhu[1], Panmao Zhai[6], Jinfeng Wang[7], Lucie Vincent[4], Aiguo Dai[8]，Yun Gao[9]，Yihui Ding[10]

1 China Meteorological Administration, National Meteorological Information Center, 46 Zhongguancun South Avenue, Haidian District, Beijing, China 100081

2 School of Atmospheric Sciences, Sun Yat-sen University, 135 Xingangxi Road, Guangzhou,China, 510275

3 Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK;

Also: Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

4 Climate Research Division, Science and Technology Branch, Environment and Climate Change Canada, Toronto, Ontario, Canada.

5 Australian Bureau of Meteorology, Melbourne, VIC, Australia

6 China Meteorological Administration, China Academy of Meteorological Sciences, 46 Zhongguancun South Avenue, Haidian District, Beijing, China 100081

7 LREIS, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China 100101

8 Department of Atmospheric & Environmental Sciences, University at Albany, SUNY, Albany, New York 12222, USA.

9 China Meteorological Administration, 46 Zhongguancun South Avenue, Haidian District, Beijing, China 100081

10 China Meteorological Administration, National Climate Center, 46 Zhongguancun South Avenue, Haidian District, Beijing, China 100081

*Corresponding author:

Dr. & Prof. Qingxiang Li
School of Atmospheric sciences,
Sun Yat-sen University
135 Xingangxi Road,
Guangzhou, China 510275
Tel: 86-756-3688211
E-Mail: liqingx5@mail.sysu.edu.cn; liqx@cma.gov.cn

Abstract

A new dataset of integrated and homogenized monthly surface air temperature over global land for the period since 1900 (China Meteorological Administration global Land Surface Air Temperature (CMA-LSAT)) is developed. In total, 14 sources have been collected and integrated into the newly developed dataset, including three global (CRUTEM4, GHCN, and BEST), three regional and eight national sources. Duplicate stations are identified, and those with the higher priority are chosen or spliced. Then, a consistency test and a climate outlier test are conducted to ensure that each station series is quality controlled. Next, two steps are adopted to assure the homogeneity of the station series: (i) homogenized station series in existing national datasets (by National Meteorological Services) are directly integrated into the dataset without any changes (50% of all stations), and (ii) the inhomogeneities are detected and adjusted for in the remaining data series using a penalized maximal t test (50% of all stations). Based on the dataset, we re-assess the temperature changes in global and regional areas compared with GHCN-V3 and CRUTEM4, as well as the temperature changes during the three periods of 1900–2014, 1979-2014 and 1998-2014. The best estimates of warming trends and there 95% confidence ranges for 1900–2014 are approximately 0.102±0.006 °C/decade for the whole year, and 0.104±0.009, 0.112±0.007, 0.090±0.006, and 0.092±0.007 °C/decade for the DJF (December, January, February), MAM, JJA, and SON seasons, respectively. MAM saw the most significant warming trend in both 1900–2014 and 1979–2014. For an even shorter and more recent period (1998–2014), MAM, JJA and SON show similar warming trends, while DJF shows opposite trends. The results show that the ability of CMA-LAST for describing the global temperature changes is similar with other existing products, while there are some differences when describing regional temperature changes.

Introduction

Land surface air temperature (LSAT) change is a primary measure of global climate change (Hartmann et al., 2013; Jones, 2016). IPCC AR5 has cited four global monthly land surface temperature datasets, including CRUTEM4 from the Climatic Research Unit (CRU) of the University of East Anglia, the Global Historical Climatology Network (GHCN) dataset from NOAA's National Center for Environment Information (NCEI; formerly National Climatic Data Center (NCDC)), GISTEMP from NASA's Goddard Institute of Space Studies, and the Berkeley Earth Surface Temperature (BEST) dataset. Of these datasets, the first version of the GHCN monthly air temperature dataset was developed by NCEI in the early 1990s (Vose et al., 1992). GHCN-V3, containing 7280 stations, was released in 2011, with improved quality control of duplicate data, climate anomalies, and spatial inconsistencies (Durre et al., 2007; Lawrimore et al., 2011). Homogeneity testing and adjustment of the temperature series was conducted by the automatic paired alignment approach (Menne and Williams, 2009). An independent effort in the United Kingdom produced a first release of CRUTEM in the late 1980s. Today, a global dataset of over 6000 stations is still maintained in the fourth iteration of CRUTEM (CRUTEM4, Jones et al., 2012). GISTEMP (Hansen et al., 1999) was developed on the basis of the original data of GHCN-V2, introducing data from several stations in Antarctica and data from the homogenized U. S. Historical Climatology Network (USHCN) consisting of more than 1200 stations. The BEST team combined 16 datasets to build an integrated dataset of global monthly surface air temperature, and a new algorithm has been

developed that can utilize short and/or discontinuous data. After the removal of duplicate records, this dataset incorporates 36,000 stations with various record length (Rohde et al., 2013). These datasets have provided the scientific basis for quantifying and detecting climate change over land. A large dataset, containing about 32,000 stations, has also been released as part of the International Surface Temperatures Initiative (ISTI) (Rennie et al., 2014). Whilst the ISTI dataset itself is not homogenized, the data it contains have been used as the basis for a number of datasets which have performed their own homogenization, including that of Karl et al. (2015) and the next update of the GHCN dataset (GHCNv4, not yet published at the time of writing). The Japan Meteorological Agency (JMA) also maintains a global LSAT dataset at its website (http://ds.data.jma.go.jp/tcc/tcc/products/gwp/temp/ann_wld.html), but no detailed information about the data quality control and homogenization can be found in the scientific literature, and it is only independent of GHCN after 2001. Exploiting reanalyses and using satellite data, as done in Simmonds et al., (2017) and Cowtan and Way (2014) have the potential to refine the conventional approach into the future but are outside the scope of this paper.

Although methodologies for developing all four station datasets differ (for example, CRUTEM4 incorporates homogenized national data sets where available, whereas GHCN and BEST carry out their own homogenization), they all exhibit close agreement with respect to large-scale changes in global and hemispheric LSAT. However, there are differences with respect to regional climate change, especially in

4

South America, Africa, and Asia, and for periods before 1950 (Jones, 2016). Also,

because they differ in terms of data collection, processing techniques and focus, these

datasets exhibit subtle differences when describing global-average LSAT changes,

although they have generally become closer to each other during the 20th century

(IPCC, 2007; Hartmann et al., 2013). All the groups use much the same input data,

but there are still some differences between regions and continents in terms of stations

and data quality. Moreover, they employ different approaches to ensure data quality,

to interpolate and develop gridded products, to construct the global/regional climate

change series, and to calculate / communicate the error assessment. All four datasets

have relatively low station densities (through GTS, detailed discussion in Section 5)

over Asia compared with the United States and Europe, especially for large countries

such as China, Russia, and India (Li et al., 2016; Jones, 2016). The China

Meteorological Administration (CMA) has proposed a plan to develop global baseline

temperature and precipitation datasets to fulfill the needs for regional (especially

Asian) and global climate monitoring and climate change studies (Li, 2013).

In this study, a new dataset of integrated and homogenized monthly surface air

temperature over global land (referred to as the China Meteorological Administration

global Land Surface Air Temperature dataset, CMA-LSAT) is developed and applied

to build a new global homogenized LSAT change series for the period since 1900.

The CMA-LSAT dataset follows a similar philosophy to the CRUTEM4 dataset,

in that it employs homogenized datasets developed at the national or regional level

where possible, with other sources being used only where suitable homogenized

national or regional datasets do not exist. This recognizes that national institutions are likely able to perform homogenization more effectively than is possible in a global dataset, principally due to better access to metadata, and access to more potential reference stations (as most countries have substantial domestic networks whose data are not transmitted internationally). It is considered that the advantages of being able to draw on this national-level information outweighs any disadvantage which may arise from the differences in the actual homogenization techniques used in different parts of the dataset.

The most substantial advance in this dataset, relative to other global datasets cited earlier, is the improved station coverage, especially in Asia. Compared to existing datasets, CMA-LSAT shows significant improvements in data coverage in most countries of Asia, especially in China and its neighboring regions, while its station coverage in Africa and South America is comparable to that of the existing global datasets such as CRUTEM4 and GHCN-V3.

The remainder of this article describes the development of the dataset and provides comparisons with other existing global datasets. Section 2 describes the main data resources and the integration principle. Data quality control and homogenization are highlighted in Sections 3 and 4, respectively. Section 5 summarizes and discusses the characteristics of the merged datasets at the regional scale. Section 6 summarizes the new evaluation of large scale surface air temperature change trends. Discussions and conclusions are presented in Sections 7 and 8, respectively, along with plans for future updates.

1. Data sources and pre-processing

1.1 Data sources and merging hierarchy

Considerable efforts have been devoted to homogenize, create and compare climate datasets over China by scientists from CMA, Chinese universities, and the Chinese Academy of Sciences (Ding and Dai, 1994; Li et al., 2004a; 2009; 2010a; 2015; 2016; Feng et al., 2004; Zhai et al., 2004; Li and Yan, 2009; You et al., 2011; Xu et al., 2013; Cao et al., 2013; Sun et al., 2014; Wang et al., 2014, Yin et al., 2015). However, relatively few of these studies (Wang and Zhou, 2015; Wang et al., 2016) have extended the global or hemispheric domain. In 2015, the National Meteorological Information Center (NMIC) released a station dataset of monthly LSAT over global land (referred to as CMA-LSAT). In this paper, we document this dataset, perform a preliminary analysis, and report results from early comparisons with other data groups.

A total of fourteen sources of LSAT station data has been collected and integrated to develop the CMA-LSAT dataset, including three global sources (CRUTEM4, GHCN-V3, and BEST), three regional sources, and eight national sources. Table 1 summarizes all the sources currently in the CMA-LSAT dataset. Some sources only provide data at the monthly timescale, while others have daily data (which were used to derive monthly data used in CMA-LSAT). Most station records include daily maximum (T_{max}) and minimum (T_{min}) temperatures, but in some cases only daily mean temperature data are available. At the time this project started, the

ISTI database was unavailable to us, thus it was not included in the current version of CMA-LSAT. We plan to include ISTI data in our future versions. Most of the Chinese station data are available only after 1900, and pre-1900 temperature data are also subject to a higher level of uncertainty as many countries had not yet standardized their instrument shelters (Trewin, 2010; Parker, 1994). Because of these, we chose 1900 as the start year of the CMA-LSAT dataset, which is updated continuously in near-real time (see Appendix for details) although data up to 2014 were used in this paper.

A number of different methods have been used by various countries to calculate daily (and hence monthly) mean temperature (T_m). While many countries use the mean of the T_{max} and T_{min}, others use the mean of the values at fixed hours, e.g., the mean of the temperatures at 0h, 06h, 12h and 18h local time. There are systematic differences between mean temperatures calculated by these methods (Trewin, 2004) and hence it is preferable to use a consistent method where possible. However, in some cases, only the mean temperature (not the maximum or minimum) are available and hence it is necessary to use whichever method was used for the calculation of the mean temperature source data. But we calculated T_m whenever the T_{max} and T_{min} are available, even if the T_m is also available.

Table 1 Summary of data sources used in CMA-LSAT

Sources	Priority	Name/Country	Time scale	Raw/QC /Hom*	T_{max}	T_{min}	T_{avg}	Reference
Country		China	monthly	Hom	Y	Y	Y	Xu et al., 2013; Li et al., 2010a
		USA	monthly	Hom	Y	Y	Y	Menne and Williams, 2009

	1	Russia	daily	QC	Y	Y	Y	—
		Canada	monthly	Hom	Y	Y	Y	Vincent et al., 2012
		Australia	daily	Hom	Y	Y	Y	Trewin, 2013
		Korea	daily	Raw	Y	Y	Y	—
		Japan	monthly	QC	Y	Y	Y	—
		Vietnam	daily	Raw	Y	Y	Y	—
		SCAR	Monthly	Raw	N	N	Y	—
Regional	2	ECA&D	monthly	Raw	Y	Y	Y	Klein et al., 2002
	3	HISTALP	monthly	Hom	N	N	Y	Auer et al., 2007
Global	4	GHCN-M v3 (raw)	monthly	QC	Y	Y	Y	Durre et al., 2007
	5	BEST	monthly	QC	Y	Y	Y	Rohde et al., 2013
	6	CRUTEM4	monthly	QC/Hom	N	N	Y	Jones et al., 2012

*QC, quality controlled; Hom, homogenized.

Given the historical nature of data collection, sharing, and rescue, there are many cases where an individual station identifier exists in multiple data sources (potentially duplicate station records). In addition, owing to different collection and reprocessing techniques, the duplicate records do not necessarily have identical temperature values for the same station even though they are based upon the same fundamental measurements (Rennie et al., 2014). Therefore, a hierarchy of all the source datasets must be created before merging. Sources with higher priority take precedence over lower priority sources when more than one record for the same station and same time period exists. The priority order is based on a number of criteria. Due to the emphasis this dataset places on regional climate change, national weather sources (National Meteorological Service, NMSs) and hydrological service sources (National Meteorological and Hydrological Service, NMHSs, country sources) are the most

desirable and are assigned the highest priority. Among them, the American (Menne and Williams, 2009), Japanese, Canadian (Vincent et al., 2012), Australian (Trewin, 2013), and Russian datasets have been developed or published by their national meteorological data centers, whereas the Vietnamese and South Korean datasets are obtained by exchange between countries. For the country sources, different countries incorporate different stations, so they do not include all the same stations, so the eight sources are all given the highest priority of 1（Table 1）. In this case, a higher priority is given to the regional source. ECA&D (Klein et al., 2003; Wijngaard et al., 2003) and HISTALP (Auer et al., 2007) both cover the European region. ECA&D was recently updated when the importance of such provenance was explicitly recognized and given high priority of 2. HISTALP is a multinational dataset of high quality, which has been put together by the Austrian meteorological service; it was given the priority of 3 because it only contained the monthly mean average temperature. The global sources are assigned a lower priority than above sources. GHCN-M is given the priority of 4 for its regular updates with monthly mean maximum and minimum temperatures. BEST is given a higher priority of 5 than CRUTEM4 because CRUTEM4 does not provide monthly means of daily maximum and of daily minimum temperatures (but CRU does incorporate these variables in their CRU TS series of datasets, Harris et al., 2014). These are preferred over monthly mean temperature because they can be directly used to calculate the monthly mean on a globally consistent basis, and because there is compelling evidence that many data artifacts affect daily maxima and minima differently (Williams et al., 2012).

1.2 Data integration

Once all of the sources had been collected and formatted, the data were merged into a single comprehensive dataset. The merge process was based upon metadata matching and data equivalence criteria. The merge process started from the highest priority data source (target) and ran progressively through the other sources (candidates). Each candidate station was compared to all target stations in two steps.

In the first step, each candidate station was run through all the target stations and two metadata criteria were calculated for identifying matching stations. The first metadata criterion considers the likelihood that the same station from two sources has different values for the longitude, latitude, and elevation of the station (e.g., coordinates rounded to one or two decimal places). Therefore, using the latitudes and longitudes, the geographical distance between the candidate station and each of the target stations was computed. If the distance was less than 5 km and the height difference was less than 50 m, the first criterion was met. The second metadata criterion considers the likelihood that the station names also differ, particularly for countries that were once colonial and have subsequently gained independence, or where the phonetic spelling of names may differ between sources. Therefore, the station name similarity was calculated using the Jaccard Index (JI) (Jaccard, 1901; Rennie et al., 2014), which is defined as the intersection divided by the union of two sample sets, A and B:

$$JI = \frac{A \cap B}{A \cup B} \tag{1}$$

If *JI* reaches 0.8, the second criterion is met. If the candidate station meets both of the

1 above metadata criteria, it is considered to match well with the target dataset and data

2 comparison is continued to the second step. Otherwise, the candidate station is

3 determined to be unique and is added to the target dataset as a new station. This

4 process is not perfect (e.g., it is possible that two duplicate records may be added as

5 different stations if the station names used differ substantially enough not to meet the

6 *JI* criterion), but to refine further would require substantial manual intervention and,

7 in some cases, access to locally-held metadata unavailable to the authors.

