Exposure, instrumentation, and observing practice effects on land temperature measurements

Blair Trewin*

To monitor climate change adequately and determine the extent to which anthropogenic influences are contributing to observed climate change, it is critical to have land temperature data of a high standard. In particular, it is important to have temperature data whose changes reflect changes in the climate and not changes in other circumstances under which the temperatures were taken. There are numerous factors that can affect land temperature records. Among the most common are changes in instrumentation, changes in local site condition *in situ* (through urbanization or for other reasons), site relocations, and changes in observing practices. All have the potential, if uncorrected, to have impacts on temperature records at individual locations similar to or greater than the observed century-scale global warming trend. A number of techniques exist to identify these influences and correct data to take them into account. These have been applied in various ways in climate change analyses and in major data sets used for the assessment of long-term climate change. These techniques are not perfect and numerous uncertainties remain, especially with respect to daily and sub-daily temperature data. *WIREs Clim Change* 2010 1 490–506

The observed land temperature record is a fundamental indicator of global climate change. To monitor climate change adequately and determine the extent to which anthropogenic influences are contributing to observed climate change, it is critical to have land temperature data of a high standard. In particular, it is important to have temperature data whose changes reflect changes in the climate and not changes in other circumstances under which the temperatures were taken.

There are numerous non-climatic factors that can influence land temperature measurements. Some of the most significant include changes in the location of the observation site, changes *in situ* in the nature of the land surface and/or local environment around an observation site (whether because of urbanization or for some other reasons), and changes in the practices used for taking observations.

As has been noted by a number of authors,[1] the best way to avoid inhomogeneities—that is, changes in a time series that do not reflect changes in the climate, but rather outside influences—in a temperature data set (or other climate data sets) is to keep the data set homogeneous. In practice, this has been difficult to achieve at most stations, even in the modern era when the importance of homogeneous data for detection of climate change has been recognized [as illustrated by the issues associated with the introduction of automatic weather stations (AWSs), as discussed in section *Instrumentation*]. In this context, it is important to note that almost all temperature data used for the analysis of climate change come from observations which were originally established for other reasons, such as operational weather forecasting or the support of aviation, and that very few stations have been established for the specific purpose of monitoring climate change. The priorities for establishing a site for monitoring long-term temperature changes may conflict with those other needs; for example, a site whose primary purpose is to support aviation is likely to be established in the location most representative of the airport runways, regardless of any other site considerations, and a site established to support marine forecasting is likely to be established in an exposed coastal or island location.

*Correspondence to: b.trewin@bom.gov.au

National Climate Centre, Australian Bureau of Meteorology, Melbourne, Victoria 3001, Australia

DOI: 10.1002/wcc.46

Only in recent years has a station network, the US Climate Reference Network (CRN),[2] been established for the specific purpose of providing a stable platform for climate change monitoring, and it will be many years before that network produces meaningful long-term results.

WHAT STANDARDS EXIST FOR TEMPERATURE MEASUREMENTS?

What International and National Standards Exist?

The World Meteorological Organization (WMO) issues standards for the selection of an observation site.[3,4] These state that observations should 'be representative of an area in accordance with its application'; for example, synoptic observations should be in a site broadly representative of the wider region and not in a distinctive local environment such as a frost hollow.

Specific recommendations of the WMO include:

- An instrument enclosure at least 10×7 m, with a surface covered with short grass or, if grass does not grow, a surface representative of the locality.
- Not on steeply sloping ground or in a hollow.
- Well away from obstructions.

No specific standards are established for thermometer screens, with the guidelines providing for both naturally ventilated and aspirated screens, and for both wood and plastic (or similar material). The standard height for temperature measurements is between 1.25 and 2 m above ground.

Various national standards exist (two examples in the English language which are publicly available are those of the Australian Bureau of Meteorology[5] and the Meteorological Service of Canada[6]). These generally use the WMO standards as a basis but are more specific (e.g., specifying the exact type of thermometer screen to be used—in the case of Australia, a wooden Stevenson screen with its base 1.1 m above ground level).

As the Australian Bureau of Meteorology standards note, 'it is inevitable that some localities for which observations are essential may not have sites that fully conform to the criteria...'. In practice, most observation networks include numerous sites which do not fully conform to standards in some way or other. If one is interested in long-term change, a station's absolute conformance with standards is less important than its consistency over the long term; a stable, but unrepresentative, hilltop site which has been in the same position for 100 years is likely to be more useful for climate change analyses than one on flat, open ground that has moved several times over the century.

Which Temperature Measurements Are of Interest?

A number of different temperature measurements are of interest in studies of climate change. The most analyzed variable is mean annual or monthly temperature, defined using either fixed-hour observations or daily maximum and minimum temperatures (an issue discussed more fully in section *Observing Practice Changes*). Mean daily maximum and minimum temperatures, and the diurnal temperature range derived from these, have also been the subject of numerous analyses.

A more recent area of interest[7] has been that of temperature extremes of various kinds. Most of these extremes (e.g., number of days with temperatures above the 90th or below the 10th percentile, or number of days with frost) are derived in some form from daily maximum and minimum temperatures, making daily data important for this application. A number of other indices (e.g., growing season length) have also been based on daily data. There have been, to date, very few studies of changes in other temperature variables, such as temperatures at fixed hours, although hourly data were used as a basis for a study of changes in seasonal mean temperature in Canada.[8]

MAJOR FACTORS LEADING TO INHOMOGENEITIES IN TEMPERATURE RECORDS

Instrumentation

Changes in instrumentation have had an important influence on long-term temperature records. For temperature, the fundamental observing technology – of liquid-in-glass, manually read thermometers and self-registering maximum and minimum thermometers – remained largely unchanged' from the 19th century into the 1990s. More significant have been changes in the way that thermometers have been exposed to the atmosphere and sheltered from direct or indirect solar radiation.

In general terms, there have been two major changes in instrumentation that have affected much of the global observing network. The first was the introduction of standardization into the observing

FIGURE 1 | A wall-mounted Kingston screen at Parry Sound, Canada prior to its removal in 1935.

network in the late 19th and early 20th centuries. Prior to the late 19th century, few standards existed for instrument shelters (or lack thereof), and wall-mounted thermometers were common, as were many different types of shelters (Figure 1). A transitional phase then followed, with free-standing louvered shelters (the Stevenson screen or minor variations on it, such as the US Cotton Region Shelter) becoming a standard in most countries by the early 20th century, sometimes as the original standard, sometimes after a period of another locally standardized shelter such as the Glaisher stand. Some tropical countries (especially British colonies) retained a large, thatched-roof 'thermometer house' through the early part of the 20th century. By the 1920s, though, the Stevenson screen was in almost universal use, except for a handful of locations.[9]

The second major change has been the introduction of AWSs. While these have existed in some form for many years, it was only from the late 1980s that they started to make a widespread appearance at long-term stations used for climate change analysis. The way in which this change has been implemented has varied from country to country. Some countries have exposed AWS temperature probes in the same screens as their manual predecessors; others have introduced a new instrument shelter (either a screen of similar design to the Stevenson screen but using synthetic materials rather than wood, or a completely new screen design) at the same time as they introduced automatic sensors.

A comprehensive review of the various instrument shelter changes of the late 19th and early 20th century was carried out by Parker.[10] He found that despite substantial discrepancies at individual sites, there was little overall bias in mean temperatures over land from this cause in most areas since 1900, although in tropical areas there was a warm bias in the order of 0.2°C into the early 20th century. Prior to that time, biases tended to be screen- and location-specific, although numerous pre-Stevenson exposures were inadequately sheltered from solar radiation (more often indirectly, e.g., from rereflection from the ground, than directly) and thus tended to have a warm bias on clear days, especially in summer. This is also consistent with the results obtained for typical pre-1870 thermometer exposures in central Europe by Böhm et al.,[11] and for the Montsouri (French) stand by Dettwiller[12] and has been reinforced by other more recent studies,[13–15] either using historic data or modern-day replications of historic instrument exposures.

In particular, a number of studies[16,17] found that the Glaisher stand, in widespread use in English-speaking countries prior to Stevenson screen introduction, had a warm bias in maximum temperatures, ranging from 0.2 to 0.6°C in annual means and reaching up to 1.0°C in mean summer maximum temperatures and 2–3°C on some individual hot days. Minimum temperatures tended to have a cool bias of 0.2–0.3°C all year. These results were based on, among others, a 60-year set of parallel observations at Adelaide, Australia (Figure 2). A warm bias in maximum temperatures, particularly on sunny days and/or in summer, was also common to numerous other pre-Stevenson exposures.

