Climatic Change (2010) 100:59–69
DOI 10.1007/s10584-010-9836-3

Estimation of global temperature trends: what's important and what isn't

Phil D. Jones · T. M. L. Wigley

Received: 28 January 2010 / Accepted: 11 March 2010 / Published online: 19 May 2010
© Springer Science+Business Media B.V. 2010

Abstract In this essay we discuss the development of and estimation of uncertainties in the global surface temperature record. We briefly discuss the similarities in and differences between the records from the institutions that produce such series. We then consider the numerous issues that must be addressed to enable accurate estimates to be derived. We consider these in their order of importance with respect to the record: biases in the sea surface temperature data, exposure of land-based thermometers before about 1900, urbanization effects in some series, and, finally, the homogeneity of individual land-based records.

1 Introduction

Three groups routinely update gridded datasets of surface temperature and produce time series of global and hemispheric average temperatures. The three groups are: the UK Meteorological Office Hadley Centre/Climatic Research Unit (HadCRUT3; Brohan et al. 2006; http://www.cru.uea.ac.uk/cru/data/temperature/ and http://hadobs.metoffice.com/hadcrut3/) the US National Climatic Data Center (NCDC; Smith et al. 2008; http://www.ncdc.noaa.gov/oa/climate/research/anomalies/index.html) and the Goddard Institute for Space Studies (GISS; Hansen et al. 1999, 2001, 2006; http://data.giss.nasa.gov/gistemp/). One other group monitors land-based temperatures (Lugina et al. 2005) and another group monitors sea surface temperature (SST; Ishii et al. 2005). Surface temperature datasets are comprised of measurements from the land (from fixed location sites) and the ocean (from moving

P. D. Jones (✉)
Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK
e-mail: p.jones@uea.ac.uk

T. M. L. Wigley
NCAR, Boulder, CO 80307, USA

ships and buoys). Data from the two components are combined as anomalies from a base period. The base period, however, is different for the three data sets: 1961–1990 for HadCRUT3, 1901–2000 for NCDC, and 1951–1980 for GISS.

The three principal groups use much the same input data, but employ different ways of interpolating to produce gridded products. HadCRUT3 and NCDC both use a 5° by 5° latitude/longitude grid that is produced first over separate domains, land and ocean. These two gridded products have overlaps at coastlines and islands. The data are synthesized to give a global grid. The GISS data are derived by first averaging all the land station data into 80 approximately equal-area boxes, and then combining these with SST values from the marine areas. For the marine data, GISS use HadISST1 (Rayner et al. 1996) from 1880 to November 1981 and then the OIv2 data set from Reynolds et al. (2002) for December 1981 onwards. If there are no data for a given month in one of the grid boxes, the HadCRUT3 value is missing. NCDC, however, spatially infill using an eigenvector-based technique where this is judged to produce statistically reliable estimates. The GISS equal-area boxes effectively provide some infilling in data sparse areas, so only a few of the 80 are completely missing for all months. Despite these differences in the methods used to combine the basic data, the hemispheric- and global-scale time series are very similar (see the trends calculated over three different periods in Table 3.3 of Trenberth et al. 2007, and also comparisons undertaken with radiosondes and satellite estimates of lower tropospheric temperatures in Karl et al. 2006).

The purpose of this essay is to discuss the principal reasons for the similarities at large spatial scales, and then discuss the numerous issues that need to be considered to ensure reliability and to assess the accuracy of the monthly and annual estimates (both at the grid-box scale and for hemispheric and global averages).

2 Similarity and homogeneity of large area-average time series

A primary reason for the close similarity between the three independently derived data series is that the three groups use much the same raw (monthly mean) input data. While there are some differences in these input data, these are essentially random, so, when averaged over large areas, the differences tend to cancel out. Another factor is that temperature data series are highly spatially correlated. Thus, even though there are records from thousands of sites, the effective number of sites is much less than this. Estimates (using both observational data and globally complete climate model data) indicate that the effective number of independent sites at the monthly timescale is about 100 (see Jones et al. 1997). Thus, provided input data sets have at least 100 well-spaced sampling points for which the data are relatively free of non-climatic biases, even if the locations of these sites are different, they will lead to very similar large-scale area averages. For annual or decadal averages the required number of well-spaced sites can be substantially less than 100.