8 In the second step, a data comparison was made between the candidate station

9 and a target station for certain stations that passed the metadata threshold. These

10 stations were mainly from sources that had not been adjusted or could not be adjusted

11 regularly. For example, Korean or Vietnamese national sources had higher priority,

12 but the final year of data for these sources was only 2007 or 2011. As these data were

13 obtained by exchange between countries, they could not be updated at present.

14 Therefore, a data comparison was performed between the same stations in the Korean

15 source and other lower priority sources such as GHCN, CRUTEM4, or BEST. For

16 reliable data comparison, there was a minimum overlap threshold of 60 months

17 between the two stations (Rennie et al., 2014). If this threshold was met, data

18 comparison was performed using the index of agreement (*IA*) (Willmott, 1981). A

19 modified version of the *IA* (Willmott et al., 1985; Legates and McCabe, 1999), where

20 the squared term was removed, was used:

21
$$IA = 1.0 - \frac{\sum_{i-1}^{n} \left| T_i - C_i \right|}{\sum_{i=1}^{n} \left| C_i - \overline{T} \right| + \left| T_i - \overline{T} \right|} \quad , \tag{2}$$

22 where T_i and C_i are the corresponding monthly values for the target and candidate

1 stations, respectively, and $\overline{T}$ is the mean value for the target station. If the *IA* reaches

2 0.8, the candidate station is merged with the target station, with the lower-priority

3 source being used only where the higher-priority source is unavailable.

4

2. Quality control

6 Despite quality control, the use of various methods leads to different quality

7 problems in integrated datasets. Similar to the process used for GHCN-V3

8 (Lawrimore et al., 2011), we employed a three-step quality control process; Table 2

9 shows the results.

10 Step 1: Check for climate outliers. Monthly anomalies (relative to the

11 1961-1990), higher than five times the standard deviation of the monthly mean of the

12 raw data at each station were considered as outliers, which accounted for 54

13 (0.0007%), 39 (0.0008%), and 129 (0.0026%) station months, respectively, for T_m,

14 T_{max}, and T_{min} for all the station data. These anomalies were treated as missing data.

15 In this study, normals get calculated with at least 10 of 30 years.

16 Step 2: Check for spatial consistency. At a given time, Z_i is considered as an

17 outlier and excluded if

18 $$\left| Z_i - \overline{Z_{ij}} \right| > 3.5\sigma_i \quad , \text{ and} \tag{3}$$

19 $$\left| Z_i - \overline{Z_{ij}} \right| > 2.5 \quad , \tag{4}$$

20 where Z_i is the normalized air temperature anomaly in $^\circ$C at the target station i, Z_{ij}

21 is the normalized air temperature anomaly in $^\circ$C at neighboring station j (not

22 exceeding 20) within 500 km of the target station, $\overline{Z_{ij}}$ is the mean averaged over the

selected neighboring stations, and σ_i is the standard deviation of the normalized air temperature anomalies at all the selected neighboring stations at that time. The test results showed that monthly T_m, T_{max}, and T_{min} have spatial inconsistency problems for 349 (0.004%), 170 (0.003%), and 505 (0.010%) station months among all station data, respectively. These values were treated as missing data.

Step 3: Check for internal consistency. Most data sources contain monthly T_m, T_{max}, and T_{min}. Despite some data sources having been quality controlled or homogenized, internal inconsistencies may arise for some data, such as a T_m being lower than the T_{min} or higher than the T_{max}. The test results showed that internal inconsistency occurred for approximately 1,544 station months, accounting for 0.03% of the full database. Where the internal consistency check was failed, T_m, T_{max} and T_{min} were all treated as missing data.

Table 2 Results for each quality control step (unit: Station month)

Steps	Results of QC (Unit: Station month)		
	T_m	T_{max}	T_{min}
First step (check for outliers)	54 (0.0007%)	39 (0.0008%)	129(0.0026%)
Second step (Spatial consistency check)	349 (0.004%)	170 (0.003%)	505 (0.010%)
Third step (Internal consistency check)	1544 (0.020%)	1544 (0.030%)	1544 (0.031%)

3. Data homogenization

It is important in observational studies that the data used are homogenized, i.e.,

not containing artificial changes due to changes in instruments, sampling time, or station location (Jones et al., 1985; Peterson and Vose, 1998; Li and Dong, 2009; Dai et al., 2011;　Trewin, 2013; Wang et al., 2014; Vincent et al., 2015). The following procedure was performed to ensure homogeneity in our dataset. This section describes the procedures used to homogenize our dataset.

As described in Section 2, a higher priority was given to national or regional data sources because we believed that individual nations or regions are most authoritative in terms of their own climate data. Then, for data sources with homogenization already applied, such as USHCN, China, Canada, and Australia, and HISTALP , or those parts of CRUTEM4 for which the data has been homogenized by NMSs/NMHSs, or considered as homogeneous series by CRU, we adopted the data directly without additional homogenization. We realize that these various sources applied different homogeneity adjustment methods. The quantile-matching adjustment method (Wang et al., 2010; Wang and Feng 2010) was applied to produce the Chinese (Xu et al., 2013) and Canadian (Vincent et al., 2013) homogenized datasets of daily temperatures, and the percentile-matching method was used to produce the Australian homogenized dataset of daily temperatures (Trewin, 2013), whereas mean adjustment was applied to produce other homogenized datasets. However, we believe that these are the best homogenized datasets; and it is impossible for us to do better because we don't have the metadata, the expert knowledge about the data, and other data that were used to produce these national homogenized datasets (for example, concurrent hourly data were used to adjust for effects of the change in the definition of the

climatological day before statistical methods were used to produce the Canadian

homogenized dataset). We also believe that the differences induced by using different

homogenization methods are trivial in comparison with the differences in observing

practices, instrumentation, and post-observation processing used by different

nations/countries.

As has been described (e.g., Houghton et al., 2001; Jones and Moberg, 2003),

adjusting the central tendency or mean state of a climate variable is usually sufficient

to homogenize monthly time series and provide reliable estimates of trends and

variability for temperature. Therefore, the following homogenization process and

mean adjustment were applied to the remaining station series from other data sources,

such as GHCN-V3, BEST, Russia, Japan, Korea, Vietnam, and SCAR. In all, there

were 4143 station series needed to be homogenized for T_{max} and 3732 for T_{min}. T_m

was obtained from the average of T_{max} and $T_{min.}$ However, at another 1682 stations

where T_{max} and T_{min} were unavailable, T_m values were directly assessed for

homogenization.

3.1 Reference construction

It is very important to derive a homogeneous reference series that well represents

the same climatic variations as those in the candidate series. However, it is often

difficult to find a homogenous representative reference data series. We put all the

stations, including homogenized data and raw data, in the data pool for the

construction of reference series. There are three steps to derive the reference series.

In the first step, a reference series should be chosen within a representative

distance (Xu et al., 2013). For this purpose, we divided the world into seven

continents to determine different spatial representative distances, in a similar way to

Li et al. (2010a) (Table 3). The regional spatial representative distances were obtained

by the following steps: 1) calculate the correlation of the first difference series

between any two stations within 1000 km of each other; and 2) then fit a curve

through it to determine the distance at which the correlation goes below 0.8. From

Table 3, the representative distances of the regions were all within 450 km; that is,

reference stations within this distance should be chosen. The exception was found in

Antarctica; although the representative distance (derived from a small sample) of the

region is 220 km, there were very few stations within this distance, so it was

impossible to homogenize most Antarctic stations.

In the second step, the method of Peterson and Easterling (1994) was used to

establish the reference time series (R0). Peterson and Easterling (1994) showed that

the reference series can be constructed by spatially averaging the neighboring first

difference series, weighted by the inverse square of distance, from at least three

nearby stations that are highly positively correlated with the candidate series. The

averaged difference series was analyzed to remove any abnormal data points, after

which it was converted back to data series for use as the reference series. The

advantage of this approach is that it reduces the impact from individual jumps on the

re-converted reference series, which is affected by inconsistent lengths of nearby

series and some anomalous values or undetected inhomogeneities in individual

sequences.

1   In the third step, the inhomogeneity and representativeness of R0 is tested to

2   ensure its suitability. The R0 series is first tested by visual checks or by using a

3   penalized maximal F (PMF) test (Wang, 2008a and 2008b). If R0 is obviously

4   inhomogeneous, it is reconstructed by adjusting the potential reference stations to

5   make a second homogeneous reference series R1. Then, the correlation between the

6   first difference series of R1 and the target series is calculated. If this correlation

7   reaches 0.8, the R1 series is used as the reference series. In our analysis, there were

8   about 20 and 25 stations (0.48% and 0.67%) of Tmax and Tmin respectively which

9   were excluded because of no suitable reference series available to allow the

10   construction of an R0 or R1 series meeting the correlation criterion.

11   Table 3 Average distance between any two stations within 1000 km of each other for which

12   the correlation of available annual temperature data during 1900-2014 reached 0.8 across the

13   seven regions

Region	North America	Europe	Asia	South America	Africa	Australia	Antarctica
Average distance	389	342	392	421	429	396	220

14   **3.2 Methodology for discontinuity detection and adjustment**

15   There are many studies on benchmarking of the data homogenization methods;

16   many methods are found to have exhibited good performance in different aspects

17   (Kuglitsch et al., 2012; Venema et al., 2012; Domonkos et al., 2013) with rankings of

18   specific methods often dependent on the metric(s) used for assessment. In this study,

19   we used the RHtestsV4 software package (Wang and Feng, 2010) to homogenize the

20   monthly temperature series. This package includes two algorithms for detecting

unknown changepoints: the PMTred algorithm (Wang 2008a), which is based on the

penalized maximal t test (Wang et al., 2007) and requires a reference series; and the

PMFred algorithm (Wang 2008a), which is based on the penalized maximal F (PMF)

test (Wang, 2008b) and can be used without a reference series. The RHtestsV4

package and its previous versions have been widely used to homogenize climate data

(e.g., Zhang et al., 2005; Alexander et al., 2006; Wan et al., 2010; Dai et al., 2011;

You et al., 2011; Kuglitsch et al., 2012; Vincent et al., 2012; Wang et al., 2013; Xu et

al., 2013).

The potential change points from the RHtestsV4 were then synthesized and

identified using available metadata and three other criteria. The three criteria were the

timescale consistency, spatial consistency, and elements consistency. For the

timescale consistency, the monthly scale breakpoints were compared with those on an

annual scale. For the spatial consistency, the climate trends in the base series were

compared with nearby stations to determine the breakpoints in the candidate series.

For the elements consistency, the adjusted values as well as the sensitivity of the three

(T_m, T_{min}, T_{max}) time series to artificial changes were used as the criteria to determine

the breakpoints. In general, metadata was the most direct and solid evidence; that is, if

a change point is supported by metadata, it will be retained for adjustment. We

consider a detected change point to be documented when metadata indicate a

documented change within one year before or after the detected change point.

Unfortunately, at present, detailed metadata (for those stations not derived from

national datasets) is unavailable outside the mainland of China. For other criteria,

1 change points supported by at least two pieces of evidence will be retained for
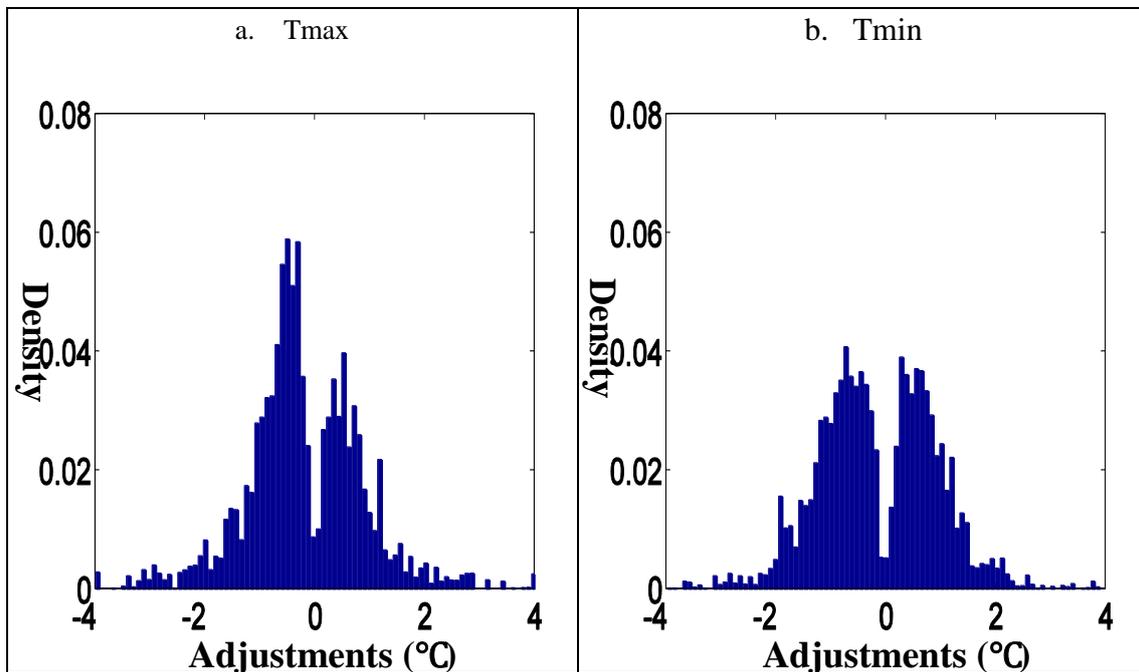
2 adjustment.

3.3 Statistics of the adjustment

4 As described above, our homogenization was not applied to the data sources that

5 have already been homogenized. In addition, stations with a length shorter than 20

6 years were included in the station dataset after being compared/tested either visually

7 or by statistics. In total, there were 4143 stations that remain to be homogenized for

8 T_{max} and 3732 for T_{min}. Table 4 lists the number of stations without any shift or with

9 shifts that have been identified and adjusted for T_{max} and T_{min}. At the 5% significance

10 level, the T_{max} and T_{min} series were found to be homogeneous at 2935 and 2456

11 stations (71% and 66%), respectively. A total of 1447 change points were detected in

12 1208 T_{max} time series and 1750 change points were detected in 1276 T_{min} time series;

13 for these stations, T_m was calculated by averaging the homogenized T_{max} and T_{min}. In

14 addition, other directly collected raw T_m series for 1682 stations were also assessed

15 for homogeneity; they were found to be homogeneous at 1092 (65%) stations, and a

16 total of 736 change points were detected in 590 T_m time series.

17 Table 4. Statistics of breaks for the T_{max}, T_{min}, and T_m temperature series during 1900-2014.

Breaks	T_{max}	T_{min}	T_m
none	2935	2456	1092
one	1002	982	462
two	184	242	113
three	14	33	12
four or more	8	19	3
total Breaks	1447	1750	736

|			
total stations	4143	3732	1682

1

2

3 Figure 1 Probability density of all mean-adjustments applied to the T_{max} and T_{min} temperature
4 series.

5   Figure 1 shows the probability density of all mean-adjustments applied to the

6 monthly T_{max} and T_{min} identified as having inhomogeneities. The vast majority of the

7 adjustments are between −1 °C and 1 °C for both T_{max} and T_{min}, with more negative

8 than positive adjustments for T_{max}. The extreme adjustments mostly range from −3 °C

9 to 3 °C for T_{max} and from −2 °C to 2 °C for T_{min}. The mean of the adjustments is

10 −0.2322 °C for T_{max} and −0.1386 °C for T_{min}. The bimodal distribution of the results

11 is similar to those discussed in Brohan et al. (2006) and Lawrimore et al (2011).