The effect of the more recent change toward AWSs has been more mixed. Many early AWSs, up to and including the 1980s, had substantial biases relative to liquid-in-glass thermometers in Stevenson screens.[18] These were also found in one of the earliest major changeovers of a national-scale network to automated sensors, the late 1980s introduction of automated sensors in smaller plastic screens to the US cooperative station network. Quayle et al.[19] found that this resulted in a substantially depressed diurnal range, with an estimated bias of −0.4°C for maximum temperatures and +0.3°C for minimum temperatures, while a later study by Hubbard and Lin[20] found that

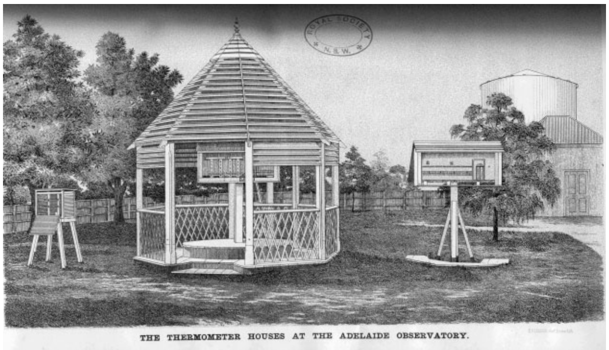

FIGURE 2 | The long-running comparison at the Adelaide Observatory, with a Stevenson screen (left), an octagonal 'thermometer house' (middle), and a Glaisher stand (right).

biases were station-specific and ranged up to ±1°C at some locations.

These biases have been substantially reduced in more recent generations of AWSs introduced by various countries over the last 20 years. A number of studies in which AWS sensors were exposed in a Stevenson screen alongside manual thermometers[21,22] have found virtually no bias in mean temperature, but a slight increase (in the order of 0.1°C) in diurnal temperature range, most likely as a result of the faster response time of automatic sensors (thus capturing more extreme fluctuations near the times of maximum and minimum temperatures). Field comparisons between wooden Stevenson screens and various other screen types introduced at the same time as AWSs, such as plastic Stevenson-type screens[23] and round multiplate ('Beehive') screens,[24–26] or between different types of 'new' screens,[27] have also mostly found minimal differences. A separate issue noted by Guttman and Baker[28] and Milewska and Hogg[29] is that in some cases replacements of a manual station by an automatic one, or other changes in instrument type, are accompanied by a site relocation (especially at airports, where a central-airfield location that is impractical for manual observations becomes feasible) and that the site relocation often has a much greater impact on temperatures than the instrument change itself.

Changes in Local Site Conditions *In Situ*

A major potential influence on observed temperatures is the state of the local environment in the vicinity of the observation site. The best known manifestation of this is the effect of urbanization on temperature records, which is well established as having a warming influence on night-time temperatures.

The impact of urbanization on estimates of observed climate change is described in Parker[30] and will not be covered in depth here. In summary, the urbanization-related uncertainty in estimates of global land temperature change is estimated at approximately 0.006°C/decade, an order of magnitude (or more) smaller than the observed temperature trend, although the local effect of urbanization can be far greater in specific locations.

A related issue is that of local conditions around an observation site. While observation standards in most places (see section *What Standards Exist for Temperature Measurements?*) state that instruments should be over short grass or a natural ground surface, and that there should be a surrounding buffer zone (e.g., Australian standards[5] state that any hard surface should be at least distance 5 *w* from the screen where *w* is that surface's width), such standards are often not observed in practice. Building in the vicinity of an observation site can be an issue even in very small towns or villages which would not normally be thought of as sufficiently large to generate a significant urban heat island (Figure 3). As is the case for many other influences on observed temperature, a site's absolute conformance with standards is less important for long-term data homogeneity than its consistency over time; a building which is too close to an observation site, but has been there unchanged for 100 years, is less of a problem than a new bitumen car park nearby would be.

Other changes in land use or land cover in the vicinity of an observation site have also been found to have an impact on observed temperatures. One notable example is that of irrigation. A number of studies[31–33] have found that the introduction of

FIGURE 3 | An observation site at Yunta, Australia (32°34′S, 139°34′E) in 1989.

irrigated agriculture in the vicinity of an observation site can have a substantial cooling impact on maximum temperatures during the growing season. Roy et al.[33] and Mahmoud et al.[32] find an average cooling impact in the order of a few tenths of a degree in irrigated regions of India and the north-central United States respectively, while Lovell and Bonfils[31] obtain much more dramatic results for the Central Valley of California, finding an estimated 5°C difference in summer maximum temperatures between fully irrigated and unirrigated lands. Other changes in land use or land cover appear to have a smaller impact, with Hale et al.[34] finding no significant change in maximum or minimum temperatures arising from either deforestation or reforestation in the vicinity of an observation site.

Site Relocations

There are few meteorological observation sites which have remained in exactly the same location for 100 years or more. Most long-term sites have moved at least once during their history, for a wide variety of reasons, including (but not limited to) changes in the principal purpose which an observing site served, the availability (or lack thereof) of observers, urban or other developments rendering a site unsuitable, or the availability of suitable communications to the site. Site relocations can be as small as a few meters, or as large as several kilometers (the point at which a change ceases to be regarded as a 'site relocation' and becomes the closing of one site and the opening of a new one is somewhat arbitrary).

Site relocations have the potential to have a substantial impact on temperature observations. Even a small site relocation can have a large impact on the local site environment as described above (e.g., a 20-m move may place instruments well clear of a building which previously affected observations), while more substantial relocations introduce the potential for changes due to mesoscale influences such as elevation changes, local topography, or proximity to the coast. This potential can be especially acute where sharp local climatic gradients exist. A number of studies[35–39] have found ridge–valley differences of 3–5°C in mean minimum temperatures (in some cases, with local relief of only 20–30 m), with differences of 10°C or more on some individual nights; Trewin[40] found that in such situations, differences tended to be largest on the coldest nights. Very sharp climatic gradients can also be found near coasts (especially where there is a large land–water temperature difference). A notable example is in the San Francisco Bay area, where mean July maximum temperatures near the open ocean are up to 6°C lower than those 10 km away near the western shore of the bay and up to 15°C lower than those 50 km inland.[41]

In a real observation network, it would be unusual for a site to be relocated from one 'extreme' location to another, and hence most site relocations would be expected to have a much smaller impact on temperatures than those suggested above. There are very few studies in the literature specifically describing the impact on temperatures of a site relocation (one exception is Patzert et al.,[42] who found a 0.8°C change in mean maximum temperatures from a move in the principal downtown observation site at Los Angeles); most such impacts, if they are documented at all, tend to be documented in internal and largely inaccessible reports within meteorological agencies, although a number of examples have recently been collated and published by the Joint CCl/CLIVAR/JComm[a] Expert Team on Climate Change Detection and Indices (ETCCDI).[43]

Two examples of site relocations are shown in Figures 4 and 5. Figure 4 shows an example from a site which was moved from a very built-up location, in the center of a town of population approximately 5000, to an open location at an airport outside the town area. Figure 5 shows a case where the instruments were moved from the bottom of a small hill to higher ground several meters away.

Menne and Williams,[45] Torok,[46] and Syrakova and Stefanova,[44] in papers describing the development of homogenized data sets (discussed further in section *Methods Used for Data Sets Used in Climate Change Detection*), presented information on the size adjustments made to station data (many, but not all, of which would arise from site relocations). Menne and Williams reported some adjustments as large as 4°C, although most are less than 2°C, while the largest single site-related adjustment reported in the Australian data set of Torok is 2.3°C, and for the Bulgarian set of Syrakova and Stefanova is 1.2°C. This gives an indication of the largest site relocation impacts likely to be experienced in an operational network. Vincent[47] gives a case study, as part of a paper describing a larger data set, of a site relocation which resulted in a −1.6°C change in mean maximum temperature, although no information is presented on how typical (or atypical) this change was.

A remaining question is the extent to which inhomogeneities resulting from site relocations affect estimates of global and regional temperatures. Easterling and Peterson,[48] among others, argue that on a global or continental scale, such changes largely cancel each other out, but that they may be highly significant at a local or regional scale.

