Urbanization bias I. Is it a negligible problem for global temperature estimates?

Ronan Connolly [*1] ***, Michael Connolly*** [1]

[1] *Connolly Scientific Research Group. Dublin, Ireland.*

Abstract

Several studies have claimed that the warming bias introduced to global temperature estimates by urbanization bias is negligible. On the basis of this claim, none of the groups calculating global temperature estimates (except for NASA Goddard Institute for Space Studies) explicitly correct for urbanization bias. However, in this article, by re-evaluating these studies individually, it was found that there was no justification for this.

There is considerable evidence that there has been global warming since the late 1970s. The urbanization bias problem is sometimes incorrectly framed as being a question of whether there has recently been global warming or not. However, the recent warming appears to have followed a period of global cooling from an earlier warm period which ended in the 1940s. So, resolving the urbanization bias problem is necessary to address issues such as how the recent warm period compared to the early 20th century warm period. If the earlier warm period was comparable to the recent warm period, then claims that recent global temperature trends are unprecedented or unusual will need to be re-evaluated.

Citation:
R. Connolly, and M. Connolly (2014). *Urbanization bias I. Is it a negligible problem for global temperature estimates?*, Open Peer Rev. J., 28 (*Clim. Sci.*), ver. 0.2 (non peer reviewed draft).
URL: http://oprj.net/articles/climate-science/28

Version: 0.2 (non peer-reviewed)
First submitted: January 8, 2014.
This version submitted: February 3, 2015.

1 Introduction

In this series of three companion papers[1, 2] we consider the extent to which the global increase in urbanization since the Industrial Revolution has biased those estimates of global temperature trends which are based on weather station records[3–9], i.e., those listed in Table 1.

The fact that urban areas tend to be warmer than neighbouring rural areas has been known since at least the 19th century[10]. The extra warmth associated with urban areas is referred to as an *"urban heat island"* (often abbreviated to UHI), since it is confined to locations in or near urban areas. Urban areas still only comprise about 1% of the Earth's surface, and so this effect does not have a major effect on global temperatures. However, a large percentage of weather stations are located in or near urban heat islands, and as a result, current calculations of global temperatures using weather records may be disproportionately biased by the urban heat island effect.

In this paper (Paper I), we reassess the oft-cited claim that the growth in these urban heat islands has only had a small or negligible effect on the calculation of global temperature trends.

In Paper II, we assess the adjustments applied by the only group that explicitly attempts to correct their estimates for urbanization bias, the Goddard Institute for Space Studies[3, 11, 12]. We identify several serious problems with these adjustments and find that they introduce about as many biases as they remove[1].

In Paper III, we attempt to assess the extent of urbanization bias in two of the main weather station data sets used for estimating global temperature trends - the U.S. Historical Climatology Network and the Global Historical Climatology Network[2]. We find that many of the stations in these datasets are potentially affected by urbanization bias, particularly

*Corresponding author: ronanconnolly@yahoo.ie. Website: http://globalwarmingsolved.com

Urbanization bias	Name	Research group	Ref.
Attempt adjustment	GISTEMP	National Aeronautics and Space Administration's Goddard Institute for Space Studies (NASA GISS)	[3]
Believe negligible	CRUTEM4	University of East Anglia's Climate Research Unit (CRU)	[4]
Believe negligible	GSTA	National Oceanic and Atmospheric Administration's National Climatic Data Center (NOAA NCDC)	[5]
Believe negligible	Lugina et al.	Carbon Dioxide Information Analysis Center (CDIAC)	[6]
Believe negligible	JMA	Japan Meteorological Agency's Tokyo Climate Center (JMA)	[7]
Believe negligible	BEST	Berkeley Earth Surface Temperature (BEST)	[8]
Believe negligible	CMA	China Meteorological Agency (CMA)'s National Meteorological Information Center	[9]

Table 1: *List of recent global temperature estimates based on weather station records.*

amongst the stations with the longest and most complete records.

This means that, with the currently available data, it is very difficult to calculate global temperature trends from the weather records without incorporating a substantial amount of urbanization bias. In this paper, we reassess a number of studies which have reached the opposite conclusion. These are the ten sets of studies listed in Table 2, i.e., Hansen & Lebedeff, 1987[13]; Wigley & Jones, 1988[14]; Jones et al., 1990[15]; Easterling et al., 1997[16]; Peterson et al., 1999[17]; Peterson, 2003[18]; Parker, 2004[19] & 2006[20]; Efthymiadis & Jones, 2010[21]; Wickham et al., 2013[22] and the Hansen et al., 1999[11]; 2001[12] and 2010[3] studies. For various reasons, each of these studies have claimed that the magnitude of urbanization bias in the current global temperature estimates is either small or negligible. On the basis of this remarkable claim, none of the groups except for the Goddard Institute of Space Studies currently explicitly attempt to correct their estimates for urbanization bias, as can be seen from Table 1. However, in this paper, we systematically re-evaluate each of these studies, and in all cases find that their conclusion is invalid.