12   Data inhomogeneities also decrease the spatial consistency of estimated trends in

13 annual mean temperatures; this could be improved with homogenization (Li et al.,

14 2004a; Dai et al., 2011; Vincent et al., 2012; Trewin, 2013; Xu et al., 2013). Figure 2

1 shows the distribution of the original and adjusted T_{max} and T_{min} trends for those

2 stations that have been detected as inhomogeneous, with the trends being calculated

3 over the available record during 1900-2014. The temperature trends derived from the

4 homogenized data series have improved spatial consistency compared to those

5 derived from the original data series. The improvement is particularly noticeable in

6 regions with large warming or cooling trends in North America, South America,

7 Africa, Europe, and Asia for both T_{max} (upper panel) and T_{min} (lower panel). It is

8 worth noting that we have not made any adjustments to the Antarctic data, so there is

9 no change for this area due to a lack of available neighbors for using the approach

10 discussed in section 4.1.

a. Trend difference from raw T_{max}

b. Trend difference from homogenized T_{max}
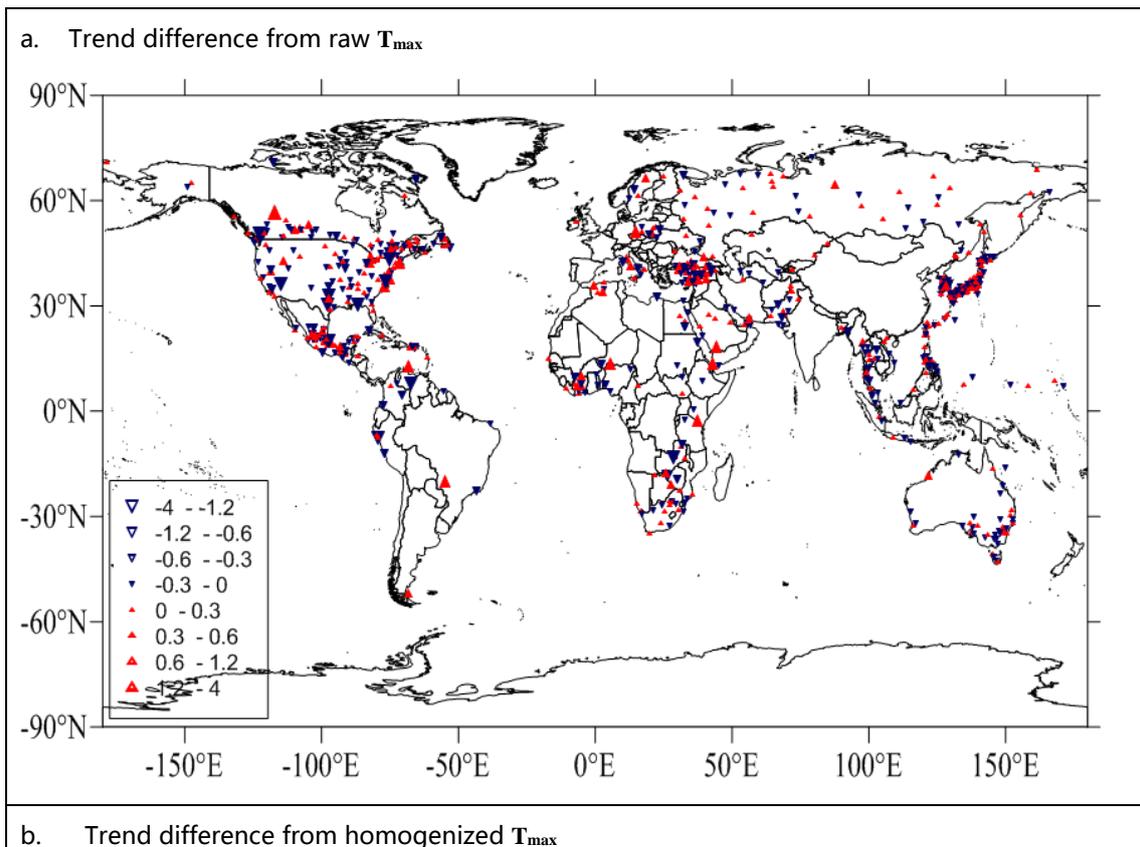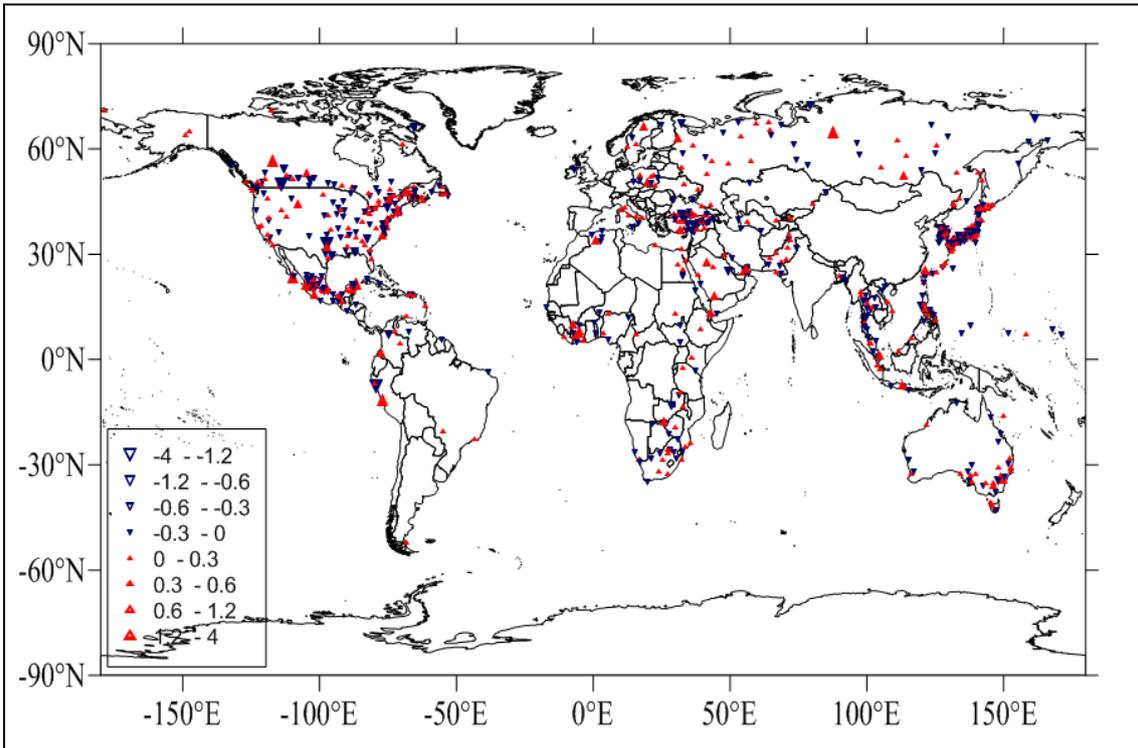
c.   Trend difference from raw T$_{min}$
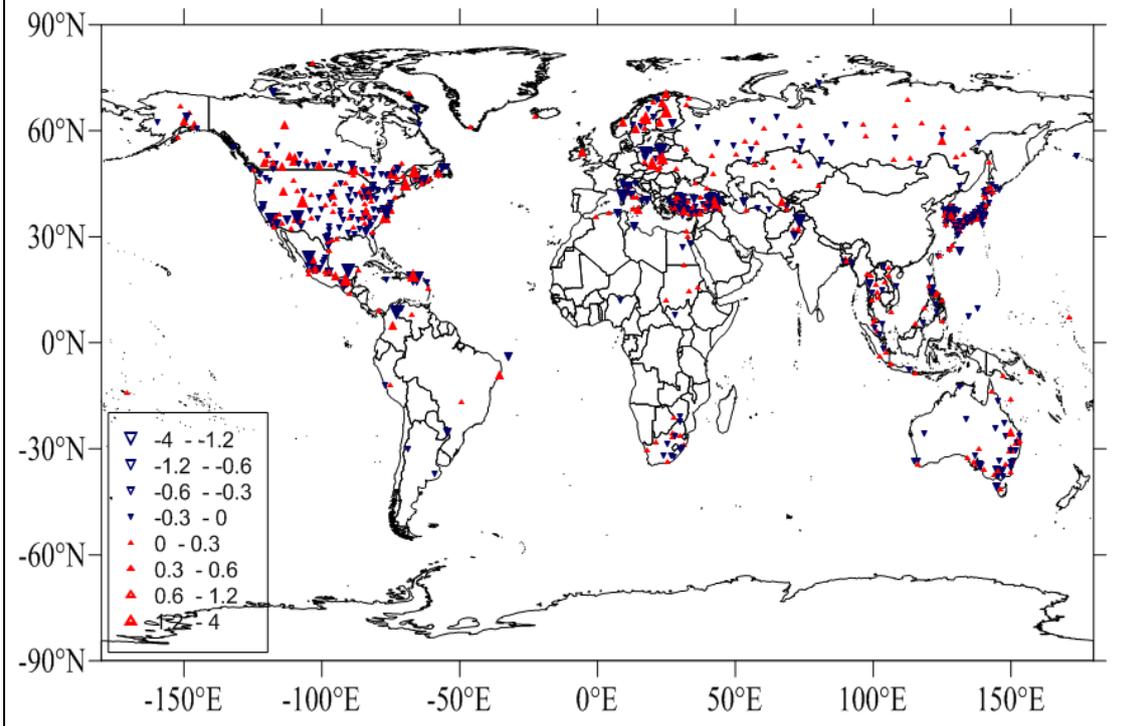
d.   Trend difference from homogenized T$_{min}$
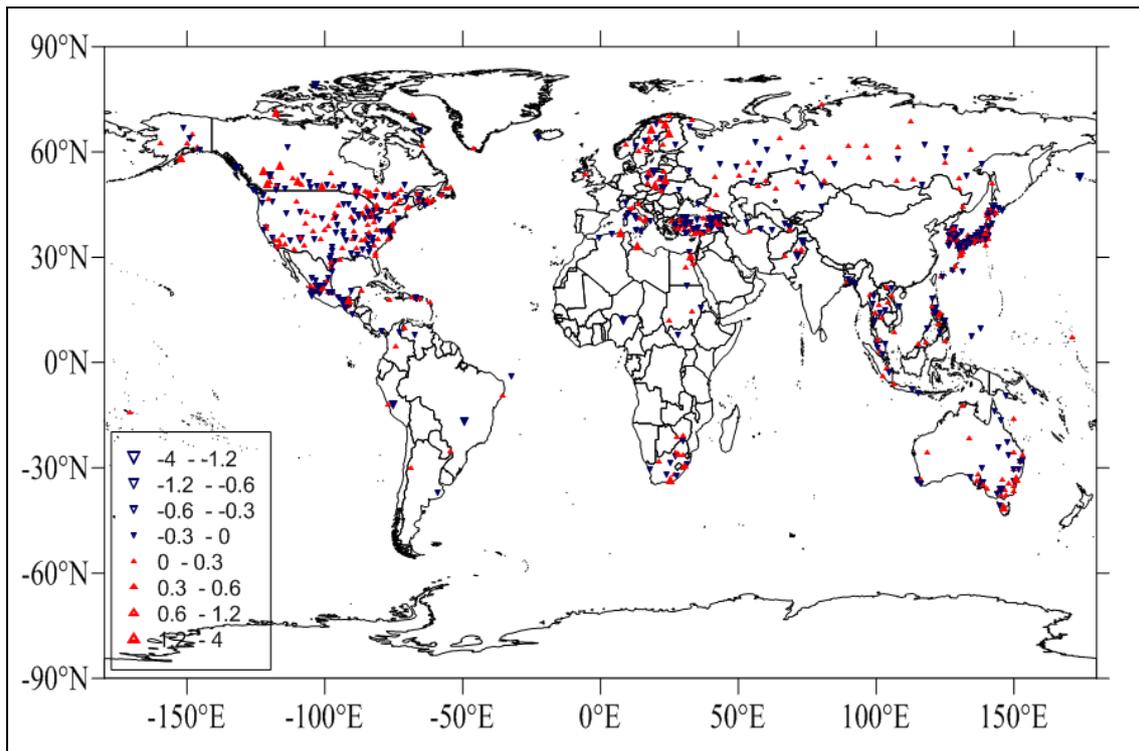
Figure 2 Trend difference between each station and its nearest neighbor station in time series of annual means of the raw and homogenized monthly means of daily maximum (T_{max}) and of daily minimum (T_{min}) temperatures (homogeneous stations are not shown; trends are calculated over whatever length of record is available). The triangle size is proportional to the magnitude of trend difference. Units: °C/decade.

3.4 Urbanization

It is well known (e.g., Oke, 2004) that urban sites are generally warmer than their rural surroundings. The nature of the urban-rural temperature difference has been widely documented in the literature, with much of the focus on the largest differences between individual sites at sub-daily timescales.

For studies of the potential urban influence on global-scale temperature trends, our interest is not in extreme differences, or in assessments of the urban heat island (UHI) for individual locations, but rather in the way that changes in urban influences over time may affect long-term trends in large area averages based on many station records (Peterson, 2003; Parker, 2004, 2006). Peterson and Owens (2005) also note

that, in the context of a large data set, any UHI influence operates in conjunction with other influences on local-scale climate observations, including elevation differences, distances from major water bodies, observation time changes and weather types.

Existing global data sets take a range of approaches to urbanization influences. The GISTEMP data set currently includes an urbanization adjustment, based on a comparison of trends at urban stations with those at nearby rural stations (Hansen et al., 2010). Other global data sets do not include an explicit urbanization adjustment, in many cases, urban influences on temperatures at a station will be manifested as a step change (e.g. when a new building is constructed near an observing site) and will be adjusted for as part of the general homogenization process.

At the global scale, urbanization impacts on estimates of global mean land surface temperatures have been found to be negligible. The urbanization adjustments applied to GISTEMP only influence global mean LSAT by a maximum of 0.01 °C (Hansen et al., 2010), whilst Berkeley Earth (Wickham et al., 2013) also found a minimal influence, reinforcing similar conclusions in AR4 of IPCC (IPCC, 2007).

Urbanization impacts on temperature can be regionally important, especially in areas such as eastern China where rapid urban growth is occurring. This has been assessed by a number of authors (e.g., Li et al. 2004 and 2010b, Zhou et al. 2004, Jones et al. 2008, Yan et al. 2009, Ren et al. 2005, Wang et al. 2015) with varying conclusions, but all find that urbanization has partially contributed to observed warming over China, with estimates ranging from 5% to 40% of the total observed warming. Hansen et al. (2010) also found an enhanced urbanization influence in the

southwest United States, another area which has experienced rapid urbanization in recent decades, with an influence of up to 0.1 °C on area averages.