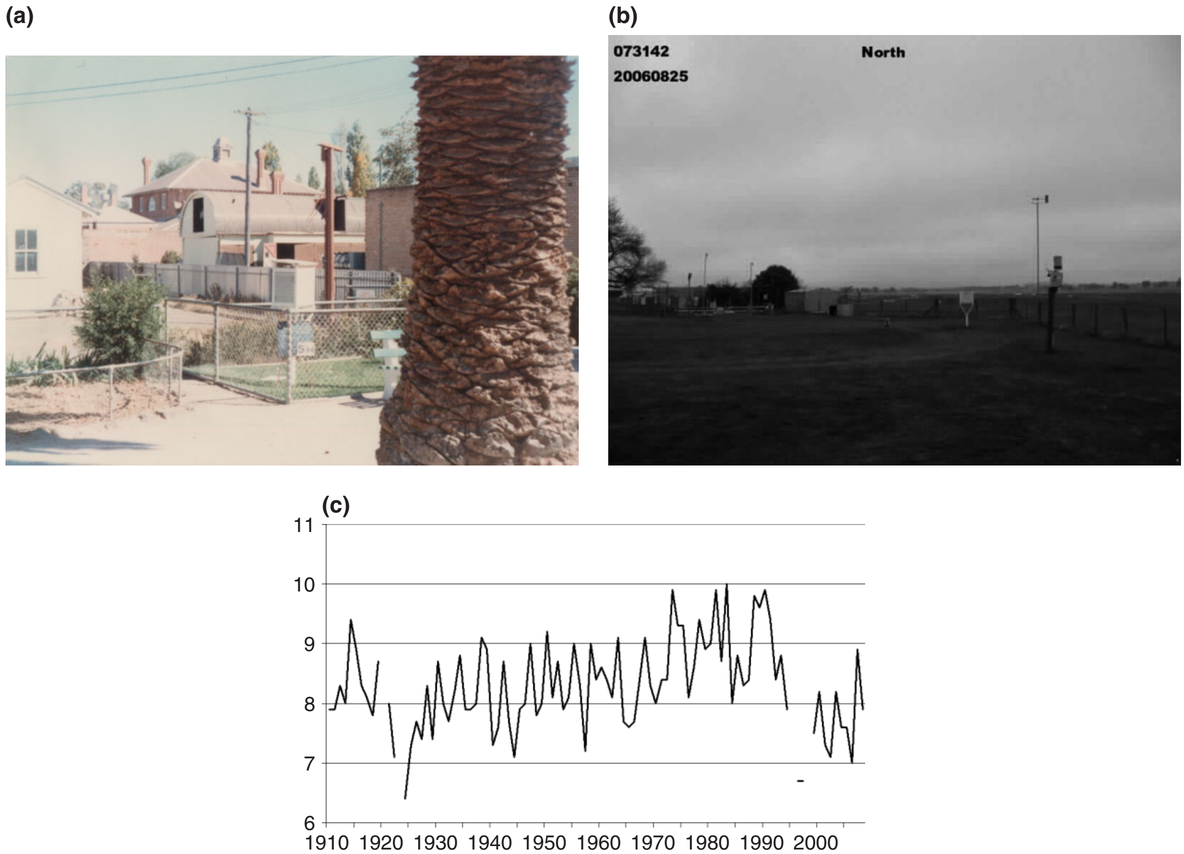

FIGURE 4 | The observation site at Cootamundra, Australia (34°38′S, 148°02′E), (a) before and (b) after a move of 1.7 km in 1995. (c) Mean annual minimum temperatures (°C) at Cootamundra before and after the move.

The potential for a large-scale bias exists, however, if there is a widespread pattern of a particular type of site relocation in an individual country or region. A possible example of this is the establishment of many observing stations at airports during and after the Second World War, as aviation (and the influence of weather information on aviation) grew in importance. If these airport sites generally replaced town center sites, as occurred in some countries,[49] this would be expected to cause a general negative bias in mean minimum temperatures.

Observing Practice Changes

Changes in observation practice can affect temperature data. While temperature measurements are not subjective in the way that, for example, cloud and visibility measurements are, and are therefore less vulnerable to observer biases, there are still a number of observation practice changes that can affect the data; most notably, changes in the method of calculating mean temperature, changes in observation times, and changes in units or data precision.

Algorithms for the Calculation of Mean Temperature

Mean temperature, as noted earlier, is probably the most fundamental temperature variable used in climate change analyses. This requires the calculation of the mean daily temperature (or, equivalently, calculation of a monthly mean from quantities measured daily). The 'true' daily mean can be considered as the integral of the temperature curve averaged across 24 h. It is only in recent years, with the availability of high-resolution data from automatic sensors, that it has become practical to measure this integral, and a variety of methods have therefore evolved to approximate it. These can be placed in three broad categories,[50] all of which are in widespread use:

(**a**) The mean of the daily maximum and minimum temperatures (widely used in English-speaking countries).

(**b**) The mean of a number of regularly spaced observations, e.g., four 6-h or eight 3-h

(a)

(b)

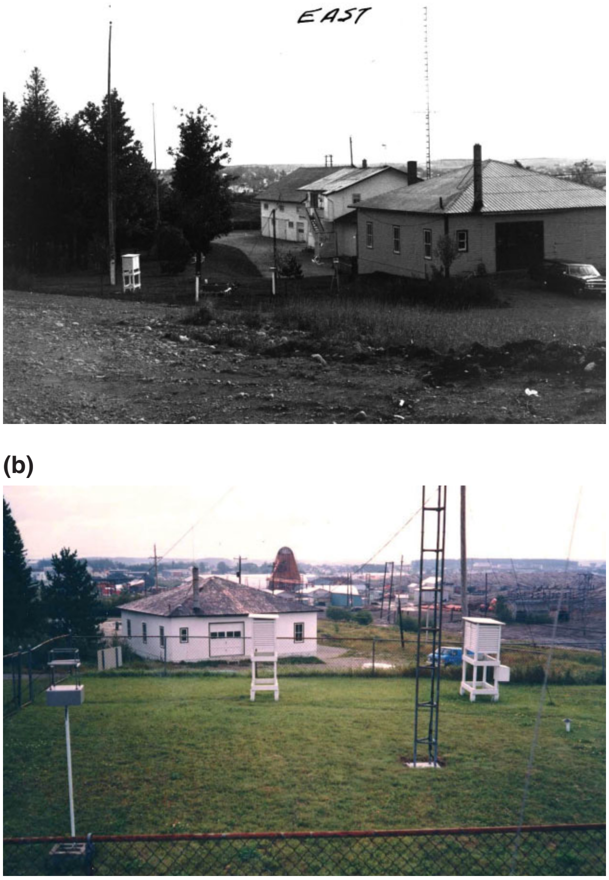

FIGURE 5 | The instrument enclosure at Amos, Canada (48°34′N, 78°07′W), (a) before and (b) after a site move from low to high ground in 1963. This move was found to increase mean minimum temperatures by 1.3°C.[44]

observations (used in China, the former Soviet Union, and numerous other countries).

(c) A mean, sometimes weighted, of temperatures measured at fixed hours [e.g., $(T_{0700} + T_{1400} + 2T_{2100})/4$, where T_n is the temperature at time n], sometimes also incorporating daily maximum and minimum temperatures (widely used in continental Europe and Latin America).

WMO and other formal guidance have changed over the years. The 1983 second edition of the WMO *Guide to Climatological Practices,* which predated much appreciation of the importance of climate data inhomogeneity, recommended the use of method (b). The 1990 Intergovernmental Panel on Climate Change (IPCC) First Assessment Report recommended the retention of whichever practice had been used at a station historically, while the third edition of the WMO *Guide* (due for release shortly at the time of writing) is expected to recommend the use of method (a) (Ian Barnes-Keoghan, personal communication). This guidance, in whatever form, appears to have had very little influence on national-level policies. Method (a) is the least accurate approximation of the 'true' mean—an Austrian study[51] found biases relative to the true mean of up to 0.8°C in individual months at particular stations, and approximately 0.2°C on a network-wide basis, compared with 0.4 and 0.1°C for method (c)—but is also the method which is most likely to be able to be used consistently in a long-term data set, as maximum and minimum temperatures are measured at almost all temperature stations, whereas many stations only report once or twice daily (or do not have digitized fixed-hour data) and therefore do not have the necessary fixed-hour observations for the implementation of method (b) or (c).

The algorithm used for the calculation of mean temperature will have an effect on long-term temperature measurements if there is a change from one algorithm to another, either explicitly[52] or implicitly through, for example, an effective 1-h shift in observation time through the introduction of daylight savings time. Victoria et al.[53] noted a shift of −0.18°C arising from a 1938 change in algorithm in Brazil. There have been a number of studies[50,51,54–56] which have assessed systematic differences between two or more of the methods described above. In general, these have found that differences tend to vary according to the nature of the location (e.g., its position in its local time zone, or the extent to which it is influenced by sea breezes) and the season, but that differences between methods of 0.1–0.2°C on a network-wide basis, and 0.5–1.0°C at the most extreme individual sites, are typical.

Changes in Time of Observation

Changes in the time of observation, such as the change in the time of the evening observation from 1900 to 2100 in Austria,[51] have an obvious impact on fixed-hour observations. Less obviously, they also have an impact on daily maximum and minimum temperatures, which arises because, if the time of observation is in the early morning, the coldest nights are likely to be counted twice, once against the end of one observation period and once against the start of the next (conversely, an observation time in the afternoon will lead to hot days being double-counted).

This issue has received particular attention in the United States, where no firm standards exist for the observation time at the majority of stations, and changes of observation times are a major source of inhomogeneities in temperature data, especially as there has been a tendency over time for stations to shift from afternoon/evening to morning observations. A number of studies[57–59] have found that a change from afternoon to early morning observations typically produces a shift in the order of −1°C in mean temperatures calculated using daily maxima and minima, and Karl et al.[58] developed a model to determine the expected shift that would result from a given change in the time of observation, according to the specific location and season. (This shift will also be a function of interdiurnal temperature variability and would therefore be expected to be less in climates where that variability is less than it is in much of the United States, especially in winter.)