The format of this paper is as follows. In Section 2 we review some of the evidence in the literature that urbanization bias is a systemic problem for weather station-based estimates of global temperature trends, and discuss some of the challenges inherent in adequately resolving this problem. In Section 3, we will consider some of the flaws that are common to more than one of the studies we are re-assessing. In Section 4, we will re-assess each of the studies in Table 2 in turn. Finally, in Section 5, we offer some concluding remarks.

2 The urbanization bias problem

Since Howard's studies in the early 19th century of the city of London, U.K., it has been known that urban areas tend to be warmer than neighbouring rural areas[10], i.e., they demonstrate an *"urban heat island"* (often abbreviated to UHI). Although Howard identified many of the factors still used to explain this phenomenon[10], the exact relationship between these factors is complex and varies from location to location. As a result, it is still the subject of considerable research[23–26].

Nonetheless, from the schematic in Figure 1, we can understand the basic problem it introduces to analysing global temperature trends. If a weather station was initially located in a rural (or even modestly urbanized) area, but over the years, the surrounding area became more urbanized, then the weather station would at some point begin to be affected by the associated urban heat island. This would introduce an artificial warming *"urbanization bias"* into the station records. If the area continues to become more urbanized, this will have a tendency to increase the magnitude of the urban heat island, and as a result, the bias in the station records would become greater over time[27].

Urban areas still only cover $\sim 1\%$ of the Earth's land surface area, so genuine global temperature trends are probably not seriously affected by the increases in urban heat islands. However, many of the longest and best maintained weather records are those kept in or near urban areas. This is partly because, before recent advances in automation, weather stations needed a staff to make measurements, and

References	Study title	Section
Hansen & Lebedeff, 1987[13]	Global trends of measured surface air temperature	4.1
Wigley & Jones, 1988[14]	Do large-area-average temperature series have an urban warming bias?	4.2
Jones et al., 1990[15]	Assessment of urbanization effects in time series of surface air temperature over land	4.3
Easterling et al., 1997[16]	Maximum and minimum temperature trends for the globe	4.4
Peterson et al., 1999[17]	Global rural temperature trends	4.5
Peterson, 2003[18]	Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found	4.6
Parker, 2004[19] and 2006[20]	Large-scale warming is not urban; A demonstration that large-scale warming is not urban	4.7
Efthymiadis & Jones, 2010[21]	Assessment of maximum possible urbanization influences on land temperature data by comparison of land and marine data around coasts	4.8
Wickham et al., 2013[22]	Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications	4.9
Hansen et al., 1999[11]; 2001[12]; and 2010[3]	GISS analysis of surface temperature change; A closer look at United States and global surface temperature change; Global surface temperature change	4.10

Table 2: *Studies which have concluded urbanization bias only has a small or negligible effect on global temperature estimates.*

maintain equipment. They therefore have tended to be located in areas which are relatively easy to access, i.e., close to where people live. For instance, Ren et al., 2008 note that in China, although there are some *relatively* remote stations, the observers of these *"arduous stations"* (and their families) usually live in towns or cities. So, even these stations tend to be *"...located in small towns or some sites with better traffic conditions."*[28].

Hence, many of the stations used for global temperature estimates have been exposed to increasing urbanization over the period covered by their records. For this reason, it is likely that urbanization bias has inadvertently introduced artificial warming trends into the current weather station-based global temperature estimates. If so, they would not be representative of actual global temperature trends[29–34].

Many researchers have suggested that we should be detecting *"anthropogenic global warming"*[1] due to an increase in atmospheric carbon dioxide concentrations[35]. The anthropogenic global warming *"hindcasts"*[2] of the current Global Climate Models (GCMs) bear some similarity to the current global temperature estimates[36]. This has led to the popular assumptions that (a) most of the global temperature trends of recent decades are a result of anthropogenic global warming, and (b) the Global Climate Model projections for future global temperature trends are somewhat reliable.

