Global Land Surface Air Temperature Dynamics since 1880

Jinfeng Wang[1*], Chengdong Xu[1], Maogui Hu[1], Qingxiang Li[2], Zhongwei Yan[3] and Phil Jones[4,5]

1. LREIS, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
2. National Center of Meteorological Information, CMA, Beijing 100081, China
3. Institute of Atmospheric Physics, CAS, Beijing 100029, China
4. Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, UK
5. Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author email address:

JFW: wangjf@lreis.ac.cn

ABSTRACT

The geographical extent, magnitude, and uncertainty of global climate change have been widely discussed and have critical policy implications at both global and local scales. In this study, a new analysis of annual mean global land surface air temperature since 1880 was generated, which has greater coverage and lower uncertainty than previous distributions. The Biased Sentinel Hospitals Areal Disease Estimation (BSHADE) method, used in this study, makes a best linear unbiased estimation (BLUE) when a sample is small and biased to a spatially heterogeneous population. For the period of 1901–2010, the warming trend was found to be 0.109°C/decade with 95% confidence intervals between 0.081°C and 0.137°C. Additionally, warming exhibited different spatial patterns in different periods. In the early 20th century (1923–1950), warming occurred mainly in the mid-high latitudes of the Northern Hemisphere, whereas in the most recent decades (1977–2014), warming was more spatially extensive across the global land surface. Compared with other common methods, the difference in results appears in the areas with few stations and in the early years, when stations had sparse coverage and were unevenly distributed. Validation, which was performed using real data that simulated the historic situation, showed a smaller error in the BSHADE estimate than in other methods. This study produced a new database with greater coverage and less uncertainty that will improve the understanding of climate dynamics on the Earth since 1880, especially in isolated areas and early periods, and will benefit the assessment of climate-change-related issues, such as the effects of human activities.

Key words: global; land surface air temperature dynamics; biased observations; best linear unbiased estimate (BLUE)

1. Introduction

Temperature is a key metric for assessing the state of the climate. The extent, magnitude, and uncertainty of global surface temperature change have been highly related to policy-making and public affairs on both global and local scales. According to the Intergovernmental Panel on Climate Change, the last three decades are the warmest period since the mid-19th century, and the warming is unequivocal and unprecedented (Hartmann et al., 2013). Many studies indicate that global warming will negatively impact human activities, natural environments, and ecosystems, such as ice melting, sea level rise, floods and droughts, the spread of disease, human health, species extinction, etc. (Gething et al., 2010; Hansen et al., 2006; McMichael et al., 2006; Patz et al., 2005; Rahmstorf, 2007; Walther et al., 2002). These studies have directed the focus of science towards explaining the driving forces behind the rapid warming of the Earth, and today there is widespread agreement that human activity is the dominant cause for the increase of greenhouse gases, although uncertainty of its relative contribution still remains (Bindoff et al., 2013; Qin, 2014; Santer et al., 1996; Stott et al., 2000). It is essential to construct a spatial analysis of the global land surface temperature at a large scale and with less uncertainty from the limited and even biased observations made since 1880. Doing so will enable a thorough understanding of the pace of climate change and its effects on human activity at both a global and local basis.

Currently, maps of global land surface air temperature using instrumental records have been developed mainly by four groups: the UK Met Office Hadley Centre and the University of East Anglia Climatic Research Unit (CRUTEM4), the National Oceanic and Atmospheric Association's (NOAA's) National Center for Environmental Information (NCEI), the NASA Goddard Institute for Space Studies (GISS), and the Berkeley Earth Surface Temperature Project (Berkeley) (Jones, 2016). The results published by these groups correspond with each other after 1900 (Hansen et al., 2010; Hartmann et al., 2013), while there are greater differences between their results before the early 20th century, although similar data sources were used (Jones and Wigley, 2010; Lawrimore et al., 2011; Vose et al., 2005). The differences are mainly caused by the various groups using different approaches to remove the inhomogeneities of the dataset and deal with the issue of sparsely distributed stations, which is an important uncertainty source in global or regional (i.e., continental) mean temperature estimation in these early decades (Jones, 2016; Brohan et al., 2006; Hansen et al., 2010; Jones et al., 2012; Jones and Wigley, 2010). The influence of sparse data coverage first appeared before 1950 (Lawrimore et al., 2011), and estimation error decreased as station coverage become more dense.

The influence of sparse station coverage on the observed climate is also evident in recent years due to international exchange of data and station closures. This reduction in station numbers is much more significant in Africa and South America. The sparse coverage of stations results in sample bias when the population is spatially heterogeneous. By sample bias, we mean that the sample's histogram is different from that of the population's. A biased sample will lead to a biased estimate if the sample bias is not accounted for (Wang et al., 2012).

In order to solve this problem, we used the Biased Sentinel Hospitals Areal

Disease Estimation (BSHADE) method in the estimation of the land surface air temperature anomaly and uncertainty for China between 1900–2006 (Wang et al., 2014; Wang et al., 2011; Xu et al. 2013; Hu et al. 2013). In theory, the method has the potential to remedy station bias resulting from sparse coverage when the population is spatially heterogeneous and simultaneously accounting for the characteristics of spatial autocorrelation.

Using station data on China's annual temperature anomaly from 1900–2006, the BSHADE method exhibits a smaller error variance of estimation than traditional methods, especially for periods with sparse station coverage (Wang et al., 2014).

The present study aims to reconstruct the dynamic of temperature anomalies for the global land surface from 1880–2014 using BSHADE and the CRUTEM4.4.0.0 station data. The findings are expected to improve the understanding of historical temperature change since 1880, at both the global and local scales.

The remainder of this paper is organized as follows. In Section 2, the data and methods are described. In Section 3, the results are presented, including: (1) the geographical distribution of global land surface air temperature anomalies; (2) the global land surface air temperature anomaly series; (3) a trend map of global land surface temperature; and (4) validation of the estimation. Section 4 includes a discussion and conclusions.