In some other countries, observation time changes have been introduced on a national basis, with three examples being the 1961 change from 0600 to 0000 UTC for minimum temperatures in Canada,[8,60] the 1964 change from 0000 to 0900 UTC local time at Bureau of Meteorology-staffed stations in Australia,[61] and a 1938 shift from 0800 to 1900 UTC for minimum temperatures in Norway.[62] In the Norwegian case, impacts on mean minimum temperatures were found to be up to 1.5°C in some regions and seasons, while in Canada the change was found to introduce a cold bias into minimum temperatures which averaged −0.2°C in western Canada and −0.8°C in eastern Canada.

Data Precision

WMO standards recommend that temperatures should be recorded to the nearest 0.1°C. Like many such standards, these are often not followed in practice. The United States generally records to the nearest whole degree Fahrenheit, while numerous stations elsewhere only report to the nearest 0.5 or 1°C. Even at stations which nominally report to the nearest 0.1°C, observer biases may be present; for example, at all Australian stations analyzed, values ending in .0 and .5 were overrepresented in the data set.[61]

Such changes in data precision should have no systematic impact on mean temperatures. They can, however, have an impact on quantities derived from fixed thresholds (e.g., the number of days above 30°C), especially in climates where temperature variability is low.[43,63] This can also lead to inhomogeneities in such indices when standards change (e.g., a shift from 1 to 0.1°C precision in Spain; Ref 63), or where there has been a changeover from Fahrenheit to Celsius. [For example, at Eddystone Point in Australia (Figure 6), rounding temperatures to the nearest whole degree Celsius leads to a −16% bias in the number of days with maximum temperatures below 15°C.] Zhang et al.[63] address this problem by adding small random values (taken from a distribution with zero mean) to data points, but note that even where it is not addressed, the effect on the homogeneity of indices is mostly small.

TECHNIQUES FOR MAINTAINING HOMOGENEOUS RECORDS AND TREATING INHOMOGENEITIES

As noted earlier, the ideal is to maintain homogeneous temperature records, and, if this is not feasible, to implement any changes in such a way as to allow consistent long-term records to be maintained. The first formal international statement in this field came from the WMO in 1986,[64] when they called for national meteorological services to define reference

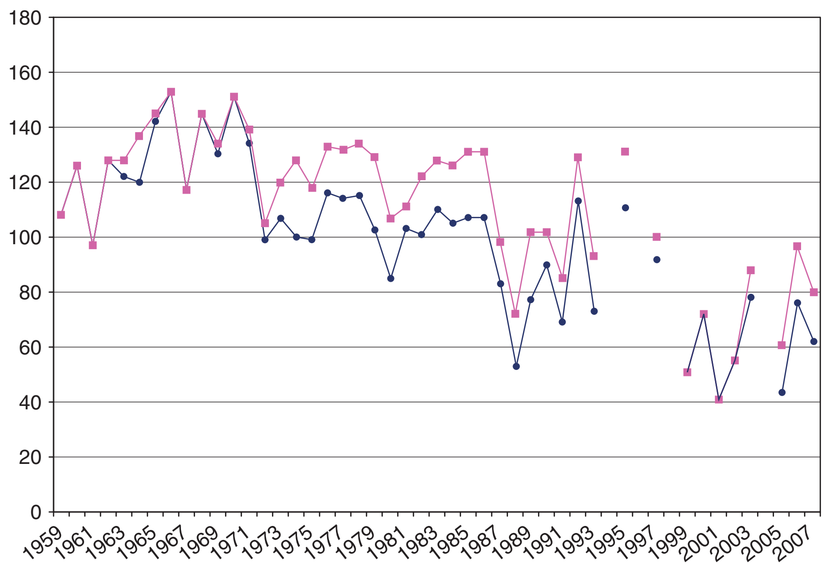

FIGURE 6 | Number of days with maximum temperatures below 15.0°C (pink) and 14.5°C (blue) at Eddystone Point, Australia (40°59′S, 148°21′E). Note the very close correspondence between 1998 and 2003, and before 1972, when most values were rounded to the nearest degree Celsius or degree Fahrenheit, respectively.

climatological stations (RCSs), with a recommended density of 2–10 stations per 250,000 km^2. It was recommended that these stations be selected on the basis of being permanent, having long records, preferably located in an environment unaffected by densely populated or industrialized areas, and having reliable instruments and observers (AWSs were not considered at that time). By 1993, 75 countries had defined a set of RCSs[65] (Canada having done so as early as 1966). A more recent international network following similar principles is the Global Climate Observing System (GCOS) Surface Network (GSN),[66,67] and national and international networks continue to be defined.[68]

A set of 10 principles for maintaining long-term climate records was enunciated by Karl et al.[69] These included the importance of maintaining observations with a long uninterrupted record, of archiving metadata, of giving long-term climate requirements sufficient priority in network and instrument design, and, most importantly, of establishing the effects on the climate record of any changes (of the types described in section *Major Factors Leading to Inhomogeneities in Temperature Records* above) prior to their implementation, for example, through a period of parallel observations. A similar (although not identical) set of principles was endorsed by the UN Framework Convention on Climate Change in 1999 and now forms part of WMO guidance.[67] Street et al.[70] take the view that parallel observation programs are a preferable approach, with statistical homogenization an alternate but less preferred option. Neither Karl et al. or WMO makes specific recommendations on the appropriate minimum length of a parallel observations program. Some national-level policies are more specific; for example, Australia recommends a length of 5 years, with 2 years as a minimum where at all possible (Karl Monnik, pers. comm.).

These principles are only recent developments and, even now, their implementation is far from universal. As a consequence, statistical and other techniques to detect inhomogeneities in temperature data, and to adjust data to, as far as possible, remove these inhomogeneities, are essential and will remain so for the foreseeable future.

It is not the purpose of this article to carry out a detailed discussion of statistical techniques for the treatment of inhomogeneities. For detailed reviews and assessments of such techniques, the reader is referred to Peterson et al.[71] and Reeves et al.,[72] while some aspects are also covered by DeGaetano[73] and Ducré-Robitaille et al.[74] Rather, this article will describe some of the issues and techniques which are in practical use in various commonly used data sets.

Detection of Inhomogeneities

There are two broad methods of detecting inhomogeneities in a temperature time series; through the use of metadata which documents that a change of some kind has occurred at the station, or through statistical tests which detect a significant breakpoint in a time series based on that station's temperature data.

Metadata-based techniques, at least in principle, offer the advantage of knowing definitively that a change has occurred at a station, and often the exact date of that change. In practice, though, the use of metadata presents a number of difficulties.[71] Metadata are often incomplete, inaccurate, or missing can be open to interpretation, often present considerable difficulty in extracting relevant information in usable form (sometimes information relevant to a climate record may form only a small part of a large volume of documentation on other matters relating to a station) and can be misleading without additional knowledge (e.g., a change in station coordinates may arise from a resurvey without the station physically moving). They are also sometimes imprecise—for example, a series of station photographs taken at 10-year intervals may indicate that a change has occurred at some point in a 10-year period, but not the exact date of that change. At the global level, a further challenge is that most historical metadata resides in hard-copy documents, in the local language, in national-level archives and is therefore very difficult to access on a global scale, and multination[75] and global data sets have generally been limited to very basic metadata such as station coordinates and the population of any urban center(s) in close proximity to the site.

There is a well-developed statistical literature on the detection of breakpoints in time series. As the ability to detect a breakpoint in a time series is a function of its size relative to the variance of the time series, many techniques used for temperature data sets seek to reduce the variance of the time series by comparing the data set under review with a well-correlated reference series which are intended to capture the underlying interannual climate variability. Most commonly, this reference series will be a combination of data from one or more neighboring stations. However, this method relies on the assumption that the reference series itself is broadly homogeneous, which will not hold if there is a change which affects an entire network at the same time (e.g., a change in observation time or a change in the method of calculating mean temperature) or a large number of stations enter or leave the reference series around the same time, and in data-sparse areas it may also be difficult to find a suitable reference series.[1] Menne and

Williams[45] used an alternative method using multiple pairwise comparisons between individual stations.

Once a time series (either of a station's data or its data with respect to some reference) for testing has been developed, the next step is to carry out statistical testing for significant breakpoints. Three general techniques which have been widely used for temperature data sets are:

- The standard normal homogeneity test (SNHT) of Alexandersson,[76] used by, among others, Vincent[47] and Laughlin and Kalma.[36]
- Two-phase regression (TPR), originally developed for use in climate by Easterling and Peterson,[77] and with a number of refinements, particularly in determining the true significance of potential breakpoints.[78–82] This method is used for the Global Historical Climatology Network (GHCN) data set,[83] which underlies much of the HadCRU and National Climatic Data Center (NCDC) global analyses described in section *Methods Used for Data Sets Used in Climate Change Detection* and is also used in the RHTest software developed under the auspices of the ETCCDI.[84]
- Visual inspection of a time series (used by Jones et al.[85]).