Considering the evidence that urbanization has a minimal effect on LSAT at global and continental scales, the CMA-LSAT data set does not include any explicit urbanization adjustments. However, regional analyses based on the data set should take possible urbanization influences at the local scale into account in heavily urbanized areas.
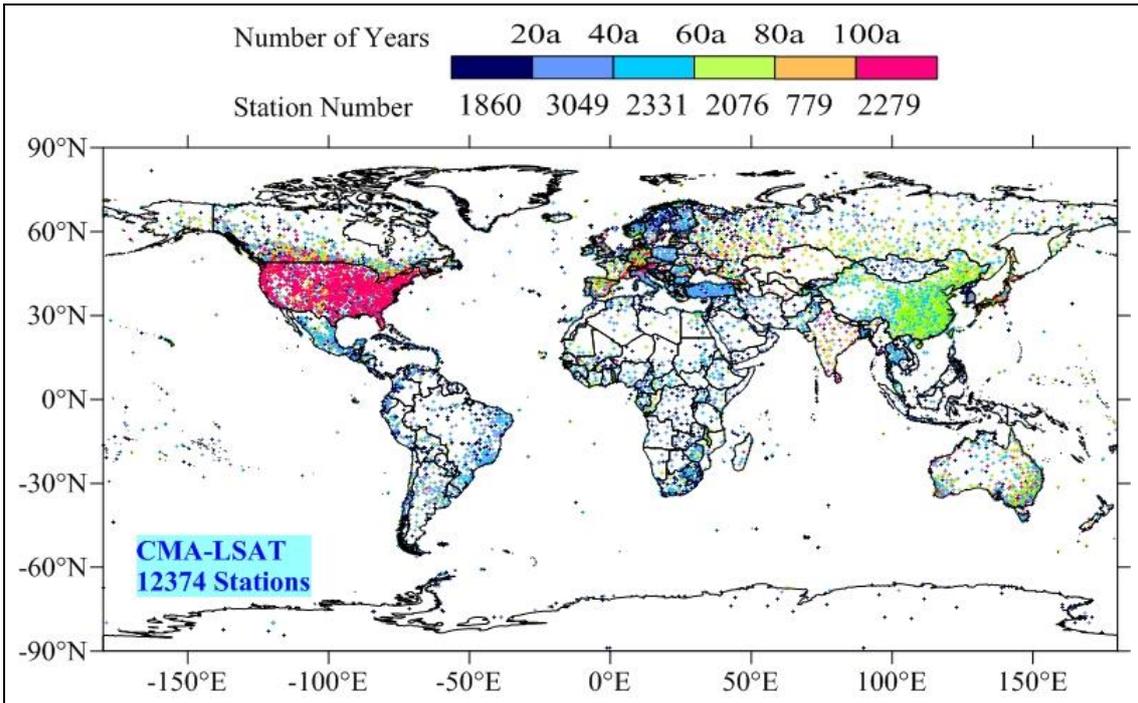
4. Data assessment and discussion

4.1 Station number and spatial-temporal distribution

Using the source hierarchy (Table 1) and data integration (Section 2.2), approximately 12,374 unique stations with monthly T_m series, 8273 stations with T_{max} series, and 7655 stations with T_{min} series were identified. Figure 3 shows the spatial distribution of the stations with monthly Tm data during 1900-2014 and their record length from CMA-LSAT and other datasets. Although the station density in CMA-LSAT is still lower than ISTI (Fig.3a, b), it is higher than that in GHCN-V3 and CRUTEM4 (Fig. 3c, d), particularly over Europe and Asia, where the length of data series increases the most in CMA-LSAT (Fig. 3e, f).

a. CMA-LSAT

Number of Years
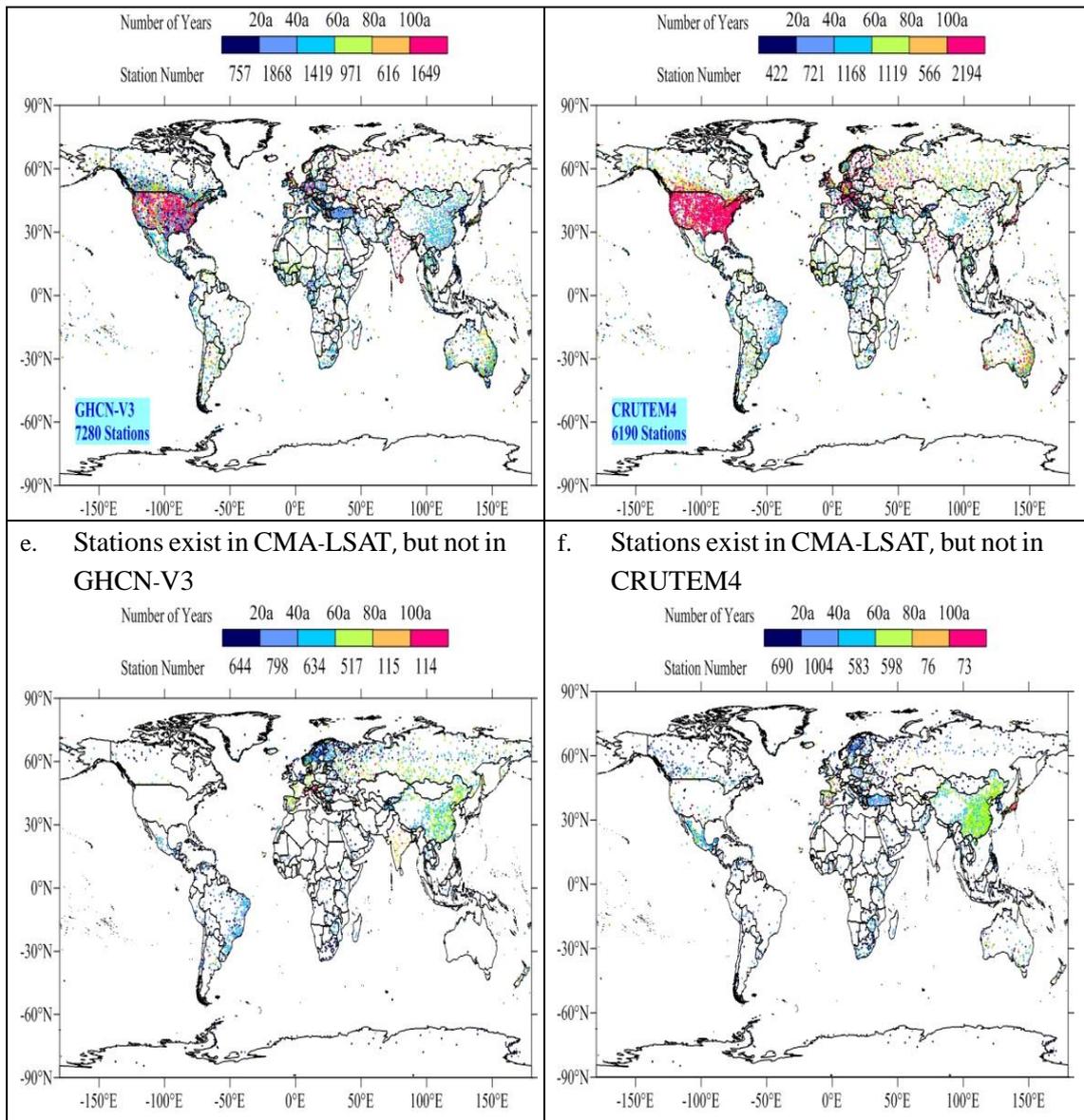
	20a	40a	60a	80a	100a
Station Number	1860	3049	2331	2076	779	2279

CMA-LSAT
12374 Stations

b. ISTI

Number of Years

	20a	40a	60a	80a	100a
Station Number	10789	7627	5288	3479	4853	24

ISTI
32060 Stations
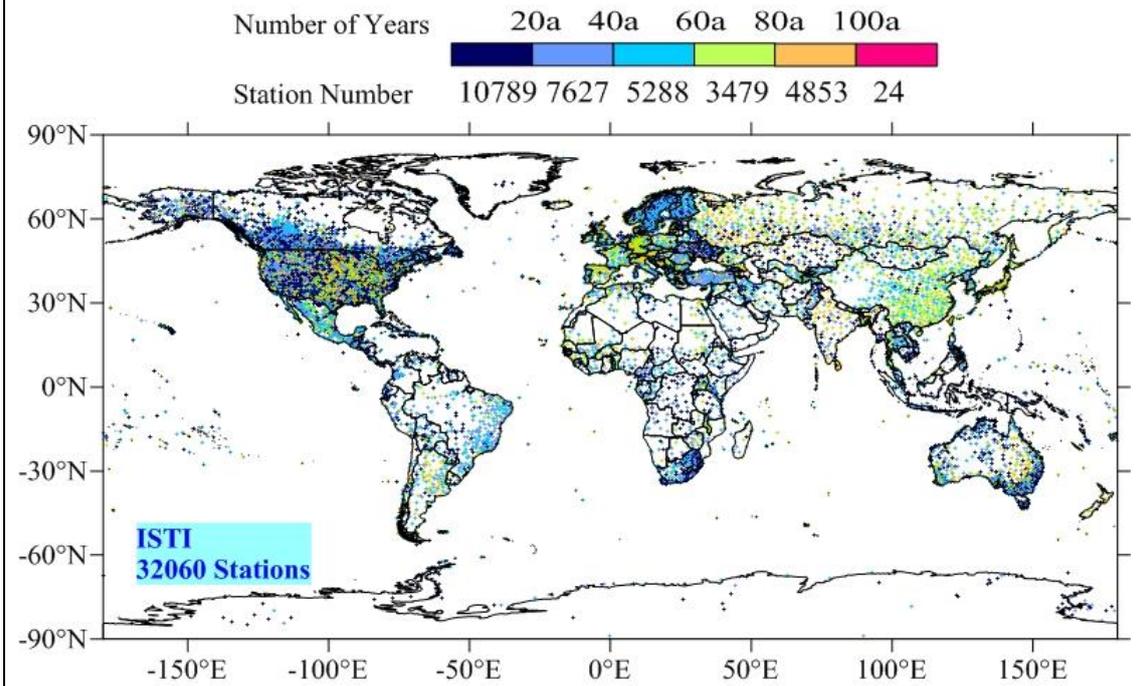
c. GHCN-V3	d. CRUTEM4

Figure 3 Spatial distribution of the stations with monthly mean temperature T_m data, and their length from the CMA-LSAT and other datasets during 1900-2014 (unit: year)

A comparison of station counts is given in Figure 4. Since 1900, there are consistently more stations in CMA-LSAT than in GHCN-V3 or CRUTEM4. Moreover, there is a significant drop in the number of stations in 1990, but this is ameliorated by many of the new sources, illustrating that this station drop-off is simply due to access, not due to widespread station closures. (The ISTI dataset also shows substantial improvement over GHCN-V3 and CRUTEM4 in recent years). It is

worth noting that for Europe, CMA-LSAT has fewer long station series than CRUTEM4, because we use ECA&D and GHCN-V3 as the main sources and the longer series in CRUTEM4 were not given preference due to their not having T_{max} and T_{min} data. This could become an important issue if CMA-LSAT is to be extended before 1900.
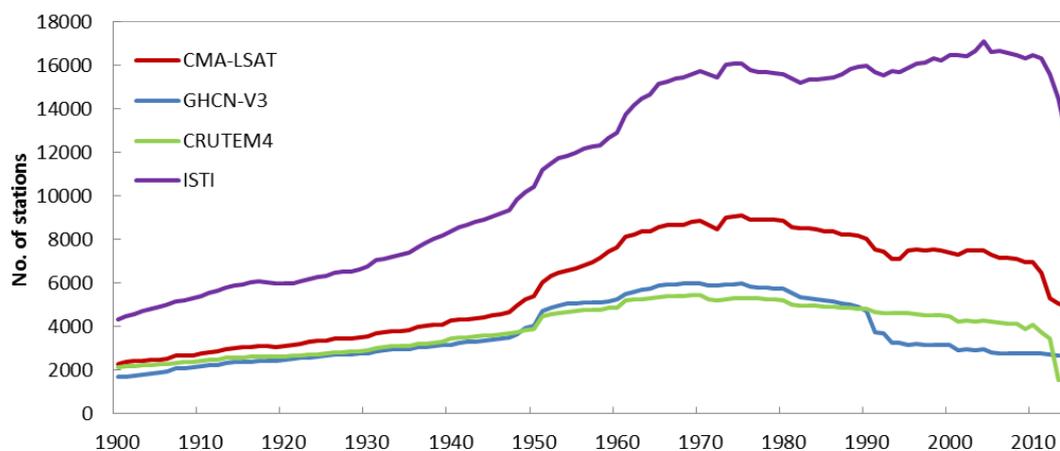

Figure 4 Station count comparison from the CMA-LSAT and other datasets

A comparison of grid boxes is given in Figure 5. Coverage is defined by one or more stations within each $5° \times 5°$ grid box. There is an increase in the global station coverage for all time periods comparing with CRUTEM4 and GHCN-V3, especially for 1990–2010s (the coverage of CMA-LSAT is only slightly lower than that of ISTI, which shows both of them have comparable data resources for near real-time updating). Compared with GHCN-V3, CMA-LSAT's coverage in the Northern Hemisphere (NH) is higher by about 10–15% from 1900 to 1950, 20% during from the 1960s to the 1980s, and 30–40% during the 1990s and 2000s (Figure 5b). The

1 improvement for The Southern Hemisphere (SH) coverage varies, for the most part,

2 between a 15% and 25% increase, with larger increases of approximately 50% over

3 the past 20 years (Figure 5c).

4
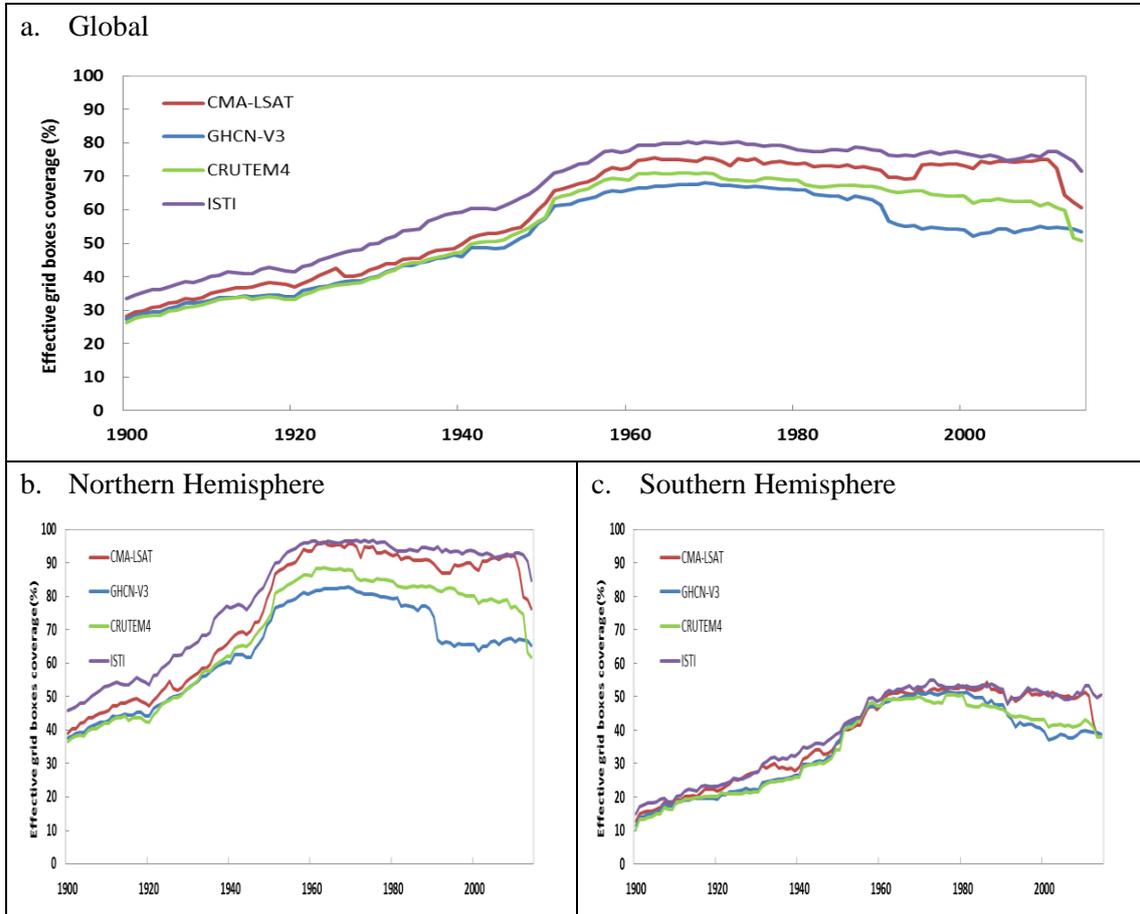
5 Figure 5 Grid box numbers for the CMA-LSAT compared to GHCN-V3 CRUTEM4 and ISTI:
6 (upper) global, (left) Northern Hemisphere, and (right) Southern Hemisphere

7 **4.2 Characteristics at the continental scale**

8   As described in Section 2.1, we spent considerable efforts to obtain a higher

9 density of stations over Asia, for example, by exchanging data with Korea, Vietnam,

10 and Japan. As shown by Fig. 6a, the number of Asian stations with different record

11 lengths in CMA-LSAT is significantly higher than that in other datasets: 202 stations

12 in CMA-LSAT have record lengths of 120–150 years, compared with 93, 57, and 117

13 stations in GHCN-V3, CRUTEM3, and CRUTEM4, respectively. Figure 6b shows

1   that the number of Asian stations in CMA-LSAT is significantly higher than those in

2   GHCN-V3 and CRUTEM4 for the entire analysis period from 1900-2014, but less

3   than those for short (less than 20 years) and long (longer than 80 years) series in ISTI.