2. Data and Methods

2.1 Station Data

The CRUTEM4.4.0.0 (Jones et al., 2012) station data, from 1880 to 2014, downloaded from the website of Met Office Hadley Centre, was employed to estimate the spatial distribution of global land surface air temperature. This dataset was constructed using monthly mean temperature data. Quality control was undertaken by checking whether a station's annual average was more than 5 times the standard deviation beyond the average (based on the period of 1941–1990), and the identified outlier records (0.096%) were deleted from the dataset. For any given year, the monthly records having no missing values were averaged to annual values.

Before the 1900s, the spatial distribution of stations was very sparse and highly biased, with the majority of stations located in Western Europe and United States, and only a few stations located on other continents. For example, stations were mainly located near the coastal areas of Africa, South America, Japan, India, and the southeast area of Australia. The stations number increased sharply during the first half of the 20th century between 1901–1960. The station number reaches its maximum in 1961–1990. However, even in recent years, the spatial distribution of stations in some areas is still sparse and uneven, such as in the Antarctic, the Arctic, and the interior of Africa and South America. Figure S1 shows the number of stations from 1880 to 2010. In the station anomaly estimation, reference series were defined as the station data from 1961–1990. Stations less than 15 years of missing data during 1961–1990 were selected, and the average temperatures in the period were estimated from the remaining records (Figure S1A).

The data under study is both spatially autocorrelated and spatially heterogeneous, and the geographical distribution of meteorological stations is highly uneven, especially in some areas and in the earlier years. An estimator's theoretical merits would apply in practice only when its assumption was identical or approximate to reality; therefore we choose to use the BSHADE algorithm in this study.

2.2 BSHADE Algorithm

In BSHADE, the continental mean anomaly $\bar{Y}$ is estimated by a weighted station average $\bar{y}$:

$$\bar{y} = \sum_{i=1}^{n} w_i y_i \quad (1)$$

where w_i (i=1,…,n) is the weight of the i-th station and is calibrated by the Eq. (S1) and observed data.

The weight w_i satisfies the unbiased condition

$$E\bar{y} = \bar{Y} \quad (2)$$

and minimum estimation variance

$$\min_w v(\bar{y}) = E(\bar{y} - \bar{Y})^2 \quad (3)$$

where E denotes the statistical expectation, v indicates statistical variance, and $\bar{Y}$ represents the true average value of an area.

Eq. (2) can be expressed as

$$E\bar{y} = E\sum_{i=1}^{n} w_i y_i = \bar{Y} \quad (4)$$

that is:

$$\sum_{i=1}^{n} w_i b_i = 1$$

where we set

$$b_i = Ey_i/\bar{Y} \quad (5)$$

$b_i = 1$ will guarantee the sample estimator $\bar{y}$ to be unbiased, while $b_i \neq 1$ will lead to $\bar{y}$ being biased. The weight w_i for each station can be calibrated by Eq. (S1), and by insert the weights into Eq. (1), the regional mean anomaly $\bar{Y}$ can be estimated by $\bar{y}$. Furthermore, the estimation variance

$$v(\bar{y}) = E(\bar{y} - \bar{Y})^2 = C(\bar{y}, \bar{y}) + C(\bar{Y}, \bar{Y}) - 2C(\bar{y}, \bar{Y}) \tag{6}$$

can also be calculated by Eq. (6), in which *C* denotes the statistical covariance.

In BSHADE, the characteristic of geographical spatial correlation is indicated by the parameters of the covariance, which is derived by the semivariogram of geostatistics theory (Isaaks and Srivastava, 1989, Chaper 16). The correlation will decrease with the increase of distance between two sites, and the relationship between spatial correlation and distance is different between continents. Some studies use a correlation distance of up to 1200 km (Hansen et al., 2006), while Lawrimore et al. found that temperatures were sufficiently correlated more than 1000 km away (Lawrimore et al., 2011). Figure S2 illustrates a semivariogram representing the relationship between the spatial correlation of the annual temperature anomaly and distance for each continent, which indicates that spatial correlations extend beyond 1000 km in all regions. In order to produce lower uncertainty in this study, 1000 km was used as the distance limitation for the neighbouring station selection in the estimation.

Meanwhile, the bias of sample is quantitatively reflected by the parameter vector $B\{b_i\}$. The parameter b_i is the ratio between the anomaly of the *i*-th station and the continental mean value. This parameter reflects the phenomenon that the mathematical expectation of the station records' mean value is not equal to the true value across the whole continent, an effect which is caused by spatial heterogeneity. The sample bias occurs more clearly in areas with few stations and high heterogeneity and in the early period when the coverage of meteorological stations was sparse and uneven. Due to BSHADE method's ability to account for the characteristics of both the spatial correlation and spatial heterogeneity of the target domain and sample bias, an objective function of errors which is minimized and remedies the biased sample problem to produce an estimate that is BLUE (best linear unbiased estimate). This happens when the assumption of a model approximates the characteristics of a population and the way of sampling. (Wang et al., 2014; Wang et al., 2011; Xu et al., 2013; Wang et al., 2012).

3. Results

3.1. Geographical Distribution of Global Land Surface Air Temperature Anomalies

Annual global land surface air temperature anomaly maps from 1880 to 2014 were developed by the BSHADE method. Each grid box is 5° latitude by 5° longitude. The results are shown in Figure S3. Before the 1900s, the projected temperature anomaly map covers all of Europe; most of North America, except for the regions near the Arctic; Asia, except for some northern areas and western parts of China; and almost the whole area of Australia. Some parts of South America and Africa are missing because too few stations were available. After 1920, there are estimated temperatures for most land areas, except some parts of interior South America and Africa, and all of Antarctica. After 1940, our temperature anomaly distribution maps cover almost all areas.