A similar situation exists for pressure data. Here, the correlation decay length is similar to that for temperature, so relatively few sites can give good area averages. For precipitation, however, the required number of data sites to produce reliable area averages is much larger than for temperature, as correlation decay lengths are much smaller.

The relatively small number of sites required to estimate large-scale area averages accurately means that, even for early parts of the temperature record when the data network was relatively sparse, area averages are reliable. A test of the adequacy of the evolving network of temperature data sites for deriving large area-average time series is provided by Figure 1.3 of Le Treut et al. (2007). Here, many of the series developed before 1990 are compared and shown to agree well. Even the record developed by Köppen (1873) for the Northern Hemisphere land masses is similar to averages developed today by HadCRUT3.

Parker et al. (2009) have also shown that the number of sites required to get a reliable area average is small (see their Fig. 1) by calculating global land averages using a limited set of 5° grid boxes and then with another set offset from the first by 10° of longitude and latitude. There are noticeable errors at the annual time scale prior to about 1900, but these become far less marked for decadal and longer time scale averages. Jones (1994) used a sparse, but more constant network of stations to show that the sparser networks available in the second half of the nineteenth century could reproduce the global average reliably on decadal timescales and so ensure the homogeneity of large-scale area-average time series. The strong spatial correlation of temperature is also important in paleo-temperature reconstructions from proxy data. Here, the numbers of sites is much fewer than for instrumental data, but reliable area averages can still be produced (see Jansen et al. 2007).

3 Issues considered in series adjustment and error assessment

The effective number of spatial degrees of freedom is a (albeit not the only) key parameter in estimating the statistical uncertainty in estimates of large-scale averages. In the most recent study by the HadCRUT3 group (Brohan et al. 2006), estimated uncertainties also account for uncertainties in homogeneity and bias adjustments to the basic data, as well as the effects of sparser sampling in the earlier years. Smith and Reynolds (2005) have also looked at the effects of sparser coverage in earlier years. Knowledge of these potential sources of error and their correlative structure is key information if the uncertainties in the global temperature record are to be reduced. The greatest potential for improvement will come from infilling data gaps in early years, particularly through the incorporation of more marine data (where improvements in metadata will also be important—as evidenced in the work of Thompson et al. 2008).

To summarize, there are three basic issues in the development of the gridded temperature products and global and hemispheric means: *homogeneity* of the basic raw station or marine time series; large-scale systematic *biases* that might affect large areas; and the lack of *coverage* in parts of the world, particularly before the 1950s. We will discuss these in order of importance for the large-scale averages: biases, coverage, and homogeneity of the individual site series. At the local (grid box) scale the order of importance would differ: coverage, then homogeneity, and finally bias. Additionally, the components of the uncertainty are independent of each other, so may be combined in quadrature (Brohan et al. 2006).

It is important to note, however, that these problems apply to the original (raw) input data. For the data that are used to produce standard area-average time series, corrections have been applied to remove, as far as possible, these potential sources

of error. The fact that three organizations have made such corrections independently is a testimony to the robustness and accuracy of the resulting homogenized data.

3.1 Biases

Biases are homogeneity issues that affect large portions of the dataset. They may be smaller in magnitude than the effects of site moves and other factors (see later), but they can be important if they are pervasive across significant fractions of the basic input data. We will discuss the three most important sources of bias, again in order of importance: methods of measuring SST; exposure issues with temperature recorded before the development of louvred (Stevenson) screens; and the time-varying effects of urbanization due to the growth of cities, and the related changes in land-use in the environments around the sites.

3.1.1 SST measurements

Any issue of homogeneity or bias in measuring SSTs could have a serious impact on global temperature estimates because almost 70% of the planet's surface is ocean. The history of marine measurements goes back to the eighteenth century and the basic meteorological measurements (not just SST, but air temperature, pressure, wind direction and speed, etc.) were entered into ship logbooks. Even before instruments, ships kept logbooks as these were essential for navigation.