Reeves et al.[72] found that the first two methods both had positive and negative attributes, depending on what priorities users had (e.g., accurately detecting the date of a changepoint, or minimizing the number of false alarms), with the post-1995 developments in the TPR technique having substantially improved its performance, and that SNHT performed especially well when good reference series were available. The third method has become largely outmoded in recent years but persists in some of the adjusted time series used in global data sets.

Adjustment of Data to Remove Inhomogeneities

In some data sets, such as the European Climate Assessment and Data set,[86] no attempt is made to adjust for inhomogeneities—instead users are informed which stations are homogeneous and which are not, and are left to make their own decisions as to how to use the data.

Many data sets, however, seek to adjust data sets to remove, as far as possible, inhomogeneities which have been identified. Historically, these adjustments have normally involved either a uniform annual adjustment[49] or adjustments calculated for each of the 12 calendar months[47,83,85] and have often been calculated by comparing station means (or their difference with a reference series) before and after an inhomogeneity—for example, Jones et al.[85] use means of interstation differences for 10 years before and after an identified inhomogeneity. Alternatively, adjustments based on a station's characteristics (such as its time of observation or the size of an urban area it is associated with) may be applied.[87]

A characteristic of such methods is that they seek to produce a data set whose monthly or annual means are homogeneous. It does not, however, follow that such an adjusted data set also has homogeneous higher order statistical properties, such as variance or the frequency of extremes. This issue was identified by Trewin and Trevitt,[88] who noted that the temperature difference between sites could be weather-dependent, with, for example, ridge–valley minimum temperature differences typically being larger on cold nights (which tended to be calm and clear) than on warm nights (which tended to be cloudy and/or windy). In recent years, a number of attempts have been made to address this problem. Some have involved explicitly testing the homogeneity of higher order statistical properties, such as mean daily variability[86] or exceedances of percentile-based thresholds,[89] while others have sought to homogenize daily data across the full range of the frequency distribution, by matching percentile points in the frequency distribution[61,90] or by other methods[91]; one of these[61] was the first known attempt to produce a homogenized national-scale data set at the daily timescale. This is currently a very active area of research, in particular through the European COST Action on 'Advances of homogenization methods of climate series: an integrated approach'.[92]

It should be noted that a number of data sets which are described as having adjustments applied at the daily level[9,13,93–96] in fact use interpolation between monthly adjustments to produce a set of calendar-date adjustments that follow a smooth annual cycle and do not use weather- or distribution-dependent daily adjustments.

A particular issue is the adjustment of very old temperature data (early 20th century or earlier), affected by inhomogeneities associated with the introduction of the Stevenson screen or similar shelters as discussed in section *Instrumentation*. Making adjustments for this change is particularly challenging because the change often occurred across a network over a fairly short period of time, and because documentation of the date of the change, the instrument exposure prior to the change, or both is often poor. In some data sets pre-Stevenson data are excluded altogether (e.g., the Australian data set discussed in section *National and Regional Data Sets* below uses a starting date of 1910 for this reason). Other studies (e.g., Ref 13 in Spain) use a standard adjustment based on an experimental comparison of a replica early exposure with a Stevenson screen.

METHODS USED FOR DATA SETS USED IN CLIMATE CHANGE DETECTION

Much of the discussion in sections *Major Factors Leading to Inhomogeneities in Temperature Records; and Techniques for Maintaining Homogeneous Records and Treating Inhomogeneities* has been at the level of the individual station. Most assessments of temperature change are based on data sets which include data from a large number of stations, either at the global, regional, or national scale. In this section, some of the techniques used in the development of those data sets, and the extent to which they might be influenced by non-climatic factors, are considered.

Global Data Sets

There are three major global data sets in widespread use for climate change analysis:

- The HadCRU data set[97,98] developed by the Hadley Centre of the UK Meteorological Office and the Climatic Research Unit (CRU) of the University of East Anglia.
- The NCDC data set[99] developed by the (US) NCDC.
- The NASA–GISS data set[87,100] developed by the Goddard Institute of Space Studies (GISS), part of the (US) National Air and Space Administration (NASA).

The land temperature components of these data sets are all based on the interpolation of station data to a regular grid. The HadCRU and NCDC data sets consist of monthly mean temperature anomalies on a 5° grid, from which global and hemispheric mean anomalies are calculated. The NASA–GISS data are based on a combination of station time series and the marine temperature data sets of the Hadley Centre and NCDC, with global and hemispheric anomalies of mean monthly temperature calculated through interpolation to a grid with latitude-varying spacing. All data sets are available via the web (the major global temperature data sets may be obtained at the following locations: NASA–GISS, data.giss.nasa.gov/gistemp; NCDC, www.ncdc.noaa.gov/oa/climate/research/anomalies/index.html; HadCRU, hadobs.metoffice.com).

The HadCRU data set uses data from a variety of sources. Some of the station data are based on the original homogenized global data set developed by Jones et al.,[85] but homogenized national-level data sets are used in preference to this where they are available. There is no explicit correction for urbanization effects, but they are used in defining an uncertainty in the data. Areal means are calculated separately for the Northern and Southern Hemispheres and then combined to create a global average.

The NCDC data set is based on the GHCN data set, which is homogenized, using neighboring-station data. Urbanization is not explicitly adjusted for, nor are urban areas explicitly excluded, but many corrections for urbanization effects have been made as part of the regular homogenization process. First, differences (the difference in values from one year to the next) have been used to incorporate data from relatively short data sets into the global analysis,[15] allowing additional coverage of otherwise data-sparse regions. Statistical methods have also been used in the grid-interpolation process to exclude excessive damping of variability from undersampled regions. The NCDC area averages are calculated globally as a single domain, which has had the effect of giving the Northern Hemisphere (which has fewer unsampled areas) more weight in the calculation of global averages than is the case for the HadCRU or NASA–GISS data sets.

The NASA–GISS data set is also based on the GHCN data set, although without the GHCN-supplied homogeneity adjustments. An urbanization correction is applied to the data (it is noted that part of the reason for not excluding urban data altogether is to allow the time series to be rapidly updated, urban data generally being more frequently updated and internationally transmitted), but homogeneity corrections are generally not otherwise applied at the station level, except where two or more stations are combined into a single record. Stations within 1200 km of each gridpoint are used in the algorithm for calculating estimated gridpoint values, which has the effect of extending the NASA–GISS analysis over data-sparse regions (especially near the poles) which the HadCRU and NCDC data sets would consider as data voids and thus gives polar regions more effective weight in the NASA–GISS analysis than in the other data sets. It is likely that this largely accounts for the different rankings of the hottest years on record in the different data sets. Year 1998, in which the most abnormal warmth was in the tropics, is the hottest year in the HadCRU data set, whereas year 2005, where the most abnormal warmth was in the Arctic, is the hottest year in the NASA–GISS and NCDC data sets.

National and Regional Data Sets

An increasing number of countries are reporting time series of national area-averaged temperatures or temperature anomalies. Moreover, 21 countries reported national temperature anomalies in the 2008 *State of the Climate* report.[101]

From the available documentation, the methods used in developing these data sets fall into three broad categories:

- (**a**) Area averages of gridded data sets derived from homogenized data from a number of long-term stations, sometimes supplemented with near-real time analyses from a broader range of stations.
- (**b**) Area averages of gridded data sets derived using all available data.
- (**c**) Averages (sometimes weighted, sometimes not) of data from a small number of long-term stations with homogenized data.

Countries which use method (a) include the United States,[102] Australia,[103] China,[104] Canada,[105] and Egypt.[106] Method (b) is used by the United Kingdom[107] and Germany.[108] The best known data set using method (c) is the Central England Temperature (CET) data set which extends back to 1659[109,110]; it has also been used for long-term time series representing Scotland and Northern Ireland,[111] and for national data sets for New Zealand[112] and Switzerland.[113] Norway has combined regional averages into a national average.[114,115] Some of these data sets explicitly exclude urban stations but most do not.

An analysis which does not fit into any of these categories is the Antarctic analysis of Chapman and Walsh,[116] who splice numerous short AWS data sets to produce an analysis over the continent, including data-sparse areas of the plateau.

Methods (a) and (c), providing station-level homogenization has been carried out properly, should produce a homogeneous data set capable of monitoring long-term temperature trends. However, the relatively small station networks used in method (c) may not be sufficient to monitor interannual variability, especially over regions larger than central England, Switzerland, or New Zealand. Janis et al.,[117] Vose and Menne,[118] and Jones and Trewin[119] all considered the question of the optimal station network to monitor temperature variability over their areas of interest, with the latter two finding diminishing returns with an increased number of stations, and networks of 100–200 stations sufficient to define temperature variability to a reasonable degree of accuracy over regions the size of Australia or the United States.