From the maps in Figure S3, we can see that there is substantial interannual spatial variability for the spatial distribution of the global mean surface air temperature anomaly. For example, in the year 2001, the areas with large positive temperature anomalies were mainly distributed over the northeast of North America, while in the next year, the areas with large positive temperature anomalies were across the Bering Strait, extending to the mid-to-high latitudes of Asia. However, in the year of 2003, the area with the largest positive temperature anomalies moves to the north, compared with the distribution of 2002, and covers higher latitude regions of Europe-Asia and North America.

Besides the global land surface air temperature anomaly, the spatial distribution of the estimation error variance for each year is also presented in Figure S3, which shows that the estimation error variance is significantly smaller in recent years than for earlier years. In addition, the high estimation error is mainly evident over areas that have few stations. For example, in the year 2001, grids with higher estimation error are mainly located over Southeast Asia and West Asia and the interior of Africa. These areas have significantly fewer stations compared with other regions.

3.2. Global Land Surface Air Temperature Anomaly Series

In addition to its application for mapping, BSHADE was also used to estimate continental and global mean temperature anomalies from 1880–2014. In order to compare the estimated results with those from the traditional methods (Jones, 1994), we also calculated results using the CAM and Block Kriging method. Using the CAM approach, anomalies are calculated for all stations within their corresponding grid box, and which are then aggregated to get a regional mean temperature (Jones, 1994). The Block Kriging method produces maps based on the spatial correlation of target fields (Cressie, 1993; Goovaerts, 1997; Isaaks and Srivastava, 1989). The bias of stations and spatial heterogeneity of population were not fully considered in the Block Kriging method. The description of the calculation process of CAM and Block Kriging is presented in supporting information (SI). Figure 1 is the estimated annual temperature anomalies.

All three series in Figure 1 agree on the overall warming trend since 1920 across global land areas. After 1920, the coverage of stations became more evenly distributed and much denser. They differ slightly more before 1920, when the meteorological stations were fewer and more unevenly distributed over global land areas, especially for the period before 1900. In the period between 1880 and 1900, the global land values estimated by the Block Kriging method are lower compared with BSHADE and CAM.

In Table 1, the overall trends of the various temperature series for different time periods are compared. The linear trends for the periods of 1901–1950, 1880–2010, 1901–2010, 1951–2010, and 1979–2014 have been calculated for BSHADE, Block Kriging and CAM with 95% confidence intervals (CI) (Table 1). The confidence intervals of the linear trends were estimated using the generalized least squares technique within each period. The effects of serial autocorrelation in the models' residuals were accounted for (Gujarati, 2003). In the period of 1880–2010, the temperature warms by 0.092–0.108°C/decade, as estimated by the three methods. In the same period, the overall trend estimated by BSHADE was 0.096°C (95% CI: 0.075°C – 0.117°C). This trend is similar to that estimated by CAM but lower than that estimated by Block Kriging. The linear trends in 1901–2010 with 95% CIs for BSHADE, Block Kriging, and CAM were 0.109°C ± 0.028°C, 0.115°C ± 0.029°C, and

0.104°C ± 0.026°C per decade, respectively. In addition, it appears that there is a significant difference between the first and the second halves of the twentieth century (Figure 1). For BSHADE, the 1901–1950 linear trend with 95% CI s was 0.118°C ± 0.032°C, while the trend for 1951–2010 was 0.223°C ± 0.049°C, which is significantly higher than that in the first half of the century. In the two periods, the trend for BSHADE is between the trend identified by the other two methods. For the recent years between 1979 and 2014, the warming trend calculated by BSHADE is 0.304°C (95% CI: 0.244°C –0.364°C), a value that is unprecedented for more than a century. In all these periods, the warming trend estimated by Block Kriging is higher than that estimated using the other two methods. The reason for this will be explained in the discussion section. Please take notice that the CIs are calculated under the assumptions of the methods. Some of the model assumptions, such as the assumption of the 2^{nd} order spatial stationarity in Kriging, is inconsistent with the reality. The accuracies of the estimations are compared using cross validation in Section 3.4.

In order to compare the global mean trends with the results from Berkeley, NCEI, GISS, 20th Century Reanalysis 2m air temperature (20CR) (Compo et.al., 2013), and Karl et al. (2015), the results from these products are also provided in Table 1, although these results were derived using different source station datasets and methods. These results show that in the period of 1901–2010, the temperature warmed by 0.090–0.194°C/decade, as estimated from all the series listed in Table 1. For the final period of 1979–2014 the temperature warms by 0.254–0.329°C/decade, about 3 times compared with the period of 1901–2010.

In this study, the urban heating's affect on the estimation of global temperature land average for BSHADE was analyzed as well (see details in SI). The results showed that during the period of 1901 to 2010 there was an urban heating effect of 0.03°C/100 years. This is similar with the results from previous studies (Parker 2004, 2006; Wang et al., 2017).

3.3. Trend Map of Global Land Surface Temperature

Although shown as a global average, a warming trend is readily apparent—especially in recent decades—but there are significant geographical variations. Figure 2 show distribution maps of the warming trend of global land surface air temperature estimated by the BSHADE method for the periods of 1901–1950, 1951–2010, 1901–2010 and 1977–2014. The values for each grid were calculated when the data satisfied two conditions: (1) more than 70% of records are available in the period, and (2) the start and the end decades are both available. The symbol "+" implies that estimated warming trends are significant, using a 90% CI, for that grid box. White areas were not estimated because of incomplete or missing data.

Since 1901 almost all land areas have experienced warming. The greatest rates of warming occurred in mid-continental locations rather than coastal areas. This is most notable in the mid to high-latitudes of North America and the middle latitudes of interior Asia. From Figure 1, it shows that there is an apparent difference between the first and the second half of the twentieth century. The warming trend in the two periods also exhibits very distinct spatial signatures. In the early years of 20th century (1923–1950), warming is mainly evident in the mid-to-high latitude regions of Northern Hemisphere, whereas the more recent warming (1977-2014) covers all global land areas

(Figure 2).