4   The number of Asian stations after the 1990s is significantly higher in CMA-LSAT

5   than in the other two datasets. Similar statistical results are found in both Africa and

6   South America, where the station densities are relatively low (Figure 6c-f). Not

7   surprisingly, the station numbers in CMA-LSAT are fewer than those in ISTI in

8   Africa and South America. From this point of view, ISTI shows strong potential as a

9   new data source for future upgrades of CMA_LSAT.
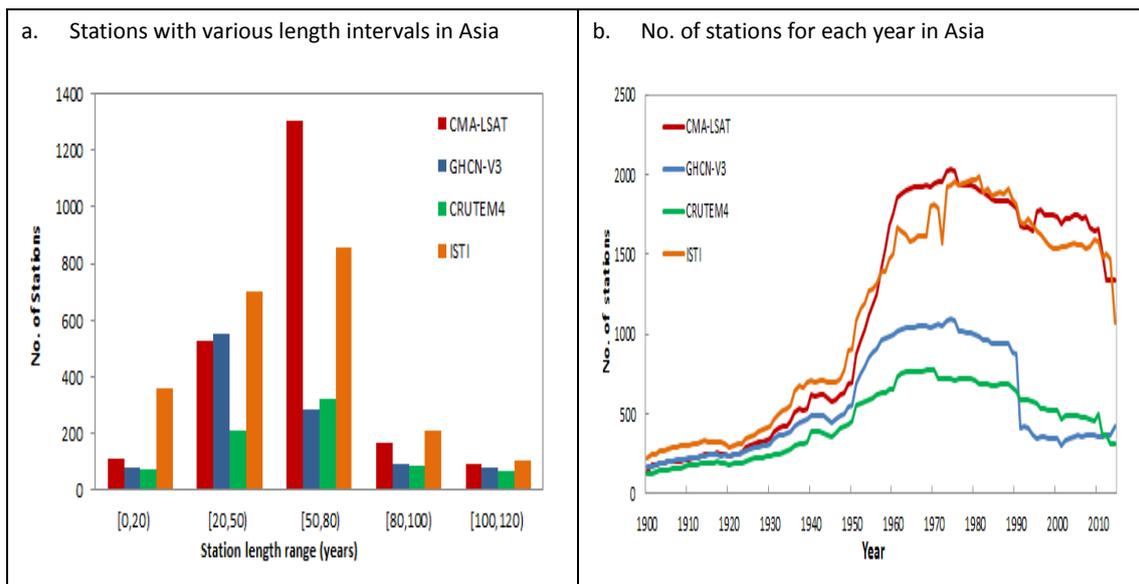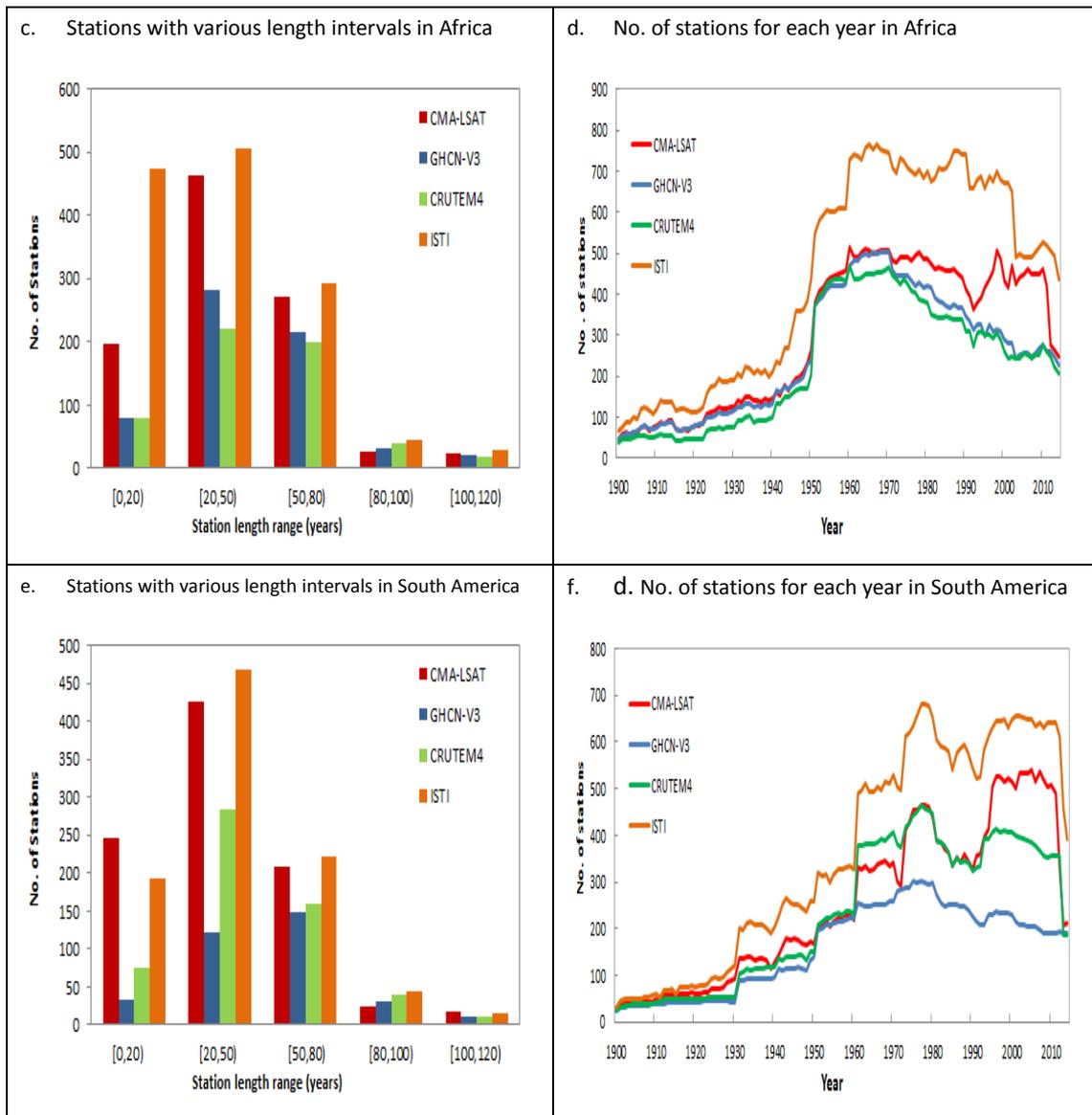
a. Stations with various length intervals in Asia

b. No. of stations for each year in Asia

3   The area averages are compared with two other analyses．For this purpose,

4   monthly anomalies were calculated relative to the reference period of 1961−1990, and

5   only those stations with annual mean values available for at least 10 years during

6   1961−1990 were used. Comparing with CRUTEM4 use of stations with lengths of at

7   least 15 years data, stations with lengths of at least 10 years were chosen in

8   CMA-LSAT to expand the coverage of normal and use more stations. This resulted in

9   a total number of 9765 stations (8300 in the NH and 1465 in the SH). Following Jones

1 (1994), gridding of the temperature anomalies was made by averaging all station

2 anomaly values within each 5° latitude × 5° longitude grid. The use of a base period

3 in this type of study appears to eliminate a large number of series, but almost all the

4 series that cannot produce the 1961-90 average are mostly short and recent series (the

5 average length of these short stations is 18 years, comparing with 61 years when those

6 longer ones were used). Regional mean LSAT anomaly time series were constructed

7 based on the method of Jones (1994) and Jones et al. (1999) by averaging with

8 area-weights, using the cosine of the central latitude of each grid box as the weight

9 coefficient. Eight regions are defined, following Jones and Moberg (2003), for the

10 seven continents of the world (Asia, Africa, South America, Europe, North America,

11 Australia, and Antarctic) plus the Arctic. Here, the regional series of the eight

12 continents/regions are first chosen for comparison (Figure 7a–h). The following

13 global or hemispheric mean LSAT anomaly time series are established using the same

14 method.

15 Figure 7a shows the Asian SAT anomaly time series during 1900–2014.

16 Compared with the other two analyses, the CMA-LSAT Asia series is slightly cooler

17 during the 1920s–1930s and slightly warmer during recent decades. However, in

18 general, the SAT change trends are very similar (0.120, 0.114, and 0.118 °C/decade

19 for CMA-LAST, CRUTEM4, and GHCN, respectively) (Table 5), except for constant,

20 very slight underestimates in recent years for CRUTEM4 and GHCN.

a. Asia: 5-60°N，60-180°E	b. Africa: 35°N-40°S，20°W-45°E
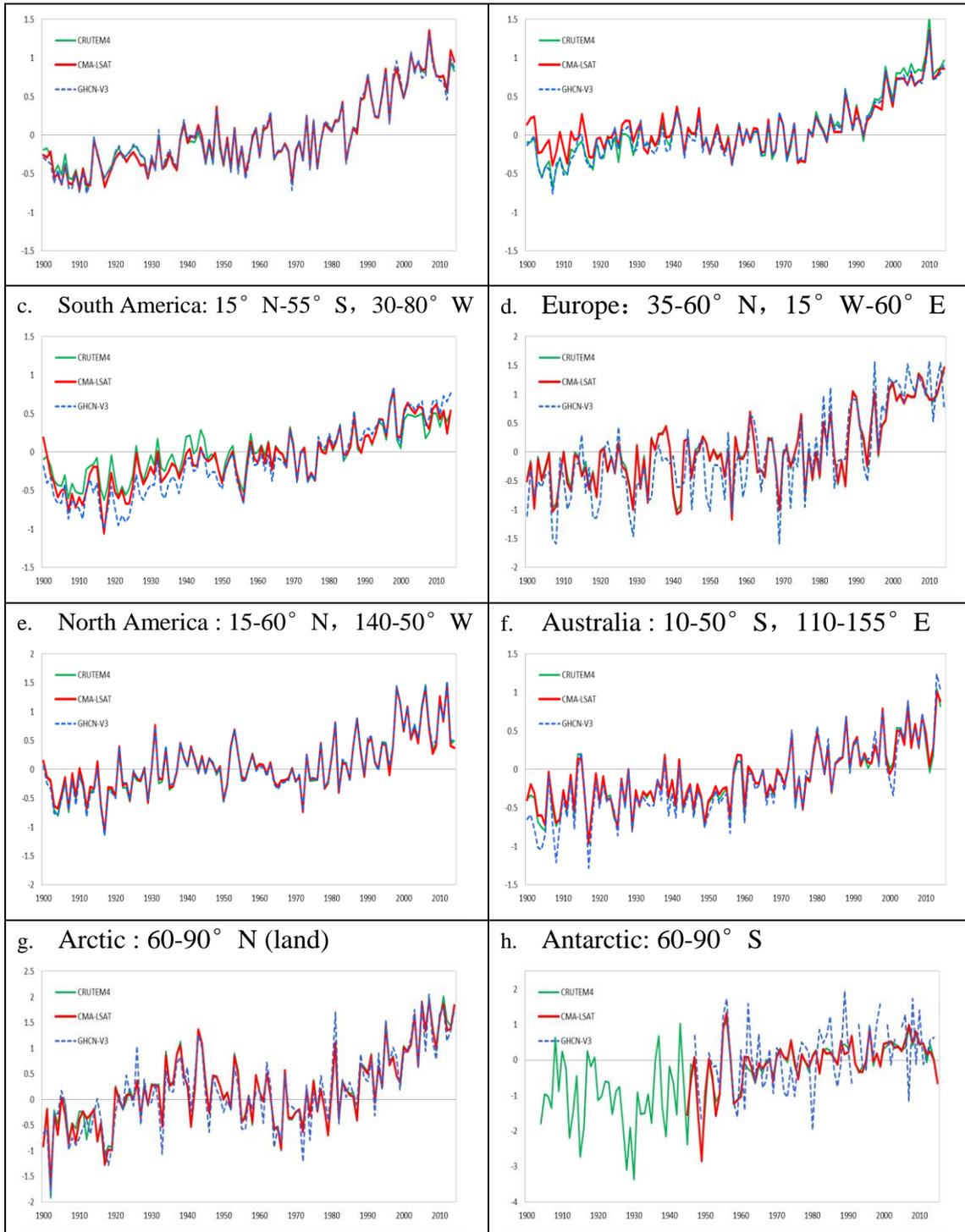
1 Figure 7. Annual land surface air temperature anomalies for different regions (Asia, Africa,
2 South America, Europe, North America, Australia, Arctic, and Antarctic (a)-(h), respectively)
3 during 1900–2014 (relative to the 1961–1990 mean).

4 Table 5 Linear trends of annual mean LSAT for different regions during different periods. All
5 trends are significant at the 5% level. Unit: °C/decade

				1900–2014			1951–2014
Asia	Africa	South America	Europe	North America	Australia	Arctic	Antarctic

CMA-LAST	0.1203	0.0681	0.0952	0.1284	0.0814	0.0835	0.1318	0.1281
CRUTEM4	0.1143	0.0992	0.0740	0.1221	0.0874	0.0874	0.1309	0.1404
GHCN-V3	0.1183	0.0913	0.1241	0.1577	0.0944	0.1054	0.1376	0.1500

1   Figures 7b–h show the SAT anomaly time series for the other seven regions

2 during 1900–2014. For Africa (Figure 8b, Table 5), the greatest differences between

3 the analyses occur at the beginning of the records. The CMA-LSAT series is slightly

4 warmer during 1900–1950 (about 0.2 °C warmer during 1900–1920s), but is slightly

5 cooler since the 1990s. For South America (Figure 7c, Table 5), there is a greater

6 divergence between the series, except in the climate normal period (1961–1990). The

7 CMA-LSAT series is more consistent with CRUTEM4 than GHCN-V3, which is

8 mainly due to the sharp drop in the number stations after 1990 in GHCN-V3. For

9 Europe (Figure 7d, Table 5), CMA-LSAT used the same homogeneity dataset as the

10 main data source as CRUTEM4, so CMA-LSAT is more consistent with CRUTEM4.

11 The notable differences from the GHCN-V3 arise from the fact that GHCN-V3 has

12 fewer stations in Europe than HISTALP and CRUTEM4. For Australia (Figure 7e,

13 Table 5), there are slight differences between the other two datasets, which is likely

14 due to the homogenization of the data by different groups [e.g., CMA-LSAT uses

15 homogenized ACORN data (Trewin, 2013) from the Australia Bureau of Meteorology,

16 as does CRUTEM4]. For North America (Figure 7f, Table 5), CMA-LSAT is

17 consistent with the other two datasets (except at the beginning and end of the records)

18 because all three datasets use the USHCN homogenized dataset as the main data

19 source and CMA-LSAT uses the newly developed, second generation of

20 homogenized data for Canada (Vincent et al., 2013). It is worth pointing out that these

high latitude areas tend to show a faster temperature increase, however, the

meteorological observations in high latitude areas tend to be shorter, and the

shorter-term trends for the SAT in higher latitude tend to be larger. If these short-term

data series have been added to the whole dataset directly, this could cause some of the

warming bias in global/hemisphere SAT trends. Thus CMA-LSAT does not contain

that many more stations in high latitude regions like Arctic and Antarctic. For the

Arctic (Figure 7g), the CMA-LSAT series is more consistent with CRUTEM4 than

GHCN-V3 during most of 1900–2014. For the Antarctic (Figure 7h, Table 6),

CMA-LSAT is consistent with CRUTEM4 after 1960. There are many fewer stations

in the Antarctic, and therefore larger variances, before 1960; thus, as for GHCN, we

only retain the series after the 1940s. CRUTEM4 before the mid-1940s is based upon

a single station (Orcadas).