The first SST measurements were taken using wooden buckets tied to a rope. A sample of sea water was hauled onto the ship's deck and the temperature measured. In the earliest years these measurements came mostly from voyages of exploration. By the early nineteenth century the whole array of measurements including SST became a routine part of life at sea (Maury 1855). The advent of steamships in the nineteenth century led to ships increasing their speed and deck height above the sea surface. These changes probably led to more wooden buckets being broken on the sides of ships, so, by the late nineteenth century, many SST measurements were made with canvas buckets, which were more flexible and much cheaper to construct. The use of canvas buckets continued on most merchant and naval vessels up to about 1940.

When SST data were first examined in detail by climatologists, it was soon realized that the method of measurement might influence the results. Between the wars there were a number of comparisons made of different measurement techniques on research vessels and on cruises: i.e., comparisons of different types of bucket, as well as with thermometers fitted on the engine cooling-water intake pipes of ships. More comparisons were undertaken in the 1970s, and it was at this time that an extra code was added into both logbooks and transmitted data to indicate how the measurement was made.

The different thermal properties of the buckets: wooden, canvas, and also, more recently, rubber, mean that to use these data it is necessary to determine their relative biases, and to develop a history of which types of bucket were used. Bucket-type biases have been comprehensively assessed by Farmer et al. (1989) and Folland and Parker (1995).

With regard to changing instrumentation, a basic assumption is that wooden buckets dominated in the nineteenth century, canvas buckets from 1900 to 1941,

and engine intake measurements from then on. These were not, of course, abrupt changes, but spatially variable transitions over time, so correcting for these changes is not a simple task. The importance of these measurement technique biases is evident from the average size of the adjustment across the world's oceans – canvas bucket measurements need to be raised by 0.4°C between 1900 and 1941 compared to engine intakes. The main cause here is the evaporative cooling of the sea water between the times of sampling and reading of the thermometer. The Farmer et al. (1989) and Folland and Parker (1995) procedures provide corrections that can be applied for each part of the ocean with different values during the seasonal cycle. Temperatures measured in wooden buckets before the 1890s must also be raised relative to engine intake measurements, but by smaller amounts than for canvas buckets. Uncertainties in these adjustments are also incorporated in the overall error range accompanying each grid-box or larger-scale value. These uncertainties are dependent on the size of the adjustments, so are larger for the canvas as opposed to wooden buckets.

There are still some residual issues regarding measurement method changes. First, recent work has suggested that SST data for the period 1945 to 1960 are too cold (Thompson et al. 2008). It turns out that many of these measurements were taken by British ships, which seem to have continued their canvas bucket method of sampling. Second, since about the late 1970s there have been major changes to the marine observing system, with the principal aim of improving weather forecasts and seasonal climate predictions. Satellites began to measure SSTs at this time, and fixed buoys have been deployed in the equatorial Pacific to help ENSO forecasts. Further, since the late 1990s a large number of drifting buoys have been regularly deployed across the world's oceans. Little consideration was given to the continuity of measurements at the time these instrumental changes and additions were made. As a consequence, when detailed comparisons have been made, some inhomogeneities have been discovered. Most of these are relatively small and affect only parts of the data, but it is likely that further adjustments will be required to improve overall homogeneity. For example, it seems that the new drifters estimate SST values slightly lower than ships by 0.1 to 0.2°C, so their use might introduce a spurious cooling in the record.

It may come as a surprise that such problems occur. The Global Climate Observing System (GCOS) has set up a number of Monitoring Principles, one of which is for parallel measurements to be made when changes are made to the site, method of measurement, instruments, etc. Despite being agreed by all operators, however, parallel measurements are rarely undertaken. Instead, problems associated with such changes are generally sorted out after the fact, once sufficient observations become available to estimate the required adjustments.

3.1.2 Exposure of thermometers

The problem of thermometer exposure, primarily to avoid the direct impact of sunshine on the instruments, was solved during the mid-to-late nineteenth century with the invention of screens. Many different variants were tried, but it wasn't until the development of white-louvred screens by Stevenson around 1870 that consistent exposures were established. (Louvred screens have had different names around the world, e.g. "cotton region shelters" in the United States). Prior to these screens, thermometers were generally positioned on north wall locations of buildings in the

Northern Hemisphere, so as not to be in direct sunlight. Even so, they would have received some sun exposure during the early morning and late evenings during the summer, particularly the further north the location.