The maps of temperature anomalies estimated by BSHADE, Block Kriging, and CAM generally correspond with each other in the recent period. However, some discrepancies are present in the early period and in the areas with sparse station coverage, such as Africa, South America, East and West Australia, and North Asia (Figures 3). This indicates that the differences in the linear trends for global land surface average temperatures in the last century or longer periods for different methods are caused mainly by data availability and bias of the observations in the early periods.

3.4. Validation of Estimation

In principle, the accuracy of an estimate is determined by the properties of the population, the way of sampling, and the method of estimation, actually the match between the three, referred to as the spatial sampling and inference trinity (Wang et al., 2012). The merits of an estimator are fulfilled only if its assumption is identical to the properties of the population and the way of sampling. In this study, the population is both spatially autocorrelated (see semivarigram) and spatially heterogeneous, and the sample (meteorological stations) is highly biased (vector *B*) in remote areas and in early years. Therefore, we chose to use BSHADE, a method which takes into account both the properties of a population and biased sample to make a BLUE estimate.

Though the theoretical confidence intervals can be estimated, they depend upon the assumptions of the models. The theoretical merits of BSHADE are validated by empirical tests. A sparse network of stations was selected for analysis in each year between 1961 and 1990. The stations were chosen to match the reduced spatial coverage of stations in 1880, but the temperatures were those observed during the 1961-1990 period. The global average mean temperature for each year was computed from the sparse network and then compared with the global means computed by CAM using the full network of stations from 1961-1990. In recent decades, when there was the largest number of stations, the estimated values from the different methods are highly consistent with each other. The absolute errors in each year for 1961–1990 are calculated by the difference of the estimated and the true values (see Figure 4).

From Figure 4, the absolute errors from BSHADE, Block Kriging, and CAM were 0.16°C, 0.18°C, and 0.18°C, respectively. In order to compare the results within the same domain, the polar areas (e.g. Greenland) were not included in the Block Kriging validation. This demonstrates that the estimates of BSHADE have the smallest absolute errors compared to the other methods, which implies that, in the early years having sparse and unevenly distributed stations, the results estimated by BSHADE in this study will have the highest accuracy.

4. Conclusion and Discussion

In this study, the spatial distribution maps of global mean surface air temperature anomalies for each year from 1880 to 2014 were created using the BSHADE approach. These maps have greater spatial coverage and less uncertainty compared to existing studies. Validation was performed using a few selected stations in 1961–1990 with the same location as stations in 1880. This showed a smaller estimation error using BSHADE compared to other common methods.

The reliabilities of regional mean temperature estimation (Li et al., 2010; Peterson, 2003; Rohde et al., 2013) are determined by the combination of real land surface air temperature field, the configuration of meteorological stations, and the estimators employed, known as the spatial sampling and statistical trinity (Wang et al., 2012; Cao et al., 2013; Ge et al., 2013; Hansen et al., 2006; Jones et al., 2008; Lawrimore et al., 2011; Peterson et al., 1998; Yan et al., 2010). The discrepancy between global temperature dynamics estimated by different methods can be understood by the spatial sampling and statistical trinity.

Sparseness of stations is an important uncertainty source in global or regional mean temperature estimation. Meteorological stations are sparse and have uneven coverage in some periods and in some areas, i.e., the sample is biased to population, the histogram of the sample is different from that of the population). This occurs when the population is spatially stratified heterogeneity (Wang et al., 2016), and some strata have no sample. In this case, the sample should not be regarded as randomly drawn from a population, as is usually assumed in statistics. Thus, the mathematical expectation of the mean value of the stations' records, under the assumption of the 1st order stationary population, is not equal to the true value across the whole region. The real regional annual temperature anomalies cannot be directly represented by the samples under the assumption of random sampling. The situation is worsened in early years, especially before the end of 19th century, compared to recent years. For example, in the 1880s, existing stations were mainly located in western Europe and the northeast coasts of the USA. Although there are numerous stations available in recent years, they are uneven and sparse in some regions. For example, in the Asian continent, stations are mainly located in regions with high population density, while the mountains or plateaus.

In this study, the warming trend estimated by Block Kriging is higher than the other two methods. One of the possible reasons is that the Block Kriging estimation had more coverage than the other methods, especially in polar areas (e.g., Greenland) where the warming has been the most intense. The other reason is for Block Kriging's higher estimation is the sparse and biased station distributions in the years of the late 19th century in Africa and South America. In these areas, the mean values estimated by Block Kriging were lower than those estimated by BSHADE for the period, which results in the higher linear trends from Block Kriging. However, Block Kriging's linear trend has more uncertainty; the validation in the preceding section shows that the mean values estimated by Block Kriging in the early period have higher errors than those from BSHADE. The situation can be avoided in BSHADE due to its potential to remedy the biased sample by the value of the parameter *b*.

There is discrepancy between the CAM results and the other methods. For example, in 1880, Australia showed strong warm anomalies with CAM in the southeast of the continent, while the BSHADE method showed slight anomalies. However, there is an overlap of their error bars, where the 95% CI of CAM and BSHADE were [-0.055, 3.35], [0.25, 0.63] respectively. One of the reasons for the discrepancy is that only local stations within a box of 5° latitude by 5° longitude were used in the estimation of average land surface air temperature anomaly in each grid. Meanwhile, spatial correlation information was not used in CAM.