In summary, although CMA-LSAT shows some differences for the

continental/regional SAT series compared with GHCN and CRUTEM, the three

sources still reflect good consistency. With detailed metadata and a full understanding

of regional climate changes, the direct use of homogenized data produced by domestic

meteorological data centers would be most likely to improve the accuracy of

regional/national climate change detection.

5. Comparisons of large-scale surface air temperature changes

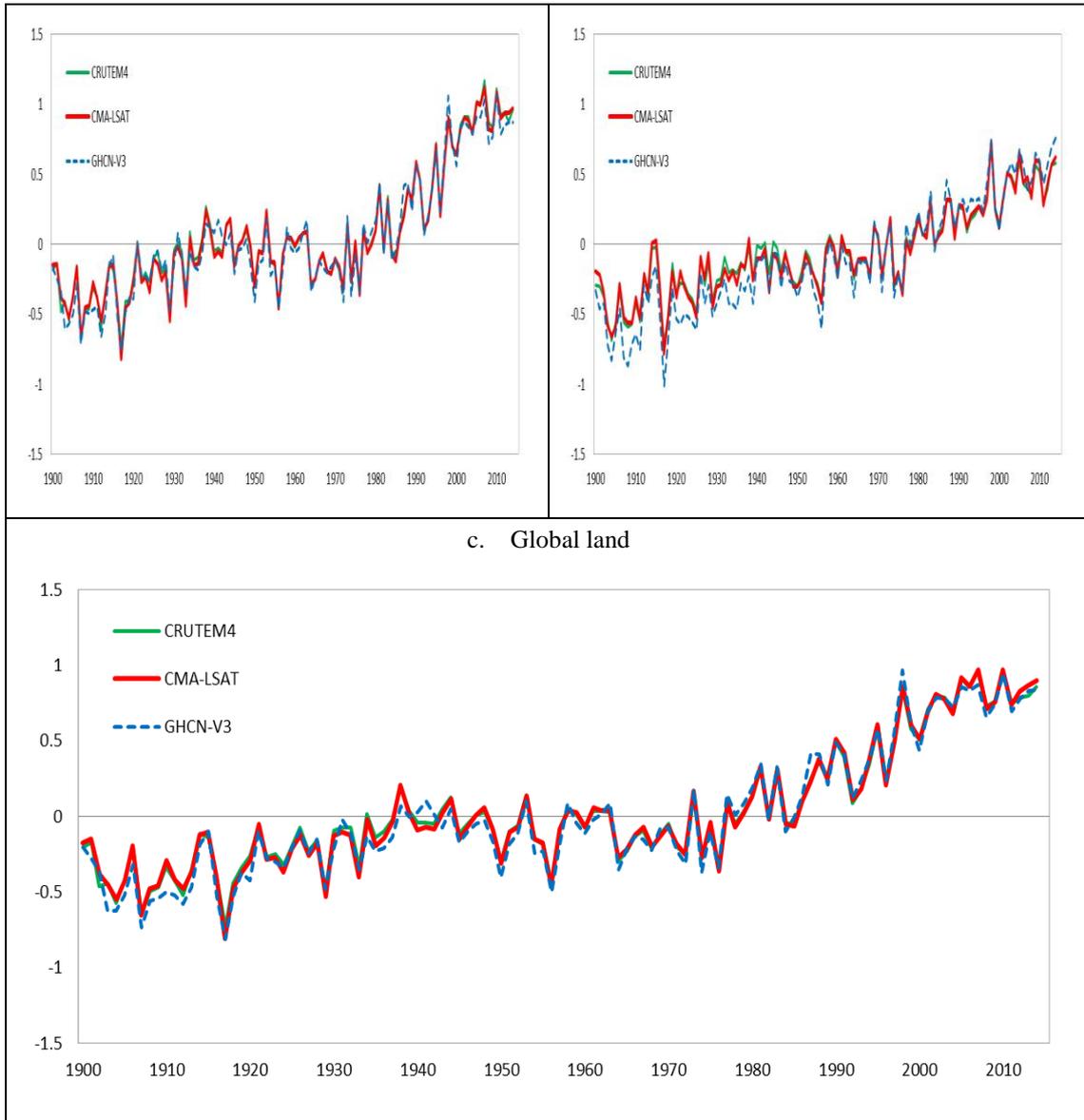### 5.1 Annual

Annual mean LSAT anomaly time series for both hemispheres and the entire

globe during 1900–2014 are shown in Figure 8. The temperature anomaly curves

1 show extremely high similarity with those reported in previous studies (Hartmann et

2 al., 2013). The linear trends in annual mean LSAT for the NH, the SH and the whole

3 globe are 0.107, 0.083, and 0.102 °C/decade, respectively; all are statistically

4 significant at the 1% confidence level (Table 6). Much of the hemispheric and global

5 warming occurred in two distinct periods: from the 1910s to the late 1930s and from

6 the early 1980s to the mid-2000s. The relatively cool periods or stable periods

7 appeared in the 1900s, the 1940s–1970s, and over the last 10 years (2005–2014).

8 These results are very similar to those found in previous studies (e.g., Hansen et al.,

9 2006; Smith et al., 2008; Jones et al., 2012; Jones, 2016).

10   The annual warming is larger in the NH (0.107 °C/decade) than in the SH

11 (0.083 °C/decade). However, the SH lands exhibit a slight warming from the early

12 1950s to the early 1970s, in contrast to the NH lands, which witness a slight cooling.

13 The hemispheric warming that began in the early 1980s is much more remarkable in

14 the NH than in the SH. It is also clear that the global mean LSAT change depends to a

15 larger extent on that from the NH, owing to there being more land (and hence grid

16 boxes in our LSAT dataset) in the NH than in the SH, as well as a higher proportion

17 of land grid boxes with available data.

a. Northern Hemisphere land	b. Southern Hemisphere land

c. Global land

Figure. 8. Annual mean LSAT anomalies (°C) during 1900–2014 for the Northern Hemisphere (a), Southern Hemisphere (b), and entire globe (c) (compared with GHCN-V3 and CRUTEM4)

From 1979 to 2014, the mean LSAT anomalies in the NH, the SH, and the entire globe experienced unprecedented and highly significant annual warming trends, reaching 0.305, 0.142, and 0.272 °C/decade, respectively.

Table 6 also gives the linear trends of annual mean LAST for the periods 1998–2014, since a number of studies have looked at this period because of the debate about "hiatus". Slower warming trends are observed in the 1998-2014 period for the

NH, SH and globe (0.150, 0.120 and 0.124 °C/decade respectively). These warming trends, although still stronger than those observed over the full 1900-2014 period, had been interpreted by many as a "hiatus" (e.g., Slingo et al., 2013), although some more recent analyses (e.g. Karl et al., 2015;Lewandowsky et al., 2015; Mann et al., 2016) question the existence of such a "hiatus" in any significant sense, or note that any slowdown was driven primarily by the oceans (Dai et al., 2015) and was less evident in LSAT. More pronounced "slowdowns", or even local cooling, in the 1998-2014 period are evident in specific regions, such as parts of North America, central and eastern Asia, and northern Australia (Figure 9c), with China showing a particularly pronounced slowdown (Li et al., 2015; Duan et al., 2015; Zhai et al., 2016).

The warmest years in the CMA-LSAT temperature record are concentrated in the later part of the data set, with 15 of the 16 warmest years in the global LSAT record being the 15 years from 2000 to 2014. Furthermore, the World Meteorological Organization has reported that 2015 was warmer than any year of the pre-2015 period, and that 2016 is very likely to be warmer still (WMO, 2016a, 2016b). 2015 was also reported to be the warmest year on record in China's national data set (Zhai et al., 2016).

The spatial distributions of annual mean LSAT trends for the periods 1900–2014, 1978–2014, and 1998–2014 are shown in Figure 9. There is spatially coherent warming at the global scale during 1900–2014, although the warming rates in most regions are below 0.2 °C/decade (Figure 9a). During 1979–2014, however, the global land surface warming trends are clearly higher than those of the entire time period,

1   with particularly large trends occurring at high latitudes of the NH (Figure 9b).

2   During the recent period of slower warming (1998–2014), a strong incoherence in the

3   global LSAT changes can be seen, with abnormal warming in Arctic areas

4   neighboring the Eurasian Continent and the North Atlantic Ocean, and substantial

5   cooling in North America, eastern and central Asia, northern Australia, and southern

6   Africa (Figure 9c) , which is quite similar with the previous studies based on both the

7   other observational and satellite datasets (Cowtan and Way, 2014). However, trends

8   over such a short period have a large uncertainty associated with them, especially in

9   regions with large interannual temperature variability such as continental interiors at

10   mid- to high latitudes in Asia and North America, and hence these geographic patterns

11   should be interpreted with caution.

1

Table 6 Linear trends and their 95% confidence ranges of annual mean LSAT for the Southern Hemisphere, Northern Hemisphere, and entire globe during

three different periods. *indicate trends that are significant at the 5% level.   Unit: °C/decade

|	1900-2014			1979-2014			1998-2014		
	NH	SH	Glo	NH	SH	Glo	NH	SH	Glo
DJF	0.113±0.011*	0.074±0.006*	0.104±0.009*	0.281±0.052*	0.126±0.024*	0.244±0.042*	-0.140±0.147	0.147±0.088	-0.107±0.123
MAM	0.124±0.009*	0.080±0.007*	0.112±0.007*	0.366±0.037*	0.079±0.037*	0.284±0.031*	0.149±0.095*	-0.005±0.143	0.147±0.087*
JJA	0.091±0.007*	0.083±0.005*	0.090±0.006*	0.300±0.031*	0.152±0.026*	0.267±0.026*	0.225±0.086	0.039±0.095	0.178±0.075
SON	0.095±0.009*	0.087±0.006*	0.092±0.007*	0.314±0.042*	0.193±0.024*	0.287±0.035*	0.142±0.106	0.252±0.068*	0.183±0.082
ANN	0.107±0.007*	0.083±0.005*	0.102±0.006*	0.305±0.030*	0.142±0.021*	0.272±0.025*	0.150±0.056	0.120±0.081	0.124±0.057

4

1
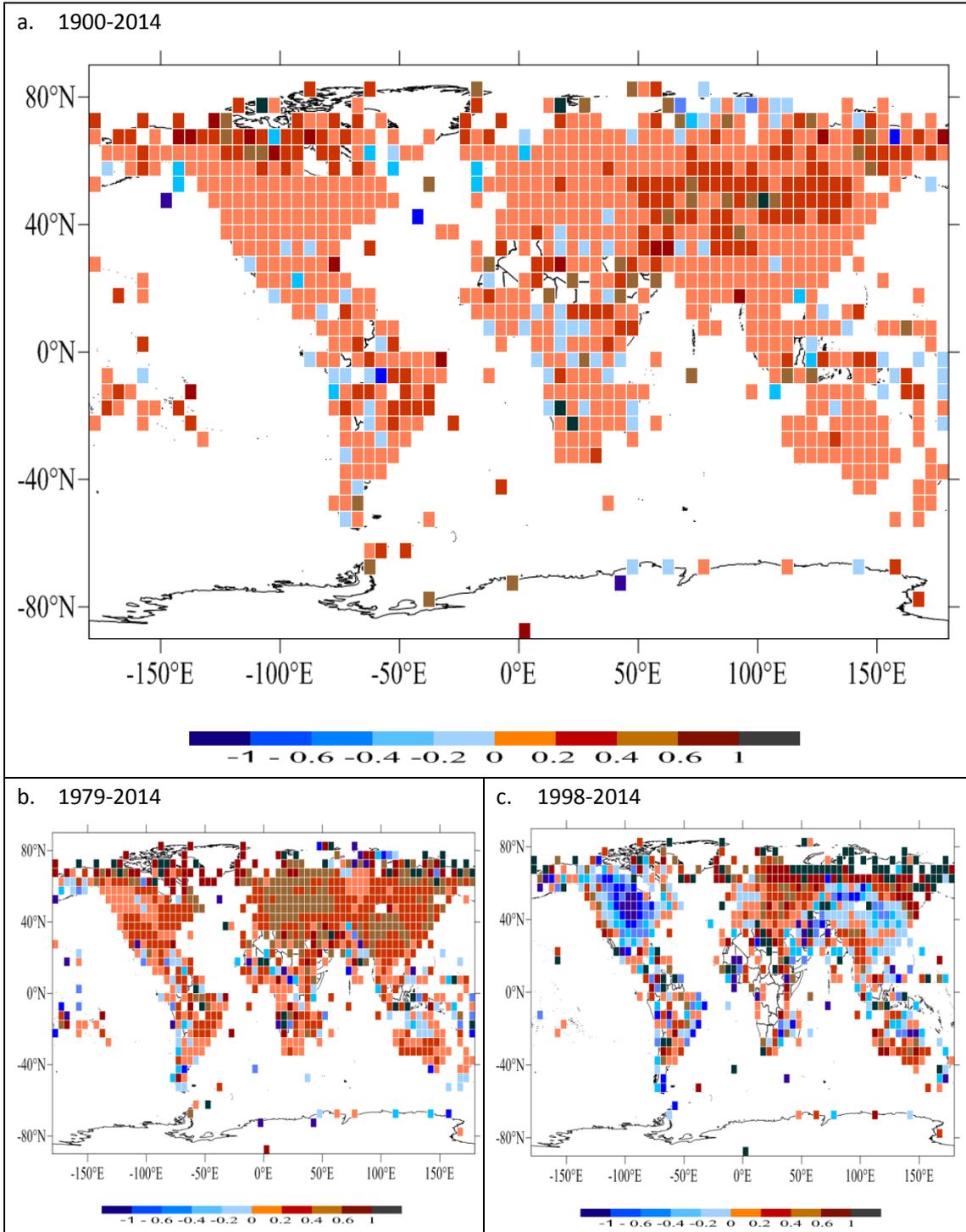
Figure 9 Trends in global land-surface air temperature from CMA-LSAT over three different periods (white area indicates missing data): 1900–2014 (a), 1979–2014 (b), and 1998–2014 (c). Units: °C/decade.
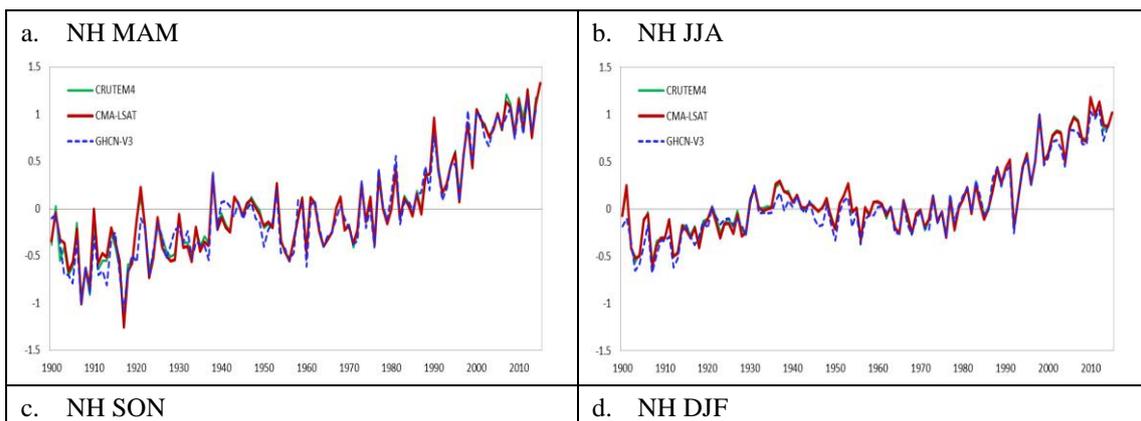
5.2 Seasonal