Method (b) relies on the implicit assumption that station-level inhomogeneities will largely cancel each other out, and that there are no major changes to the network (e.g., the establishment of new stations in data-sparse high mountain locations) that are likely to create biases in gridded analyses. The former assumption may hold as long as there are no national-level changes in observing methods (as discussed in section *Site Relocations*); the latter is probably valid in countries with dense networks over their whole territory over a long period (which is the case in both the UK and Germany), but would be more doubtful over larger areas with substantial data voids.

WHAT CHALLENGES AND UNCERTAINTIES REMAIN?

Great progress has been made on addressing the effect of external influences on land temperature measurements over the last 20 years, and numerous data sets exist which allow temperature trends over a century or more to be analyzed without significant influence from non-climatic factors.

Nevertheless, a number of challenges remain. Many adjustments to data inhomogeneities in global data sets have a substantial uncertainty attached, because of limited accessibility of metadata and the sparseness of globally distributed data that could be used in the development of reference series. National-level analyses typically have much more access to metadata and comparison data, and the approach followed in the HadCRU data set, of incorporating national-level homogenized data sets where they exist, is a promising one. However, despite the progress made in developing capacity for climate change analysis in developing countries through initiatives such as the ETCCDI workshops,[84] it is likely that homogenization of data from many parts of the world will still have to be carried out at the global level for the foreseeable future.

The effective use of early instrumental records (prior to the early 20th century) remains a challenge. While numerous studies have quantified the biases arising from particular types of pre-20th century instrument exposures in an experimental setting, much remains to be done to assess the effect that such instrument changes have had across a full observing network at the national or international scale. In some cases this problem may prove largely intractable because of a lack of documentation of historical instrument exposures. Effective communication of this issue is also important, as raw pre-20th century temperatures measured in exposures which are not consistent with more recent standards, and trends based on them, are sometimes reported in the public arena.

The homogenization of daily and sub-daily data, which is necessary to support analyses of changes in temperature variability and extremes, also remains a field with major challenges, although significant advances are likely as a result of the Cooperation in Science and Technology (COST) action currently in progress in Europe. Simply developing a long-term global daily data set of any kind is difficult, as many countries limit the release of historical daily data, and a homogenized global scale daily temperature data set remains a very distant goal. For the time being, it is likely that any effective global analyses of temperature extremes will be a consolidation of national or regional analyses, along the lines of Alexander et al.[7]

Reanalyses and satellite observations are not part of the scope of this review. In the context of this review, however, they potentially provide an additional tool for assessing land temperature measurements. In particular, as land temperature measurements do not normally form part of the input data for reanalyses, those reanalyses could potentially be used as an independent reference series for assessing the homogeneity of land temperature data, especially in data-sparse regions.

Further quantification of the effect of changes in land use and land cover on observed temperatures would also be of value; in particular, more rigorous separation of the impacts of the presence of an urban area *per se* from those of the land use and land cover in the immediate vicinity of the observation site, and further assessment of the impacts of nonurban changes in land use or land cover.

NOTE

[a]CCl, World Meteorological Organization (WMO) Commission for Climatology; CLIVAR, World Climate Research Programme (WCRP) project for Climate Variability and Predictability; JComm, Joint WMO, IOC [United Nations Educational, Scientific and Cultural Organization (UNESCO) Intergovernmental Oceanographic Commission] Technical Commission for Oceanography and Marine Meteorology.

ACKNOWLEDGEMENTS

Comments from Dean Collins, Branislava Jovanovic, Michael Coughlan and two anonymous reviewers on earlier drafts of this manuscript were of great assistance. Figure 1 was supplied by Ken Devine, Figure 5 by Lucie Vincent of Environment Canada, and Figures 2, 3 and 4(a) and (b) by the Australian Bureau of Meteorology. Material supplied by David Parker and Phil Jones was also of assistance.

REFERENCES

1. Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. *Guidelines on Climate Metadata and Homogenization*. WMO/TD No. 1186, Geneva: World Meteorological Organization; 2003. Available online at: http://www.wmo.int/pages/prog/wcp/wcdmp/documents/WCDMP-53_000.pdf. (Accessed 2009).

2. Baker CB, Helfert M. U.S. Climate Reference Network (USCRN): a unique national long-term climate monitoring network. *12th Conference on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, New Orleans, 20–24 January 2008.

3. Ehinger J. *Siting and Exposure of Meteorological Instruments*. WMO/TD No. 589. Geneva: World Meteorological Organization; 1993.

4. World Meteorological Organization. *Guide to Meteorological Instruments and Methods of Observation*. 7th ed. WMO No. 8. Geneva: World Meteorological Organization; 2008.

5. Australian Bureau of Meteorology. *Guidelines for the Siting and Exposure of Meteorological Instruments and Observing Facilities*. Observations Specification No. 2013.1: Melbourne: Bureau of Meteorology; 1997. Available online at http://www.bom.gov.au/inside/oeb/networks/20131.pdf. (Accessed 2009).

6. Meteorological Service of Canada. *Siting Standards for Meteorological Observing Sites*. MSC STDS 2–2001; 2001. Available online at: http://www.msc-smc.ec.gc.ca/msb/circulars/msc/pdf/msc_2_2001_e.pdf.

7. Alexander LV, Zhang X, Peterson TC, Caesar J, Gleason B, et al. Global observed changes in daily climate extremes of temperature and precipitation. *J Geophys Res* 2006, 111:D05109. doi: 10.1029/2005JD006290.

8. Vincent LA, Wijngaarden WA, Hopkinson R. Surface temperature and humidity trends in Canada for 1953–2005. *J Clim* 2007, 20:5100–5113. doi: 10.1175/JCLI4293.1.

9. Moberg A, Bergström H, Krigsman JR, Svanereb O. Daily air temperature and pressure series for Stockholm (1756–1998). *Clim Change* 2002, 53:171–212.

10. Parker DE. Effects of changing exposure of thermometers at land stations. *Int J Climatol* 1994, 14:1–31.

11. Böhm R, Jones PD, Hiebl J, Frank D, Brunetti M, et al. The Early Instrumental Warm-Bias: A Solution for Long Central European Temperature Series 1760–2007. *Clim Change* 2009; Available online at: https://naemail.wiley.com/owa/redir.aspx?C=16e999da0378497e8ff0451f0b66ea33&URL=http%3a%2f%2fwww.springer.com%2fearth%2bsciences%2band%2bgeography%2fmeteorology%2b%2526%2bclimatology%2fjournal%2f10584.

12. Dettwiller J. Evolution séculaire de la temperature de l'air à Paris. *La Meteorologie* 1978, 13:95–130.

13. Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, et al. The development of a new daily adjusted temperature dataset for Spain (SDATS) (1850–2003). *Int J Climatol* 2006, 26:1777–1802. doi:10.1002/joc.1338.

14. Klingbjer P, Moberg A. A composite monthly temperature record from Tornedalen in northern Sweden, 1802–2002. *Int J Climatol* 2003, 23:1465–1494. doi: 10.1002/joc.946.

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

16. Laing J. Maximum summer temperatures recorded in Glaisher stands and Stevenson screens. *Met Mag London* 1977, 106:220–228.

17. Nicholls N, Tapp R, Burrows K, Richards D. Historical thermometer exposures in Australia. *Int J Climatol* 1996, 16:705–710.

18. Huband NDS, King SC, Huxley MW, Butler DR. The performance of a thermometer screen on an automatic weather station. *Ag Forest Met* 1984, 33:249–258.

19. Quayle RG, Easterling DR, Karl TR, Hughes PY. Effects of recent thermometer changes in the cooperative station network. *Bull Amer Met Soc* 1991, 72:1718–1723.

20. Hubbard KG, Lin X. Reexamination of instrument change effects in the U.S. Historical Climatology Network. *Geophys Res Lett* 2006, 33:L15710. doi:10.1029/2006GL027069.

21. Gajar B, Ondras M, Kubjatkova D, Zahumensky I. Long-term data comparison of automatic weather station (AWOS) and climatological data. 1999. *2nd International Conference on Experiences with Automatic Weather Stations (ICEAWS)*, Vienna, 27–29 September 1999.

22. Spengler R. Use of AWS-data in climatology. *2nd International Conference on Experiences with Automatic Weather Stations (ICEAWS)*, Vienna, 27–29 September 1999.

23. Perry MC, Prior MJ, Parker DE. An assessment of the suitability of a plastic thermometer screen for climatic data collection. *Int J Climatol* 2007, 27:267–276. doi: 10.1002/joc.1381.

24. Barnett A, Hatton DB, Jones DW. *Recent Changes in Thermometer Design and their Impact*. WMO/TD No. 871. Geneva: World Meteorological Organization; 1998.

25. Van der Meulen JP, Brandsma T. Thermometer screen intercomparison in De Bilt (The Netherlands), Part I: understanding the weather-dependent temperature differences. *Int J Climatol* 2008, 28:371–387. doi: 10.1002/joc.1531.