Besides comparing the results from the traditional methods and BSHADE, we also compared the results from BSHADE with reanalysis data and other widely used datasets. Compo et.al. (2013) have presented the linear trend of 20CR and eight different near-global datasets constructed from land surface observations. The linear trend of spatial patterns estimated by BSHADE over the 1901–2010 and 1951–2010 periods correspond with the eight datasets (see Figures 3, S2, and S3 in the 2013 paper by Compo et.al.). The linear trend of spatial patterns between BSHADE and 20CR in the above two periods also have the same general agreement with differences in local areas such as Argentina, eastern Brazil and the midwestern United States, which may be induced by some uncertainty of 20CR caused by factors such as land use and land cover, pressure observations, and so on. Detailed regional analyses and trends between the various methods and how the improved coverage affects regional means and trends could be conducted but are outside of the scope of this paper.

This paper provides a new estimation of global land surface air temperature since 1880 with greater spatial coverage and lower uncertainty. In this study, we took the mean values of spatial correlation matrix C in Kriging and BSHADE and sample bias vector B in BSHADE. The theories behind the parameters deserve further investigation in future studies. Although BSHADE has advantages compared with traditional methods, there is potential to improve the method’s parameterizations in the future by information fusion, such as using more data sources in the method, such as tree ring data.

Acknowledgments: This study was supported by CAS (XDA05090102), MOST (2016YFC1302504), NSFC (41531179; 41421001), and CASPIFI (2015DE016) grants.

References

Bindoff NL, Stott PA, AchutaRao KM, Allen MR, Gillett N, Gutzler D, Hansingo K, Hegerl G, Hu Y, Jain S, Mokhov II, Overland J, Perlwitz J, Sebbari R, Zhang X, 2013: Detection and Attribution of Climate Change: from Global to Regional. 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* by Stocker, TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Brohan P, Kennedy J J, Harris I, Tett SFB, and Jones PD, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J Geophys Res-Atmos*, **111**.

Cao LJ, Zhao P, Yan ZW, Jones P, Zhu YN, Yu Y, Tang GL, 2013: Instrumental temperature series in eastern and central China back to the nineteenth century. *J Geophys Res-Atmos*, **118,** 8197-8207.

Cressie N, 1993: Statistics for spatial data: Wiley series in probability and statistics. *Wiley-Interscience New York*, **15,** 16.

Ge Q, Wang F, Luterbacher J, 2013: Improved estimation of average warming trend of China from 1951–2010 based on satellite observed land-use data. *Climatic Change*, **121,** 365-379.

Gething PW, Smith DL, Patil AP, Tatem AJ, Snow RW, Hay SI, 2010: Climate change and the global malaria recession. *Nature*, **465,** 342-U394.

Compo GP, Sardeshmukh PD, Whitaker JS, Brohan P, Jones PD, McColl C, 2013: Independent confirmation of global land warming without the use of station temperatures. *Geophysical Research Letters*, **40**, 3170-3174.

Goovaerts P, 1997: *Geostatistics for natural resources evaluation.* Oxford University Press.

Gujarati DN, 2003: *Basic Econometrics. 4th.* New York: McGraw-Hill.

Hansen J, Ruedy R, Sato M, Lo K, 2010: Global Surface Temperature Change. *Reviews of Geophysics*, **48,** 29.

Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M, 2006: Global temperature change. *P Natl Acad Sci USA*, **103,** 14288-14293.

Hartmann DL, Klein Tank AMG, Rusticucci M, Alexander LV, Brönnimann S, Charabi Y, Dentener FJ, Dlugokencky EJ, Easterling DR, Kaplan A, Soden BJ, Thorne PW, Wild M, Zhai PM, , 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*, Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds.). Ed., Cambridge University Press.

Hu MG, Wang JF, Zhao Y, Jia L, 2013: A B-SHADE based best linear unbiased estimation tool for biased samples. *Environmental Modelling & Software* **48**, 93-97.

Isaaks EH, Srivastava RM, 1989: *Applied Geostatistics.* Oxford University Press.

Jones PD. 1994, Hemispheric surface air-temperature variations - a reanalysis and an update to 1993, *J Climate*, **7**(11), 1794-1802.

Jones PD, 2016: The reliability of global and hemispheric surface temperature records. *Adv Atmos Sci*, **33,** 269-282.

Jones PD, Lister DH, Li Q, 2008: Urbanization effects in large-scale temperature records, with an emphasis on China. *Journal of Geophysical Research*, **113**.

Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M, Morice CP, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J Geophys Res-Atmos*, **117**.

Jones PD, New M, Parker DE, Martin S, Rigor IG, 1999: Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, **37,** 173-199.

Jones PD, Wigley TML, 2010: Estimation of global temperature trends: what's important and what isn't. *Climatic Change*, **100,** 59-69.

Karl TR, Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348,** 1469-1472.

Lawrimore JH, Menne MJ, Gleason BE, Williams CN, Wuertz DB, Vose RS, Rennie J, 2011: An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J Geophys Res-Atmos*, **116**.

Li Q, Dong W, Li W, Gao X, Jones PD, Kennedy J, Parker D, 2010: Assessment of the uncertainties in temperature change in China during the last century. *Chinese Science Bulletin*, **55,** 1974-1982.

McMichael AJ, Woodruff RE, Hales S, 2006: Climate change and human health: present and future risks. *Lancet*, **367,** 859-869.

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

Parker DE, 2006: A demonstration that large-scale warming is not urban. *J Climate*, **19,** 2882-2895.

Patz JA, Campbell-Lendrum D, Holloway T, Foley JA, 2005: Impact of regional climate change on human health. *Nature*, **438,** 310-317.

Peterson TC, 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. *J Climate*, **16,** 2941-2959.

Peterson TC, Karl TR, Jamason PF, Knight R, Easterling DR, 1998: First difference method: Maximizing station density for the calculation of long-term global temperature change. *J Geophys Res-Atmos*, **103,** 25967-25974.

Pielke R, Coauthors, 2007: Documentation of Uncertainties and Biases Associated with Surface Temperature Measurement Sites for Climate Change Assessment. *B Am Meteorol Soc*, **88,** 913-928.

Qin D, 2014: Climate change science and sustainable development. *Progress in Geography*, **33,** 874-883.