In general, assessment of the long-term homogeneity of temperature time series is facilitated using detailed station histories (referred to by climatologists as metadata) – but statistical comparisons of data from sites that are not too distant from each other can often identify inhomogeneities even when metadata are not available. With regard to metadata, the important questions are: when was the new screen installed; and what exposure was used previously? An additional question is whether any parallel measurements were made when the change occurred. The answer to the latter question is generally "never". Even if parallel measurements were made they have almost invariably been lost.

Even when metadata are available (which, unfortunately is rare), it is necessary to compare records from a number of sites in order to quantify (and so correct for) inhomogeneities. The accuracy of these corrections in early years in Europe has been questioned by some authors, particularly with measurements made during the summer (see Moberg et al. 2003, and Jones et al. 2003). The discussion here has centred on two aspects of long temperature series: (1) the warmth of summer temperatures in the pre-1850 period; and (2) the lower long-term warming rates in summer compared to the other three seasons.

Moberg et al. (2003) showed that the long-term warming in Swedish winters is consistent with changes in the circulation and the warming of SSTs in the North Atlantic, and so is likely to be reliable. The reliability of the summer data is harder to determine, however. The circulation and local SSTs influence summer temperatures considerably less, and the principal determining factor here is the radiation received (which, in principle, might be estimated using cloudiness data). However, the long-term homogeneity of cloudiness estimates from early observers is beset with even more problems than for temperatures, and so cannot help to assess the reliability of long summer temperature series.

It is important to note that these issues may be important regionally and for the early parts of individual station records, but they are of little significance for global-scale changes over the period since 1880.

Climatologists have recently begun to resolve these instrument exposure issues through the use of parallel measurements. Two examples of this type of work are studies in the Greater Alpine Region (GAR) by Böhm et al. (2010) and in Spain by Brunet et al. (2010). The former uses parallel measurements at one site in Austria, which enabled the differences between the old and modern exposure methods to be estimated and related to the directional exposure of all earlier sites in the GAR. The Spanish example rebuilt screens from nineteenth century diagrams and made modern parallel measurements, again developing correction formulae to apply to the original nineteenth century data. The results are similar in both cases – summer temperatures on average were recorded about 0.4 to 0.5°C warmer with the old, compared to the modern, exposures. These results are very similar to pioneering assessments made at Adelaide in Australia (Nicholls et al. 1996).

The impact of changes in instrument exposure depends on many factors (see Parker 1994, for a discussion) and it is difficult to generalize the above results to other areas. This issue is known to affect European data before 1870 and Australian data before 1910.

As many natural proxies (particularly trees) respond to summer temperatures, it is important to assess the scale of the problem and, if necessary, make adjustments. Temperature reconstructions from proxy data records clearly require the primary temperature data against which the proxy data are calibrated to be reliable. If the homogeneity of pre-1900 temperature records from individual sites could be improved, this could improve the reliability of such temperature reconstructions.

3.1.3 Urbanization

Station history information often shows that many sites began at locations in small towns, and that, over the last 100 years, these towns have developed into major cities. Such urban growth is likely to affect temperature records from urban sites, and warming trends from such sites are likely, on average, to be larger than if the city or town were not there. In climatology, this issue is referred to as the urbanization effect. It could potentially impact large-scale temperatures as urban areas have expanded dramatically during the twentieth century. The issue is not the urbanization effect *per se*, but whether nearby rural and urban locations show similar long-term trends. For example, city centre sites in London and Vienna are warmer than their rural counterparts, but the urban time series during the twentieth century change at exactly the same rate (see Jones et al. 2008 and Jones and Lister 2009).

There are a number of factors, however, that make the assessment of urbanization effects difficult – although, as shown below, it is likely that residual errors are small. The first factor that must be considered is that sites in urban areas are generally not in the downtown part of the city, but are more likely to be in a parkland setting or at an airport. The second is that each case probably has to be assessed individually. It is impossible to make generalizations: European cities differ from North American cities which, in turn, differ from East Asian cities.