Seasonal mean LSAT anomaly time series for the globe during 1900–2014 are

shown in Figure 10. For the first distinct global warming period, from the 1910s to the

1　late 1930s, JJA and MAM express warming characteristics; whereas DJF and SON

2　appear relatively stable (DJF shows a little cooling). All seasons express distinct

3　warming characteristics across the globe from the early 1980s to the mid-2000s. All

4　seasonal series show weak trends during the 1940s–1970s
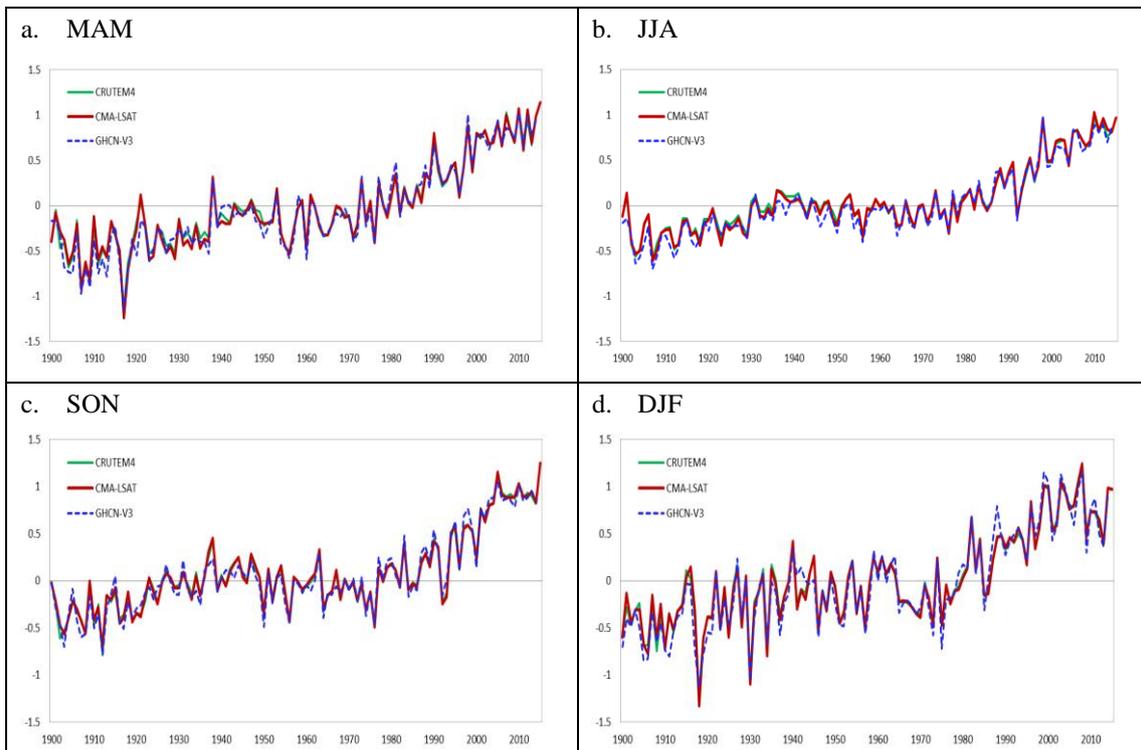
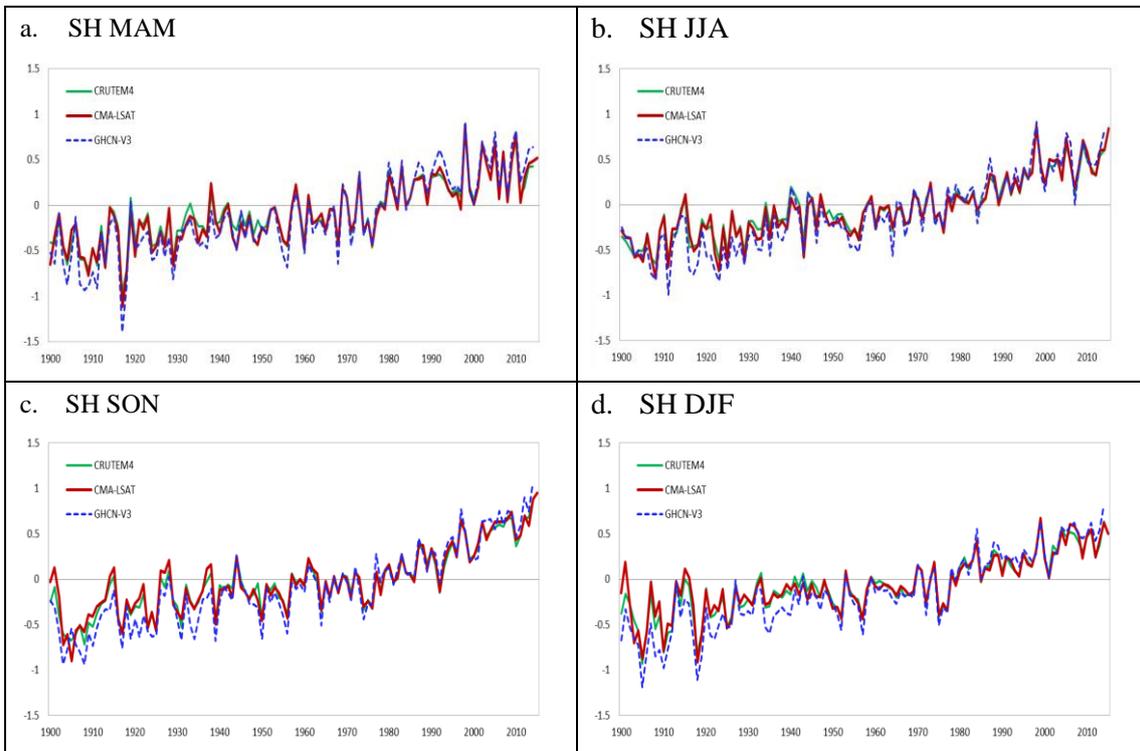5　Figure 10 Seasonal mean LSAT anomaly time series for the entire globe during 1900–2014

6
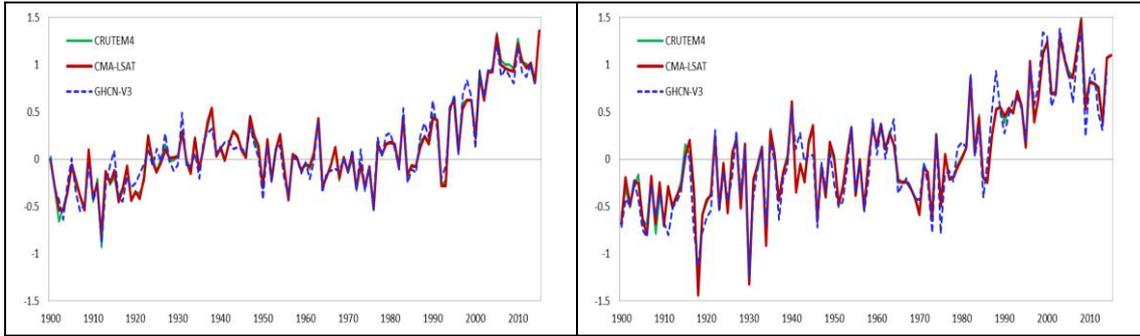
Figure 11 Seasonal mean LSAT anomaly time series for the Northern Hemisphere during 1900–2014.

For the NH (Figure 11), year-to-year variability is greatest during DJF and lowest in JJA. All seasonal series show comparable century-scale warming from the beginning of the 20th century, but there are differences between them in terms of timing. Warming is significant in all seasons during 1900–2014; it is greatest during MAM and lowest in JJA. For the SH (Figure 12), year-to-year variability shows greater similarity between the seasons, as the SH land areas are more influenced by the oceans than for the NH. Warming is greatest during JJA and lowest in DJF.

Figure 12 Seasonal mean LSAT anomaly time series for the Southern Hemisphere during

1900–2014.

Table 6 also gives the linear trends of seasonal mean LSAT for both hemispheres and the entire globe for the periods 1979–2014 and 1998–2014. From 1979 to 2014, all seasonal mean LSAT trends for the SH are weaker than those for the corresponding season/period in the NH. In all seasons (except for MAM in the SH) and annual series for both hemispheres and the entire globe, warming rates are faster for 1979-2014 than for 1900-2014. For the NH, the warming trends for 1979-2014 show slight differences between the seasons. Warming is greatest during MAM and lowest in DJF, while for the SH, warming is the lowest during MAM and greatest during SON.

DJF shows a cooling trend over the 1998-2014 period (Table 6) whilst the other three seasons show warming. The DJF cooling is confined to the Northern Hemisphere, with the Southern Hemisphere showing warming in this season. The Southern Hemisphere shows strong warming for 1998-2014 in SON, and weak trends in MAM and JJA.

6. Summary

Motivated by the need to improve station coverage over Asia and to provide real-time monitoring of LSAT, in this paper we have detailed an effort to develop the CMA-LSAT dataset of monthly LSAT from 1900-present. This data set, which is freely available from the NMIC of CMA, is a collaborative product of scientists at the CMA and many developers of some global, regional, national homogenized SAT

datasets. This new data set benefits from these datasets and from the improvements to

the dataset described in this paper. The main characteristics of the dataset include:

1) The new database used 14 sources of LSAT and it contains 12,374 stations, of

which 9765 could be used in gridding (the average start year is 1938) because they

contain sufficient data for the 1961–90 reference period to calculate the average. The

records from the remaining 2609 stations are relatively short (average length is 18

years) and start later (the average start year is 1966). Spatial coverage is improved

compared to two other datasets (GHCN-V3 and CRUTEM4) during 1900-2014 and

reaches a maximum during the 1951–1990 period. It is worth noting that there have

been consistently more stations in the CMA-LSAT dataset than the other cited

datasets since 1990.

2) For the new dataset, we used homogenized data. About 50% of these data

collected by CMA (including the "doubtful" and "suspect" stations from ECA&D)

and homogenized by us using the PMT method (Wang et al., 2008), while the rest

were obtained from existing national and regional homogenized datasets for Australia,

Canada, China, USA, HISTALP and the "useful" station series from ECA&D. The

improved homogeneity and regional coverage over Asia and other regions increase

the usefulness of this dataset for regional climate change assessment.

3) The global and hemispheric average series developed from the CMA-LSAT

dataset were also compared with the results from two other centers (NCEI: Peterson et

al., 1998; and CRUTEM4: Jones et al., 2012). As reported elsewhere (Peterson et al.,

1998; Jones et al., 1999; Folland and Karl, 2001), the trends for the NH agree very

well, whereas slightly larger differences occur in the SH (CMA-LSAT shows more similarity to CRUTEM4 than to GHCN3). Based on CMA-LSAT, the best evaluation of the trends for the global, NH, and SH land are, respectively, 0.102 ± 0.006, 0.107 ± 0.007, and 0.083 ± 0.005°C/decade during 1900–2014; 0.272 ± 0.025, 0.305 ± 0.030, and 0.142 ± 0.021°C/decade during 1979–2014; and 0.124 ± 0.057, 0.150 ± 0.056, and 0.120 ± 0.081°C/decade during 1998–2014.

Appendix A

All the datasets have been made available through the national meteorological services website (http://10.1.64.154/portal/web-home.index).

Monthly updates will be based on a three-step procedure (Figure A1), as follows. 1) The first round updates are based on hourly real-time data from the WMO Global Telecommunications System (GTS) and Integrated Surface Database (ISD) from NCEI. The daily values are calculated every day from the merged (GTS and ISD) and quality-controlled hourly datasets, and then the monthly temperature is obtained and

1 updated in the dataset at the beginning of the following month. 2) Monthly mean

2 values from CLIMAT message through GTS are provided to update the dataset on

3 about the 20th day of the following month; this is the second round of dataset update.

4 In the two steps, only the WMO stations are matched to be updated. The identification

5 of the same station is performed by the WMO station number. 3) Other monthly data

6 from existing homogenized datasets of the data sources listed in Table 1 are accessed

7 and updated to the dataset. This round of updates is not carried out regularly and must

8 be based on the update of each homogenization data source. Generally speaking, the

9 last round of data update replaces the previous two rounds if the stations are found to

10 be the same in previous rounds of data update. In the step, the identification of the

11 same station is performed by the data source symbol and the station number. It is

12 worth noting that, there is no new merging algorithm to be used during the process of

13 updating.

14
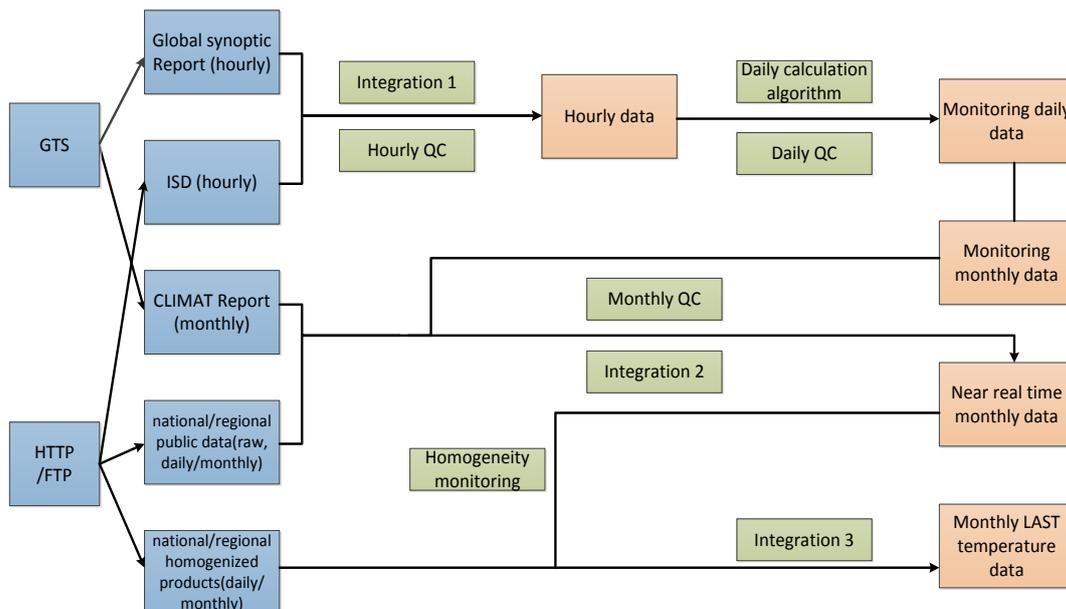
15
16 Figure A1 Near-real-time integration and updating procedures for CMA-LSAT

17

References

Alexander L. V., et al., (2006), Global observed changes in daily climate extremes of temperature and precipitation, *J. Geophys. Res.*, 111, D18116, doi:10.1029/2005JD006290

Auer I. et al., (2007) HISTALP - Historical instrumental climatological surface time series of the greater Alpine region 1760-2003. *Int. J. Climatol.*, 27:17-46

Brohan P., Kennedy, J., Harris, I., Tett, S.F.B. and Jones, P.D., (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophys. Res.*111, D12106, doi:10.1029/2005JD006548.

Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate. *J. Roy. Stat. Soc.,* **C53,** 405–425.

Cao L. J., Zhao, P., Yan, Z. W., et al., (2013) Instrumental temperature series in eastern and central China back to the 19th century. *J Geophys Res Atmos.* 118 (15), 8197-8207

Cowtan K and RG Way, (2014), Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society,* 140(683): 1935–1944, doi: 10.1002/qj.2297

Dai A., J. Wang, P. W. Thorne, D. E. Parker, L. Haimberger, and X. L. Wang, (2011) A new approach to homogenize daily radiosonde humidity data, *J. Climate*, 24, 965–991, doi:10.1175/2010JCLI3816.1.

Dai A, J. Fyfe, S. Xie, X. Dai, (2015) Decadal modulation of global surface temperature by internal climate variability, *Nature Climate Change* 5, 555–559 , doi:10.1038/nclimate2605

Ding Yihui, and Xiaosu. Dai, (1994) Temperature Variation in China during the Last 100 years, *Meteorology*, 20, 19-26.