Rahmstorf S, 2007: A semi-empirical approach to projecting future sea-level rise. *Science*, **315,** 368-370.

Rohde R, Muller RA, Jacobsen R, Muller E, Perlmutter S, 2013: A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinfor Geostat*.

Santer BD, Coauthors, 1996: A search for human influences on the thermal structure of the atmosphere. *Nature*, **382,** 39-46.

Stott PA, Tett SFB, Jones GS, Allen MR, Mitchell JFB, Jenkins GJ, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290,** 2133-2137.

Tencer B, Rusticucci M, P Jones, Lister D, 2011: A southeastern south American daily gridded dataset of observed surface minimum and maximum temperature for 1961-2000. *B Am Meteorol Soc*, **92,** 1339-1346.

Vose RS, Wuertz D, Peterson TC, Jones PD, 2005: An intercomparison of trends in surface air temperature analyses at the global, hemispheric, and grid-box scale. *Geophysical Research Letters*, **32,** 4.

Walther GR, Coauthors, 2002: Ecological responses to recent climate change. *Nature*, **416,** 389-395.

Wang J, Stein A, Gao B, Ge Y, 2012. A review of spatial sampling. *Spatial Statistics*, **2,** 1-14.

Wang J, Xu C, Hu M, Li Q, Yan Z, Zhao P, Jones P, 2014: A new estimate of the China temperature anomaly series and uncertainty assessment in 1900–2006. *Journal of Geophysical Research: Atmospheres*, **119,** 1-9.

Wang J, Coauthors, 2011: Area Disease Estimation Based on Sentinel Hospital Records. *Plos One*, **6,** 1-8.

Wang J, Zhang T, Fu B, 2016. A measure of spatial stratified heterogeneity. *Ecological Indicators*, **67,** 250-256.

Wang J, Tett SFB, Yan Z, 2017: Correcting urban bias in large-scale temperature records in China, 1980-2009. *Geophysical Research Letters*, **44,** 401-408.

Xu C, Wang J, Hu M, Li Q, 2014: Estimation of Uncertainty in Temperature Observations Made at Meteorological Stations Using a Probabilistic Spatiotemporal Approach. *J Appl Meteorol Clim*, **53,** 1538-1546.

Xu C, Wang J, Hu M, Li Q, 2013: Interpolation of missing temperature data at

meteorological stations using P-BSHADE. *J Climate*, **26,** 7452-7463.
Yan Z, Li Z, Jones P, 2010: Effects of site change and urbanisation in the Beijing temperature series 1977–2006. *International Journal of Climatology*, **30,** 1226-1234.

List of Figures

TABLE 1: Trend estimates and 95% CIs (°C/decade) during different periods.

Figure 1. Annual global land surface air temperature anomaly time series in 1880–2014 relative to 1961–1990 estimated by BSHADE, CAM, and Block Kriging, respectively

Figure 2. Trends in global land surface temperature estimated by BSHADE method for periods of 1901–2010, 1901–1950, 1951–2010 and 1951–20101977–2014..

Figure 3. Validation of the accuracy of mean temperature anomalies estimated by BSHADE, CAM, and Block Kriging using the station locations available on 1880

Figure 4 Maps of differences of average temperature anomaly in the periods 1880-1900, 1923-1950 and 1977-2014 between BlockKriging, CAM and BSHADE, respectively.

Table 1. Trend estimates and 95% confidence intervals (°C/decade) during different periods.

	1901–1950	1880–2010	1901–2010	1951–2010	1979–2014
BSHADE	0.118±0.032	0.096±0.021	0.109±0.028	0.223±0.049	0.304±0.060
CAM	0.097±0.034	0.092±0.020	0.104±0.026	0.207±0.048	0.278±0.052
Block Kriging	0.143±0.039	0.108±0.021	0.115±0.029	0.229±0.052	0.329±0.061
Berkeley (*Rohde et al.*, 2013)	0.124±0.040	0.100±0.016	0.107±0.020	0.185±0.039	0.255±0.053
[*]NCEI (*Hartmann et al.*, 2013; *Lawrimore et al.*, 2011)	0.100± 0.033	0.094±0.016	0.107±0.020	0.197±0.031	0.273±0.047
[*]GISS (*Hansen et al.*, 2010; *Hartmann et al.*, 2013)	0.098±0.032	0.095±0.015	0.099±0.020	0.188±0.032	0.254±0.049
20th Century Reanalysis (*Compo et.al., 2013*)	/	/	0.090	[#]0.134	/
Karl *et al.* (*2015*)	/	[&]0.106± 0.017	[$]0.194±0.031	/	/

Note: Berkeley used a different dataset compared with the three methods in this study. The symbol "*" indicates these trends were calculated for the periods of 1901–1950, 1880–2012, 1901–2012, 1951–2012, 1979–2012 in the cited sources. The symbol "#" indicates the trend was calculated for the period 1952–2010 in the cited sources. The symbol "&" indicates the trend was calculated for the period 1880–2014 in the cited sources. The symbol "$" indicates the trend was calculated for the period 1951–2012 in the cited sources. The symbol "/" indicates no data available.