26. Van der Meulen JP, Brandsma T. Thermometer screen intercomparison in De Bilt (The Netherlands), Part II: description and modelling of mean temperature differences and extremes. *Int J Climatol* 2008, 28:389–400. doi: 10.1002/joc.1524.

27. Sun B, Baker CB, Karl TR, Gifford MD. A comparative study of ASOS and USHCN temperature measurements. *J Atmos Ocean Tech* 2005, 22:679–686. doi: 10.1175/JTECH1752.1.

28. Guttman NB, Baker CB. Exploratory analysis of the difference between temperature observations recorded by ASOS and conventional methods. *Bull Amer Met Soc* 1996, 77:2865–2873.

29. Milewska EJ, Hogg WD. Continuity of climatological observations with automation—temperature and precipitation amounts from AWOS (Automated Weather Observing System). *Atmos Ocean* 2002, 40:333–359.

30. Parker DE. Urban heat island effects on estimates of observed climate change. *WIRES Clim Change* 2010, 1:123–133, doi:10.1002/wcc.21.

31. Lovell DB, Bonfils C. The effect of irrigation on regional temperatures: a spatial and temporal analysis of trends in California, 1934–2002. *J Clim* 2008, 21:2064–2071. doi: 10.1175/2007JCLI1755.1.

32. Mahmood R, Foster SA, Keeling T, Hubbard KG, Carlson C, et al. Impacts of irrigation on 20th-century temperatures in the Northern Great Plains. *Glob Planet Change* 2006, 54:1–18. doi: 10.1016/j.gloplacha.2005.10.004.

33. Roy SS, Mahmood R, Niyogi D, Lei M, Foster SA, et al. Impacts of the agricultural Green Revolution-induced land use changes on air temperatures in India. *J Geophys Res* 2007, 112:D21108. doi: 10.1029/2007JD008834.

34. Hale RC, Gallo KP, Loveland TR. Influences of specific land use/land cover conversions on climatological normals of near-surface temperature. *J Geophys Res* 2008, 113:D14113. doi: 10.1029/2007JD009548.
35. Bootsma A. Estimating minimum temperature and climatological freeze risk in hilly terrain. *Agric Inf Dev Bull* 1976, 16:425–443.
36. Laughlin GP, Kalma JD. Frost hazard assessment from local weather and terrain data. *Ag Forest Met* 1987, 40:1–16.
37. Lindkvist L, Gustavsson T, Bogren G. A frost assessment method for mountainous areas. *Ag Forest Met* 2000, 102:51–67.
38. Thompson RD. Some aspects of the synoptic mesoclimatology of the Armidale district, New South Wales. *J Appl Biomech* 1973, 12:578–588.
39. Young FD. Effect of topography on temperature distributions in Southern California. *Mon Weath Rev* 1920, 48:462–463.
40. Trewin BC. A notable frost hollow at Coonabarabran, New South Wales. *Aust Met Mag* 2005, 54:15–21.
41. Doggett M, Daly C, Smith J, Gibson W, Taylor G, et al. High resolution 1971–2000 mean monthly temperature maps for the western United States. *14th Conference on Applied Climatology*, Seattle, 10–16 January 2004.
42. Patzert WC, LaDochy S, Willis JK, Mardirosian T. Will the real Los Angeles stand up: impacts of a weather station's relocation on climatic records (and record weather). 2007. *14th Symposium on Meteorological Observation and Instrumentation*, San Antonio, 13–18 January 2007.
43. Expert Team on Climate Change Detection and Indices. *Examples of Inhomogeneities in Climate Datasets*. Available at: http://cccma.seos.uvic.ca/ETCCDMI/docs/Classic_Examples.pdf. Accessed 12 October 2009.
44. Syrakova M, Stefanova M. Homogenization of Bulgarian temperature series. *Int J Climatol* 2009, 29:1835–1849. doi: 10.1002/joc.1829.
45. Menne MJ, Williams CN. Homogenization of temperature series via pairwise comparisons. *J Clim* 2009, 22:1700–1717. doi: 10.1175/2008JCLI2263.1.
46. Torok SJ. *The Development of A High Quality Historical Temperature Data Base for Australia*. PhD thesis, University of Melbourne, Australia; 1996.
47. Vincent LA. A technique for the identification of inhomogeneities in Canadian temperature series. *J Clim* 1998, 11:1094–1104.
48. Easterling DR, Peterson TC. The effect of artificial discontinuities on recent trends in maximum and minimum temperatures. *Atmos Res* 1995, 37:19–26.
49. Torok SJ, Nicholls N. A historical annual temperature dataset for Australia. *Aust Met Mag* 1996, 45:251–260.
50. Trewin BC. Effects of changes in algorithms used for the calculation of Australian mean temperature. *Aust Met Mag* 2004, 53:1–11.
51. Böhm R, Auer I, Schöner W. Dangers and advantages of automatisation concerning homogeneity of long-term time series. *2nd International Conference on Experiences with Automatic Weather Stations (ICEAWS)*, Vienna, 27–29 September 1999.
52. Heino R. *Climate in Finland During the Period of Meteorological Observations*. Helsinki: Finnish Meteorological Institute, 1994.
53. Victoria RL, Martinelli LA, Moraes JM, Ballester MV, Krusche AV, et al. Surface air temperature variations in the Amazon region and its borders during this century. *J Clim* 1998, 11:1105–1110.
54. Collison P, Tabony RC. The estimation of mean temperatures from daily maxima and minima. *Met Mag London* 1984, 113:329–337.
55. Weber RO. Influence of different daily mean formulas on monthly and annual averages of temperature. *Theor Appl Climatol* 1993, 47:205–213.
56. Weiss A, Hays CJ. Calculating daily mean air temperatures by different methods: implications from a nonlinear algorithm. *Ag Forest Met* 2005, 128:57–65. doi: 10.1016/j.agformet.2004.08.008.
57. Baker DG. Effect of observation time on mean temperature estimation. *J Appl Biomech* 1975, 14:471–476.
58. Karl TR, Williams CN, Young PJ, Wendland WM. A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J Clim Appl Met* 1986, 25:145–160.
59. Vose RS, Williams CN, Peterson TC, Karl TR, Easterling DR. An evaluation of the time of observation bias adjustment in the US Historical Climatology Network. *Geophys Res Lett* 2003, 30:2046. doi:10.1029/2003GL018111.
60. Vincent LA, Milewska EJ, Hopkinson R, Malone L. Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian synoptic stations. *J Appl Met Climatol* 2009, 48:2160–2168. doi: 10.1175/2009JAMC2191.1.
61. Trewin BC. *Extreme temperature events in Australia*. PhD thesis, University of Melbourne, Australia; 2001.
62. Nordli PO. *Adjustments of Norwegian Monthly Means of Daily Minimum Temperature*. KLIMA Report 6/97. Oslo: Norwegian Meteorological Institute; 1997.
63. Zhang X, Zwiers FW, Hegerl G. The influences of data precision on the calculation of temperature percentile indices. *Int J Climatol* 2009, 29:321–327. doi: 10.1002/joc.1738.
64. World Meteorological Organization. *Guidelines on the Selection of Reference Climatological Stations*

(RCSs) from the Existing Climatological Station Network. WMO/TD No. 130, Geneva: World Meteorological Organization; 1986.