Although numerous papers have addressed urban climates and found large differences between city centre sites and rural neighbours for individual day and night temperatures, these studies are generally not relevant to the global-scale data bases described here. This is because these comparisons generally consider only days that maximise the urban/rural difference and so are not directly relevant in the context of long-term monthly averages for typical (non-city-centre) weather stations.

Despite the difficulties in correcting for urbanization effects, there are two strong arguments that indicate that any residual urbanization effects in the standard (homogenized) temperature data sets are probably very small. The first of these is that SSTs are not affected, so the similarity of warming trends from land and marine regions argues against the effect being important. Second, we can construct data sets using only rural locations. Although this restricts coverage, because of the spatial correlations noted above, we can still use these smaller networks to derive reliable large-scale averages. When compared with results using all stations, the differences are small. In some parts of the world, however, it is difficult to choose rurally-located sites.

As noted above, many assessments of urbanization effects at the large scale have considered rural-only sites and compared these to averages based on all sites, or on urban-only sites (see. e.g. Jones et al. 1990; Parker 2004, 2006; Peterson 2003; and Peterson and Owen 2005). Differences are always small, and always an order

of magnitude smaller than any long-term warming – implying that any urbanization effect is small. Locally, however, the effects may be larger, particularly in regions of rapid urban development (e.g. China; Jones et al. 2008).

In some papers, population statistics and nightlights have been used as a measure of, and hence to provide adjustments for the effects of urbanization (see e.g. Hansen et al. 2001). Jones et al. (2008), however, have shown that the use of population statistics is not a robust method.

4 Homogeneity of individual land-based records

Individual temperature records from land sites are homogeneous if the variations in the measurements result solely from regional-scale variations (at the scale of 10° by 10° of latitude/longitude) in the weather and climate. Inhomogeneities may result from many factors, some of which (instrument exposure, urbanization) have already been discussed. In addition, individual records may be affected by changes in site location, changes in the times each day the measurements are made, changes in the method used to calculate daily- and hence monthly-mean temperatures, and changes in instrumentation. While the effects of inhomogeneities vary from site to site, occasionally all the sites within a particular country may be affected. For example, for changes in instrumentation, automation of measurements across whole countries and regions has taken place during the last 20 years (e.g. for the USA in Quayle et al. 1991). It is, however, possible to identify such changes and correct for them. In summary, all of the above effects are accounted for in producing homogeneous site records.

Once inhomogeneities are identified, the raw individual site records must be adjusted to produce homogeneous time series. Adjustment factors are determined using station histories and metadata information (where this is available). Both physically-based corrections and corrections derived from objective statistical tests (comparing temperature time series from neighbouring sites) are estimated. Where necessary, adjustment factors are then calculated and the early parts of the records are made compatible with the most recent data.

For individual site records and for small-scale averages (such as at the single grid box level), homogenization is essential. At the hemispheric and global scale, however, because adjustments of both signs occur with similar frequencies, the adjustment factors tend to cancel. While there are uncertainties in adjustments at the site level, at larger scales the effects of such uncertainties are small compared to the SST biases and exposure issues noted above. This can be easily seen in a number of recent papers (e.g. Figure 4 of Brohan et al. 2006, and Figure 6 of Menne et al. 2009). Both these Figures show histogram counts of the magnitude of adjustments, with both showing bimodal distributions with peaks for both positive and negative adjustments. The overall average of adjustments, averaged over multiple sites in a region, is essentially zero. Adjustments are still made for individual sites, of course, since these are necessary to produce the best-possible gridded data.

For some countries there may be changes to the observing system that came from the National Meteorological Service (NMS) instructions to observers, or changes that have occurred in observing practices within a country. The impacts of these on large-scale averages depend on the size of the country, but they can be very important for

national temperature time series. One of the most important examples comes from the USA, as a result of the widespread change of observation times from the late afternoon to the morning. (This lead to what is referred to as the time of observation bias, TOB; Karl et al. 1986). The effect is noticeable because morning readings tend to be slightly cooler than those taken in the late afternoon. Figure 4 in Menne et al. (2009) shows the effect of the TOB for the contiguous US average from 1900, amounting to a difference of about 0.2°C between adjusted and unadjusted data during the present decade. In other words, the TOB leads to a spurious cooling trend in the unadjusted data (which, of course, is corrected for in the homogenized data).