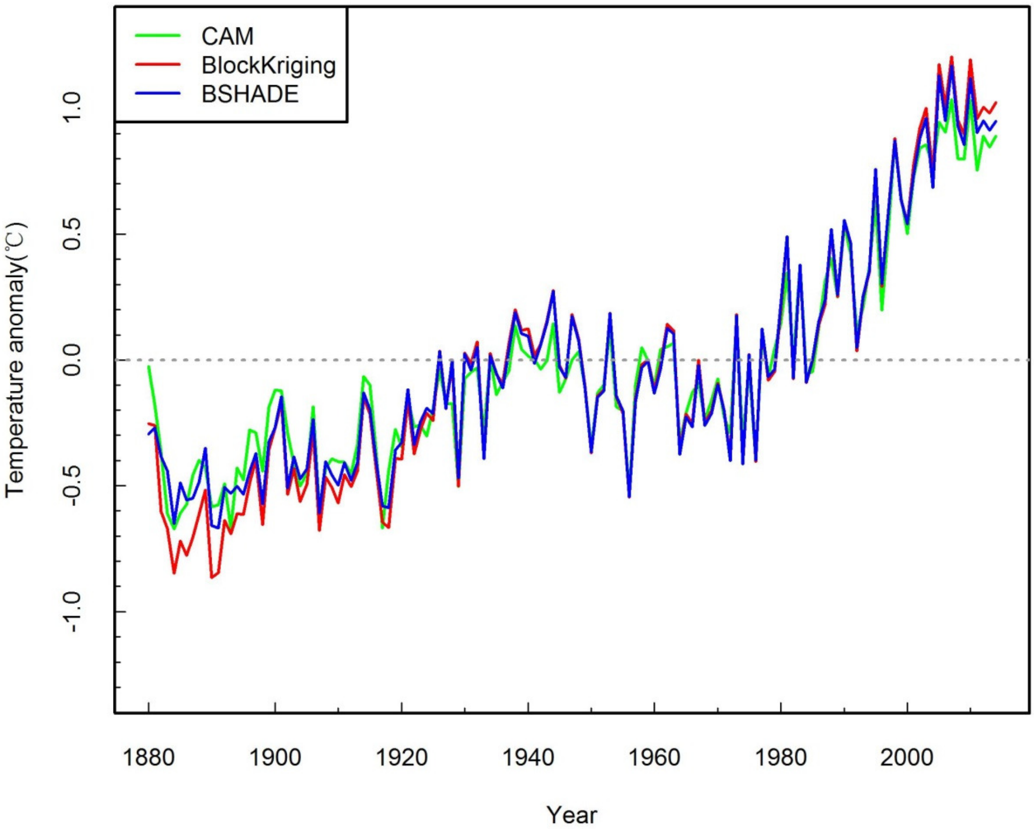

Figure 1. Annual global land surface air temperature anomaly time series in 1880–2014 relative to 1961–1990 estimated by BSHADE, CAM, and Block Kriging, respectively.

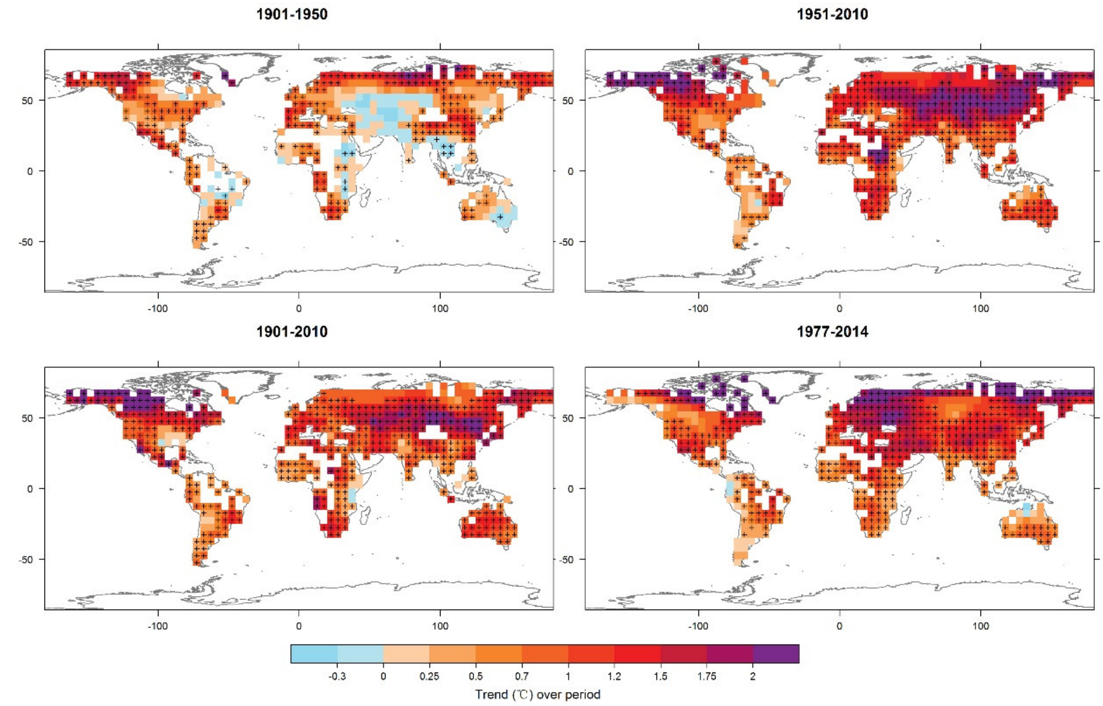

Figure 2. Trends in global land surface temperature estimated by BSHADE method for periods of 1901–2010, 1901–1950, 1951–2010 and1977–2014.