65. World Meteorological Organization. *Report of the Experts Meeting on Reference Climatological Stations (RCS) and National Climate Data Catalogues (NCC), Offenbach am Main, 25–27 August 1992*. WMO/TD No. 535, Geneva: World Meteorological Organization; 1992.
66. Peterson TC, Daan H, Jones PD. Initial selection of a GCOS Surface Network. *Bull Amer Met Soc* 1997, 78:2145–2152.
67. World Meteorological Organization. *Guide to the GCOS Surface and Upper-Air Networks: GSN and GUAN*. WMO/TD No. 1106, Geneva: World Meteorological Organization; 2002.
68. World Meteorological Organization. *Synthesis of National Reports on Systematic Observation for Climate*. WMO/TD No. 1490. Geneva: World Meteorological Organization; 2009.
69. Karl TR, Derr VE, Easterling DR, Folland CK, Hofmann DJ, et al. Critical issues for long-term climate monitoring. *Clim Change* 1995, 31:185–221.
70. Street RB, Allsopp D, Durocher Y. *Guidelines for Managing Changes in Climate Observation Programmes*. WMO/TD No. 1378. Geneva: World Meteorological Organization; 2007.
71. Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, et al. Homogeneity adjustments of in situ atmospheric climate data: a review. *Int J Climatol* 1998, 18:1493–1517.
72. Reeves J, Chen J, Wang XL, Lund R, Lu Q. A review and comparison of changepoint detection techniques for climate data. *J Appl Met Climatol* 2007, 46:900–915. doi: 10.1175/JAM2493.1.
73. DeGaetano AT. Attributes of several methods for detecting discontinuities in mean temperature series. *J Clim* 2006, 19:838–853. doi: 10.1175/JCLI3662.1.
74. Ducré-Robitaille JF, Vincent LA, Boulet G. Comparison of techniques for detection of discontinuities in temperature series. *Int J Climatol* 2003, 23:1087–1101.
75. Klok EJ, Klein Tank AMG. Updated and extended European dataset of daily climate observations. *Int J Climatol* 2009, 29:1182–1191. doi: 10.1002/joc.1779.
76. Alexandersson H. A homogeneity test applied to precipitation data. *J Climatol* 1986, 6:661–675.
77. Easterling DR, Peterson TC. A new method for detecting and adjusting for undocumented discontinuities in climatological time series. *Int J Climatol* 1995, 15:369–377.
78. Lund R, Reeves J. Detection of undocumented changepoints: a revision of the two-phase regression model. *J Clim* 2002, 15:2547–2554. doi: 10.1175/1520-0442(2002)015(2547:DOUCAR)2.0.CO;2.
79. Wang XL. Comments on "Detection of undocumented changepoints: a revision of the two-phase regression model". *J Clim* 2003, 16:3383–3385. doi: 10.1175/1520-0442(2003)016(3383:CODOUC)2.0.CO;2.
80. Wang XL, Wen QH, Wu Y. Penalized maximal t test for detecting undocumented mean change in climate data series. *J Appl Biomech* 2007, 46:916–931. doi: 10.1175/JAM2504.1.
81. Wang XL. Penalized maximal F-test for detecting undocumented mean-shifts without trend-change. *J Atmos Oceanic Tech* 2008, 25:368–384. doi: 10.1175/2007JTECHA982.1.
82. Wang XL. Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal t or F test. *J Appl Meteor Climatol* 2008, 47:2423–2444. doi: 10.1175/2008JAMC1741.1.
83. Peterson TC, Vose RS. An overview of the Global Historical Climatology Network temperature database. *Bull Amer Met Soc* 1997, 78:2837–2849.
84. Peterson TC, Manton MJ. Monitoring changes in climate extremes: a tale of international collaboration. *Bull Amer Met Soc* 2008, 89:1266–1271. doi: 10.1175/2008BAMS2501.1.
85. Jones PD, Raper SCB, Kelly PM, Wigley TML, Bradley RS, et al. Northern Hemisphere surface air temperature variations: 1851–1984. *J Clim Appl Met* 1986, 25:161–179.
86. Wijngaard JB, Klein Tank AMG, Können GP. Homogeneity of 20th century European daily temperature and precipitation series. *Int J Climatol* 2003, 23:679–692. doi: 10.1002/joc.906.
87. Hansen J, Ruedy R, Glascoe J, Sato M. GISS analysis of surface temperature change. *J Geophys Res* 1999, 104:30997–31022.
88. Trewin BC, Trevitt ACF. The development of composite temperature records. *Int J Climatol* 1996, 16:1227–1242.
89. Allen RJ, DeGaetano AT. A method to adjust long-term temperature extreme series for nonclimatic inhomogeneities. *J Clim* 2000, 13:3680–3695. doi: 10.1175/1520-0442(2000)013 <3680:AMTALT>2.0CO;2.
90. Della-Marta PM, Wanner H. A method of homogenizing the extremes and mean of daily temperature measurements. *J Clim* 2006, 19:4179–4197. doi: 10.1175/JCLI3855.1.
91. Brandsma T, Können GP. Application of nearest-neighbour resampling for homogenizing temperature records on a daily to sub-daily level. *Int J Climatol* 2006, 26:75–89. doi:10.1002/joc.1236.
92. COST. *Action ES0601: Advances in Homogenisation Methods of Climate Series: An Integrated Approach*

(HOME). 2009. Available at: http://www. homogenisation.org. Accessed 1 October 2009.

93. Bergström H, Moberg A. Daily air temperature and pressure series for Uppsala (1722–1998). *Clim Change* 2002, 53:213–252.

94. Brunet M, Saladié O, Jones P, Sigró J, Aguilar E, et al. *A Case-Study/Guidance on the Development of Long-Term Daily Adjusted Temperature Datasets*. WMO/TD No. 1425. Geneva: World Meteorological Organization; 2007. Available online at http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/ WCDMP_Spain_case_study-cor_ver6March.pdf.

95. Li Q, Zhang H, Chen J, Li W, Liu X, et al. A mainland China homogenized temperature dataset of 1951–2004. *Bull Amer Met Soc* 2009, 90:1062–1065. doi: 10.1175/2009BAMS2736.1.

96. Vincent LA, Zhang X, Bonsal BR, Hogg WD. Homogenization of daily temperatures over Canada. *J Clim* 2002, 15:1322–1334. doi: 10.1043/1520-0442(2002)015(1322:HODTOC)2.0.CO;2.

97. Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD. Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. *J Geophys Res* 2006, 111:D12106. doi:10.1029/2005JD006548.

98. Jones PD, Moberg A. Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *J Clim* 2003, 16:206–223. doi: 10.1175/1520-0442(2003)016(0206:HALSSA)2.0.CO;2.

99. Smith TM, Reynolds RW, Peterson TC, Lawrimore J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J Clim* 2008, 21:2283–2286. doi:10.1175/2007JCLI2100.1.

100. Hansen J, Ruedy R, Sato M, Imhoff M, Lawrence W, et al. A closer look at U.S. and global surface temperature change. *J Geophys Res* 2001, 106:23947–23963.

101. Peterson TC, Baringer MC, eds. State of the climate in 2008. *Bull Amer Met Soc* 2009, 90:S1–S196.

102. Menne MJ, Williams CN, Vose RS. The U.S. Historical Climatology Network monthly temperature data, version 2. *Bull Amer Met Soc* 2009, 90:993–1007. doi:10.1175/2008BAMS26131.

103. Della-Marta P, Collins D, Braganza K. Updating Australia's high-quality annual temperature dataset. *Aust Met Mag* 2004, 53:75–93.

104. Ding Y, Ren G, Zhao Z, Xu Y, Luo Y, et al. Detection, causes and projection of climate change over China: an overview of recent progress. *Adv Atmos Sci* 2007, 24:954–971. doi: 10.1007/s00376-007-0954-4.

105. Vincent LA, Gullett DW. Canadian historical and homogeneous temperature datasets for climate change analyses. *Int J Climatol* 1999, 19:1375–1388.

106. Domroes M, El-Tantawi A. Recent temporal and spatial temperature changes in Egypt. *Int J Climatol* 2005, 25:51–63. doi: 10.1002/joc.1114.

107. Perry M, Hollis D. The generation of monthly gridded datasets for a range of climatic variables over the United Kingdom. *Int J Climatol* 2005, 25:1041–1054. doi: 10.1002/joc.1161.

108. Deutscher Wetterdienst. *Information: Description of the Spatial Means*. Available at: http://www.dwd.de/. Accessed 25 September 2009.

109. Manley G. Central England temperatures: monthly means 1659 to 1973. *Quart J Roy Met Soc* 1974, 100:389–405.

110. Parker DE, Legg TP, Folland CK. A new daily central England temperature series, 1772–1991. *Int J Climatol* 1992, 12:317–342.

111. Jones PD, Lister D. The development of monthly temperature series for Scotland and Northern Ireland. *Int J Climatol* 2004, 24:569–590. doi: 10.1002/joc.1017.

112. Folland CK, Salinger MJ. Surface temperature trends and variations in New Zealand and the surrounding ocean, 1871–1993. *Int J Climatol* 1995, 15:1195–1218.

113. Begert M, Schlegel T, Kirchhofer W. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int J Climatol* 2005, 25:65–80. doi:10.1002/joc.1118.

114. Hanssen-Bauer I, Nordli PO. *Annual and seasonal temperature variations in Norway 1876-1997*. KLIMA Report 25/98. Oslo: Norwegian Meteorological Institute; 1998.

115. Hanssen-Bauer I, Førland E. Temperature and precipitation variations in Norway 1900–1994 and their links to atmospheric circulation. *Int J Climatol* 2000, 20:1693–1708.

116. Chapman WL, Walsh JE. A synthesis of Antarctic temperatures. *J Clim* 2007, 20:4096–4117. doi:10.1175/JCLI4236.1.

117. Janis MJ, Hubbard KG, Redmond KT. Station density strategy for monitoring long-term climatic change in the contiguous United States. *J Clim* 2004, 17:151–162. doi: 10.1175/1520-0442(2004)017(0151:SDSFML)2.0.CO;2.

118. Vose RS, Menne MJ. A method to determine station density requirements for climate observing networks. *J Clim* 2004, 17:2961–2971. doi: 10.1175/1520-0442(2004)017(2961:AMTDSD)2.0.CO;2.

119. Jones DA, Trewin B. On the adequacy of digitised historical Australian daily temperature data for climate monitoring. *Aust Met Mag* 2002, 51:237–250.