5 Discussion

The importance of inhomogeneities in raw surface temperature observations becomes clear when comprehensive models to estimate the uncertainties involved are developed and applied to "clean up" the data sets (e.g. Brohan et al. 2006). Factors that affect individual site records tend to be random (i.e., they can lead to positive or negative biases) and so uncertainties in any adjustments become less and less important as data are averaged over larger areas. Biases that affect multiple sites or records (such as changing measurement techniques for SSTs) can be more important, but these too are accounted for in homogenizing the data.

The impacts of sparser coverage in early decades are only important before 1880, and, even then, the impact is mostly felt in the Southern Hemisphere (Jones 1994). For the Northern Hemisphere, it is possible to derive reliable hemispheric averages from instrumental data back to about 1850. For example, Karl et al. (1994) show that global 100+ year trends become quite reliable after the 1870s based on historical sampling. Error models tell us which sources of inhomogeneity are most important for reducing uncertainties in hemispheric averages. They show that uncertainties would be most significantly reduced through the inclusion of more SST data in the nineteenth century than through adding more land station series. A number of current projects are seeking to digitize much of the British logbook material available in archives. More SST data are not only important for improving the reliability of hemispheric and global temperature series, but can help to improve infilled SST fields (i.e., where statistical methods are used to fill in data gaps and produce spatially complete data sets). Reliable infilled data sets are important for detection and attribution (D&A) work, for driving AMIP-type General Circulation Model simulations, for AOGCM validation, and for developing extended reanalyses that are ongoing or planned (Compo et al. 2006). For identifying past large-scale changes in temperature at the Earth's surface, however, the homogenized data sets currently available provide highly reliable information back into the nineteenth century and show unequivocally that the world has warmed considerably over this period.

References

Böhm R, Jones PD, Hiebl, J, Frank D, Brunetti M, Maugeri M (2010) The early instrumental warm-bias: a solution for long Central European temperature series, 1760–2007. Clim Change. doi:10.1007/s10584-009-9649-4

Brohan P, Kennedy J, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J Geophys Res 111:D12106

Brunet M, Asin J, Sigró J, Bañon M, García F, Aguilar E, Palenzuela JE, Peterson TP, Jones PD (2010) The minimization of the screen bias from ancient western Mediterranean air temperature records: an exploratory statistical analysis. Int J Climatol (conditionally accepted)

Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100 year reanalysis using only surface pressure data. Bull Am Meteorol Soc 87:175–190

Farmer G, Wigley TML, Jones PD, Salmon M (1989) Documenting and explaining recent global-mean temperature changes. Final report to the UK Natural Environment Research Council (NERC) under Contract GR3/6565

Folland CK, Parker DE (1995) Correction of instrumental biases in historical sea surface temperature data. Q J R Meteorol Soc 121:319–367

Hansen J, Ruedy R, Glascoe J, Sato M (1999) GISS analysis of surface temperature change. J Geophys Res 104:30997–31022

Hansen J et al (2001) A closer look at United States and global surface temperature change. J Geophys Res 106:23947–23963

Hansen J, Sato M, Ruedy R, Lo K, Lea DW, Medina-Elizade M (2006) Global temperature change. Proc Natl Acad Sci 103:14288–14293

Ishii M et al (2005) Objective analysis of SST and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. Int J Climatol 25:865–879

Jansen E, Overpeck J, Briffa KR, Duplessy J-C, Joos F, Masson-Delmotte V, Olago D, Otto-Bliesner B, Peltier WR, Rahmstorf S, Ramesh R, Raynaud D, Rind D, Solomina O, Villalba R, Zhang D (2007) Palaeoclimate. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis; contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp 433–497

Jones PD (1994) Hemispheric surface air temperature variations: a reanalysis and an update to 1993. J Clim 7:1794–1802