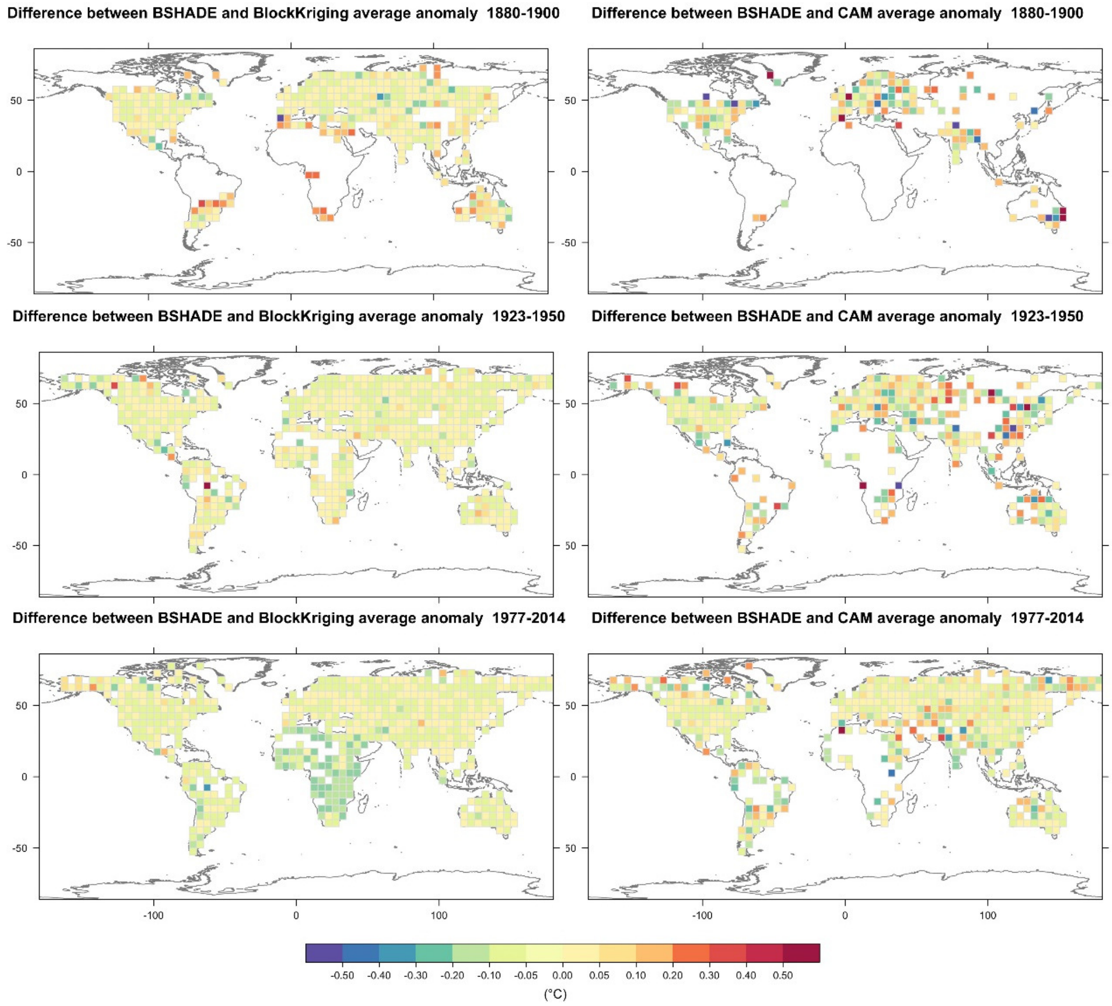

Figure 3 Maps of differences of average temperature anomaly in the periods 1880-1900, 1923-1950 and 1977-2014 between BlockKriging, CAM and BSHADE, respectively.

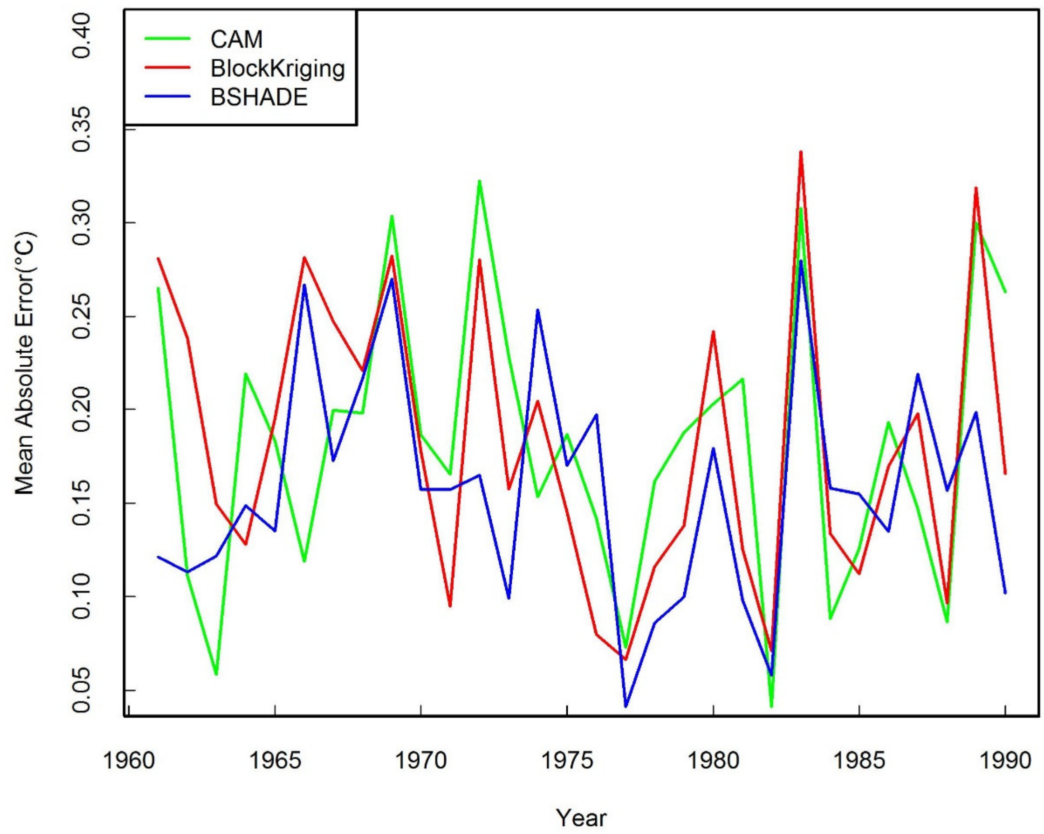

Figure 4. Validation of the accuracy of mean temperature anomalies estimated by BSHADE, CAM and Block Kriging using the station locations available on 1880.