Domonkos P., V. Venema and O. Mestre, 2013: Efficiencies of homogenisation methods: our present knowledge and its limitation. Proceedings of the Seventh seminar for homogenization and quality control in climatological databases (http://www.wmo.int/ pages/ prog/ wcp/ wcdmp/ documents/finalWCDMP78.pdf)

Duan, A. and Xiao, Z., (2015) Does the climate warming hiatus exist over the Tibetan Plateau? *Sci. Rep.* 5, 13711; doi: 10.1038/srep13711.

Durre I., Menne M.J., Vose R., (2007) Strategies for evaluation quality assurance procedures. *Journal of Applied Meteorology and Climatology*. DOI: 10.1175/2007JAMC1706.1

Fyfe J. C., N. P. Gillett, and F. W. Zwiers (2013) Overestimated global warming over the past 20 years, *Nat. Clim. Change*, 3, 767–769.

Feng S, Q Hu, and W. Qian, (2004) Quality Control of Daily Meteorological Data in China, 1951–2000: A New Dataset, *Int. J. Climatol.*, 24, 853 – 870.

IPCC. 2007. Climate Change (2007) The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. and Miller, H. L. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Guemas V., F. J. Doblas-Reyes, I. Andreu-Burillo, andM. Asif (2013), Retrospective prediction of the global warming slowdown in the past decade, Nat. Clim. Change, 3, 649–653.

Hansen J., Ruedy, R., Glascoe, J., et al., (1999) GISS analysis of surface temperature change. *J Geophys Res.* 104, 30997-31022.

Hansen J., R. Ruedy, M. Sato and K. Lo (2010) Global surface temperature change. *Rev. Geophys.*, 48, RG4004, doi: 10.1029/2010RG000345.

Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H.,(2014) Updated high-resolution monthly grids of monthly climatic observations: the CRU TS 3.10 dataset. *Int. J. Climatol.*, **34**, 623-642, DOI:10.1002/joc.3711.

Hartmann D. L., A. M. G. Klein Tank, M. Rusticucci, L. V. Alexander, S. Brönnimann, Y. Charabi, F. J. Dentener, E. J. Dlugokencky, D. R. Easterling, A. Kaplan, B. J. Soden, P. W. Thorne, M. Wild and P. M. Zhai, (2013) Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Houghton J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson (Eds) (2001), Climate change 2001, in The Scientific Basis, 881 pp., Cambridge University Press, Cambridge.

Jaccard P.(1901) Etude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Societe Vaudoise des Sciences Naturelles,* 37: 547–579.

Jones P. D., S. C. B. Raper, B. Santer, B. S. B. Cherry, C. Goodess, P. M. Kelly, T. M. L. Wigley, R. S. Bradley, and H. F. Diaz，(1985) A grid Point Surface Air Temperature Data Set for the Northern Hemisphere, TR022, Department of Energy, Washington, 251 pp.

Jones P.D., (1994) Hemispheric Surface Air Temperature Variations: A Reanalysis and an Update to 1993, *J Climate*, 7(11):1794-1802

Jones PD，M New，DE Parker，S Martin，IG Rigor, (1999) Surface air temperature and its changes over the past 150 years, *Rev Geophys*，37(2):173-199

Jones P. D., and A. Moberg，(2003) Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Climate.*, 16, 206–223

Jones P D，Lister D H，and Li Q，(2008) Urbanization effects in large-scale temperature records, with an emphasis on China. *J Geophys Res*，113, D16122, doi:10.1029/2008/JD009916.

Jones P. D., Lister, D. H., Osborn, T. J., Harpham, C.,Salmon, M., and Morice, C. P.,(2012) Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010, *J. Geophys. Res.*, 117, D05127, doi:10.1029/2011JD017139

Jones P.D, (2016) The reliability of global and hemispheric surface temperature records. *Adv. Atmos. Sci.*, 33(3), 269－282, doi: 10.1007/s00376-015-5194-4.

Karl T.R., A. Arguez, B. Huang, J.H. Lawrimore, J.R. McMahon, M.J. Menne, T.C. Peterson, R.S. Vose and H-M. Zhang (2015) Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, 348, 1469-1472, doi: 10.1126/science.aaa5632.

Klein Tank A. M. G., Wijngaard, J.B., Konnen, G.P., et al., (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *Int. J. Climatol..* 2002, 22, 1441-1453.

Kuglitsch, F. G., R. Auchmann, R. Bleisch, S. Broennigmann, O. Martius,and M. Stewart ,(2012) Break detection of annual Swiss temperature series, *J. Geophys. Res.*, 117, D13105, doi:10.1029/2012JD017729.

Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose,   and J. Rennie, (2011) An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3, *Journal of Geophysical Research*, 116, D19121,doi:10.1029/2011jd016187.

Legates DR, McCabe GJ. (1999) Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model evaluation. *Water Resources Research* 35: 233–241, doi: 10.1029/1998WR900018.

Lewandowsky, S., J. Risbey, and N. Oreskes, (2015) The "Pause" in Global Warming: Turning a Routine Fluctuation into a Problem for Science. *Bull. Amer. Meteor. Soc.* 97 (5) : 723-733, doi:10.1175/BAMS-D-14-00106.1.

Li Q., Liu, X., Zhang, H., et al., (2004a) Detecting and adjusting on temporal inhomogeneity in Chinese mean surface air temperature datasets. *Adv Atmos Sci.* 21, 260-268.

Li Q., H. Z Zhang, X. N Liu and J.-Y Huang, (2004b) Urban heat island effect on annual mean temperature during the last 50 years in China. *Theor. Appl. Climatol.*, 79, 165-174.

Li Q., Dong, W., J., (2009) Detection and adjustment of undocumented discontinuities in Chinese temperature series using a composite approach. *Adv Atmos Sci.* 26(1), 143-153.

Li Q., H. Zhang, J. Chen, W. Li, X. Liu and P. Jones, (2009) A mainland China homogenized historical temperature dataset of 1951-2004. *Bull. Amer. Meteor. Soc.*, 90, 1062-1065, doi: 10.1175/2009BAMS2736.1.

Li, Q.., W. Dong, W. Li, X. Gao, P. Jones, J. Kennedy and D. Parker, (2010a) Assessment of the uncertainties in temperature change in China during the last century，*Chin. Sci. Bull.* 55, 1974-1982, doi: 10.1007/s11434-010-3209-1

Li Q, Li,W., Si, P., et al., (2010b) Assessment of surface air warming in northeast China, with emphasis on the impacts of urbanization. *Theor Appl Climatol.* doi:10.1007/ s00704-009-0155-4

Li Q., (2013) Development of homogenized data sets in China and the possible contributions to global/regional products, International Workshop on Climate Data Requirements and Applications, March 4-8, Nanjing

Li Q., Yang, S., Xu W.H., et al., (2015) China experiences recent warming hiatus. *Geophys Res Lett*,42(3). DOI: 10.1002/2014GL062773

Li Q., Zhang L., Xu W., et al., (2016) Comparisons of time series of annual mean surface air temperature for China since 1900: observations, model simulations and extended reanalysis. *Bull Amer Meteorol Soc.,* doi:10.1175/BAMS-D-16-0092.1, in press.

Li Z., Yan, Z. W., (2010) Application of multiple analysis of series for homogenization (MASH) to Beijing daily temperature series 1960-2006. *Adv Atmos Sci.* 27(4), 777-787.

Mann M. E. et al., (2016) The Likelihood of Recent Record Warmth. *Sci. Rep.* 6, 19831; doi: 10.1038/srep19831

Menne M. J., Williams, J.R., (2009) Homogenization of temperature series via pairwise comparisons. *J. Climate*, 22, 1700-1717.

Oke T.R., （2004）*Initial Guidance to Obtain Representative Meteorological Observations Urban Sites*, Instruments and Methods of Observation Program, IOM Report No. WMO/TD 1250, World Meteorological Organization, Geneva.

Parker D.E., (1994) Effects of changing exposure of thermometers at land stations. *Int. J. Climatol.,* 14, 1-31.

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

Parker D.E., (2006) A demonstration that large-scale warming is not urban. *J. Climate,* **19**, 2882-2895.

Peterson T. C., and D. R. Easterling ,(1994) Creation of homogeneous composite climatological reference series, *Int. J. Climatol.*, 14, 671–679

PetersonT.C., Vose, R.S., (1997)　An overview of the global historical climatology network temperature data base. *Bull Amer Meteorol Soc.* 78,2837-2849.

Peterson TC, Easterling DR, Karl TR, Groisman P,Nicholls N, et al.,(1998) Homogeneity adjustments of in situ atmospheric climate data: a review. *Int J Climatol.*, 18:1493–1517.

Peterson T.C., (2003) Assessment of urban versus rural *in situ* surface temperatures in the contiguous United States: no difference found. *J. Climate,* **16**, 2941–2959.

Peterson T.C. and T.W. Owen, (2005) Urban heat island assessment: Metadata are important. *J. Climate*, **18**, 2637–2646.

Ren Guoyu，Chu Ziying，Zhou Yaqing，et al,，(2005) Recent Progresses in Studies of Regional Temperature Changes in China，*Climatic and Environmental Research*，10(4): 701-716.

Rennie J.J., et al., (2014) The International surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geoscience Data Journal.*

Rohde R., Muller, R. A., Jacobsen, R., et al., (2013) A new estimate of the average earth surface land temperature spanning 1753 to 2011 . *Geoinfor Geostat: An Overview.* doi:10.4172/gigs.1000101.

Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S. and Thépaut, J.-N. , (2017) A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. Q.J.R. Meteorol. Soc., 143: 101–119. doi:10.1002/qj.2949

Slingo, J., et al., (2013) The recent pause in global warming parts 1-3Rep., The Met Office, FitzRoy Road, Exeter, U. K.

Sun Q, Miao C, Duan Q, et al., (2014) Would the 'real' observed dataset stand up? A critical examination of eight observed gridded climate datasets for China. *Environmental Research Letters*, 9(1).015001:1-16, doi:10.1088/1748-9326/9/1/015001

Toreti, A., F. G. Kuglitsch, E. Xoplaki, and J. Luterbacher (2012), A novel approach for the detection of inhomogeneities affecting climate time series, J. Appl. Meteorol. Climatol., 51, 317–326, doi:10.1175/JAMC-D-10-05033.1

Trewin B.C., (2004) Effects of changes in algorithms used for the calculation of Australian mean temperature. *Aust. Met. Mag.*, 53, 1-11.

Trewin, B.C., (2010) Exposure, instrumentation and observing practice effects on land temperature measurements. Wiley Interdisciplinary Reviews: *Climatic Change*, 1, 409-506, doi: 10.1002/wcc.46.

Trewin B.C., (2013) A daily homogenized temperature data set for Australia. *Int. J. Climatol.* 33, 1510-1529.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T., (2012)

Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89-115, doi:10.5194/cp-8-89-2012,

Vincent L.A., Wang, X. L., Milewska, E. J., et al.,(2012) A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J Geophys Res.* 117, D18110, doi: 10.1029/2012JD017859.

Vincent L. A., X. Zhang, R. D. Brown, Y. Feng, E. Mekis, E. J. Milewska, H. Wan, and X. L. Wang, (2015) Observed Trends in Canada's Climate and Influence of Low-Frequency Variability Modes. *J. Climate*, 28, 4545–4560. doi: http: //dx.doi.org/ 10.1175/ JCLI-D-14 -00697.1

Vose R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R.Heim, T. R. Karl, and J. Eischeid, (1992) The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041, 325 pp. [Available from Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831.]

Wan H., X. L. Wang, and V. R. Swail, (2010) Homogenization and trend analysis of Canadian near-surface wind speeds, *J. Climate*, 23, 1209–1225.

Wang J., Xu C., Hu M., et al., (2014) A new estimate of the China temperature anomaly series and uncertainty assessment in 1900-2006. *J Geophys Res Atmos.* 119, doi: 10.1002/2013JD020542.

Wang J., Xu C., Hu M., et al., (2016) Global Land Surface Air Temperature Variations since the 1880, Submitted to *J Climate*

Wang X. L., Q. H. Wen, and Y. Wu, (2007) Penalized maximal t test for detecting undocumented mean change in climate data series, *J. Appl. Meteorol. Climatol.*, 46, 916–931.

Wang, X. L., (2008a) Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal *t* or *F* test. *J. App. Meteor. Climatol*, **47**, 2423–2444. DOI: 10.1175/2008JAMC1741.1.

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

Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, (2010) New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.* 49 (No. 12), 2416-2436. DOI: 10.1175/2010JAMC2376.1.

Wang X. L., and Y. Feng, (2010) RHtestsV4 User Manual, Climate Research Division, Science and Technology Branch, Environment Canada, Toronto, Ontario, Canada. 29 pp. Available online at at http://etccdi.pacificclimate.org/software.shtml.

Wang X. L., Y. Feng, and L. A. Vincent, (2013) Observed changes in onein-20 year extremes of Canadian surface air temperatures, *Atmos.Ocean*, 52(3):222-231, doi:10.1080/07055900.2013.818526, in press.

Wang, K.C. and C. L. Zhou, (2015) Regional Contrasts of the Warming Rate over Land Significantly Depend on the Calculation Methods of Mean Air Temperature. *Sci. Rep.* 5, 12324; doi:10.1038/srep12324.

Wickham, C., and Coauthors, (2013) Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications. *Geoinfor Geostat: An Overview*, **1**, doi: 10.4172/2327-4581.1000104.

Wijngaard, J. B., Klein, Tank A. M. G., Konnen, G. P., et al., (2003) Homogeneity of 20th century

European daily temperature and precipitation series. *Int. J. Climatol.*. 23, 679

Williams CN, Menne MJ, Thorne PW. (2012) Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.,* 117: D05116, doi: 10.1029/2011JD016761

Willmott CJ. (1981) On the validation of models. *Physical Geography* 2: 184–194.

Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O'Donnell J, Rowe CM. (1985) Statistics for the evaluation and comparisons of models. *Journal of Geophysical Research* 90: 8995–9005, doi: 10.1029/JC090iC05p08995.

Xu W., Q. Li,, X. Wang, et al., (2013) Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices . *J Geophys Res Atmos.* 118(17) , 9708-9720

Yan Z，Z Li，Q Li，et al.，(2009) Effects of Site-Change and Urbanisation in the Beijing Temperature Series 1977-2006. *Int J Climat*，doi：10.1002/joc.1971 .

Yin H, Donat M G, Alexander L V, et al., (2015) Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *Int J Climatol.*, 35(10): 2809–2827. DOI: 10.1002/joc.4174

You Q., S. Kang, E., Anguilar, et al., (2011) Changes in daily climate extremes in China and its connection to the large scale atmospheric circulation during 1961–2003, *Clim Dyn* , 36: 2399–2417.DOI 10.1007/s00382-009-0735-0

Zhai P., Q. Chao, and X. Zou, (2004) Progress in China's climate change study in the 20th century, *J. Geograph. Sci.*, 14, 3-11, doi: 10.1007/BF02841101.

Zhai Panmao, Yu Rong, Guo Yanjun, Li Qingxiang, Ren Xuanjuan,Wang Yaqiang, Xu wenhui, Liu Yanju, and Ding Yihui, (2016) The strong El Niño in 2015/2016 and its dominant impacts on global and China's climate, *J Meteor. Res*, 74(3):309-321, doi:10.11676/qxxb2016.049

Zhang X., G. Hegerl, F. W. Zwiers, and J. Kenyon, (2005) Avoiding inhomogeneity in percentile-based indices of temperature extremes, *J. Climate*, 18, 1641–1651, doi:10.1175/JCLI3366.1.

Zhou L, R.E. Dickinson, Y. Tian, J. Fang, Q. Li and R.K. Kaufmann, (2004) Evidence for a significant urbanization effect on climate in China. *Proc. Natl. Acad. Sci.* USA, 101, 9540-9544.