Jones PD, Lister DH (2009) The urban heat island in Central London and urban-related warming trends in Central London since 1900. Weather 64:323–327

Jones PD, Groisman PY, Coughlan M, Plummer N, Wang W-C, Karl TR (1990) Assessment of urbanization effects in time series of surface air temperature over land. Nature 347:169–172

Jones PD, Osborn TJ, Briffa KR (1997) Estimating sampling errors in large-scale temperature averages. J Climate 10:2548–2568

Jones PD, Briffa KR, Osborn TJ (2003) Changes in the Northern hemisphere annual cycle: implications for paleoclimatology? J Geophys Res 108:4588

Jones PD, Lister DH, Li Q (2008) Urbanization effects in large-scale temperature records, with an emphasis on China. J Geophys Res 113:D16122

Karl TR, Williams CN Jr, Young PJ, Wendland WM (1986) A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. J Clim Appl Meteorol 25:145–160

Karl TR, Knight RW, Christy JR (1994) Global and hemispherical trends: uncertainties related to inadequate sampling. J Climate 7:1144-1163

Karl TR, Hassol SJ, Miller CD, Murray WL (eds) (2006) Temperature trends in the lower atmosphere: steps for understanding and reconciling differences. A report by the climate change science program and subcommittee on global change research, Washington, DC, 180 pp. http://www.climatescience.gov/Library/sap/sap1-1/finalreport/default.htm

Köppen W (1873) Über mehrjährige Perioden der Witterung, insbesondere über die 11-jährige Periode der Temperatur. Zeitschrift der Österreichischen Gesellschaft für Meteorologie, Bd VIII, pp 241–248, 257–267

Le Treut H, Somerville R, Cubasch U, Ding Y, Mauritzen C, Mokssit A, Peterson T, Prather M (2007) Historical overview of climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 93–127

Lugina KM et al (2005) Monthly surface air temperature time series area-averaged over the 30-degree latitudinal belts of the globe, 1881–2004. In: Trends: a compendium of data on global change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, USA. http://cdiac.esd.ornl.gov/trends/temp/lugina/lugina.html

Maury MF (1855) Wind and current charts, 7th edn. US Navy, Philadelphia
Menne MJ, Williams CN Jr, Vose RS (2009) The United States Historic Climate Network monthly temperature data: version 2. Bull Am Meteorol Soc 90:993–1007
Moberg A, Alexandersson H, Bergström H, Jones PD (2003) Were south Swedish summer temperatures before 1860 as warm as measured? Int J Climatol 23:1495–1521
Nicholls N, Tapp R, Burrows K (1996) Historical thermometer exposures over Australia. Int J Climatol 16:705–710
Parker DE (1994) Effects of changing exposure of thermometers at land stations. Int J Climatol 14:1–31
Parker DE (2004) 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
Parker DE, Jones PD, Peterson TC, Kennedy J (2009) Comment on "Unresolved issues with the assessment of multidecadal global land surface temperature trends" by Roger A. Pielke Sr. et al. J Geophys Res 114:D05104
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, Owen TW (2005) Urban heat island assessment: metadata are important. J Climate 18:2637–2646
Quayle RG, Easterling DR, Karl TR, Hughes PY (1991) Effects of recent thermometer changes in the cooperative station network. Bull Am Meteorol Soc 72:1718–1723
Rayner NA, Horton EB, Parker DE, Folland CK, Hackett RB (1996) Global sea ice and sea surface temperature data set, 1903–1994, Hadley Centre Clim Res Tech Note, CRTN 74, Met Office Hadley Centre, UK
Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. J Climate 15:1609–1625
Smith TM, Reynolds RW (2005) A global merged land and sea surface temperature reconstruction based on historical observations (1880–1997). J Climate 18:2021–2036
Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). J Climate 21:2283–2293
Thompson DWJ, Kennedy JJ, Wallace JM, Jones PD (2008) A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. Nature 453:646–649
Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Klein Tank A, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, Soden B, Zhai P (2007) Observations: surface and atmospheric climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of Working Group 1 to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 235–336