Global Climate Model projections have been used to justify major policy changes on an international basis, e.g., Ref. [37]. However, if it turns out that a significant fraction of the apparent global warming in recent decades is due to urbanization bias, then the hindcasts of the Global Climate Models which had been thought to have been successful, would actually have been unsuccessful. This would raise serious doubts over the supposed reliability of the models. In turn, this might remove the justification for those policies which have been based on the models. With this in mind, it is important to carefully consider the urban heat island problem.

Figure 2 shows each of the global temperature estimates[3] from 1880 to present of Table 1, which do not

[1]The synonym "man-made global warming" is sometimes used. Unfortunately, the more general terms "global warming" and "climate change" are sometimes **mistakenly** treated as synonyms for anthropogenic global warming.

[2]A hindcast is the opposite of a forecast, i.e., a retrospective "prediction" of what was expected to have occurred in the past.

[3]Since 2006, the Japan Meteorological Agency only publish their land-and-ocean global temperature estimates, but we

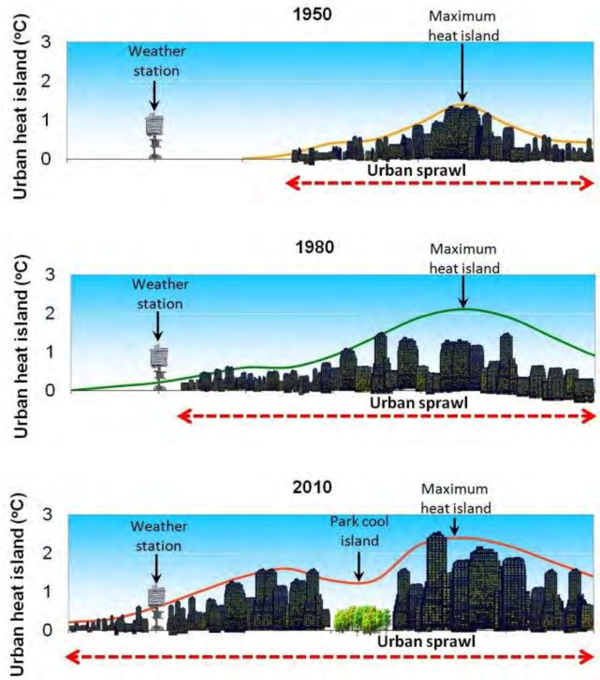


Figure 1: *Schematic diagram of urban heat island development at a hypothetical weather station which was initially rural, but became surrounded by urban sprawl from a neighbouring town over time. The curved lines suggest the magnitude of the urban heat island at each location.*

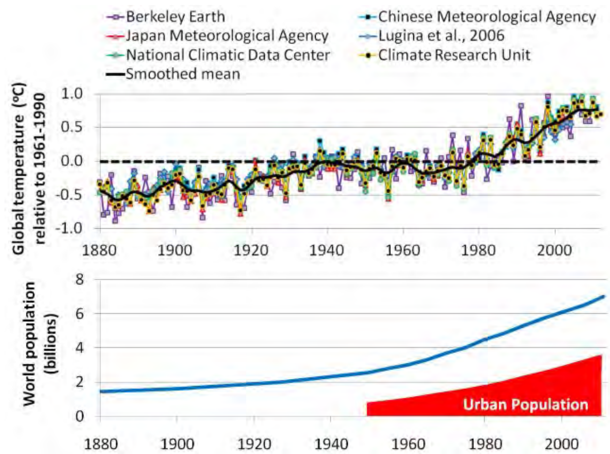


Figure 2: *Comparison between the various (land-only) global temperature estimates which do not correct for urbanization bias, relative to 1961-1990 (top) and the world population growth since 1880 as well as the urban population growth since 1950 (bottom). The solid black line represents the 11-point binomial smoothed mean of all global temperature estimates. World population figures taken from About.com Geography and urban population figures taken from U.N. Population Division).*

explicitly correct for urbanization bias, i.e., all except the Goddard Institute of Space Studies estimate which we discuss separately in Paper II[1]. Figure 2 also shows the world population (from 1880 on), and the urban population (for 1950 on). We see that, coinciding with an apparent "global warming" trend, there has been a substantial increase in population. Particularly since the mid-20th century, this population growth has increasingly been in urban areas. In other words, the world's population is becoming increasingly urbanized.

Since the longest and best-kept station records tend to be those located in or near human settlements, many of the weather stations used for constructing the global temperature estimates of Figure 2 will have witnessed at least some degree of urbanization over the course of their record. Therefore, it is plausible that a substantial fraction of the apparent "global warming" of Figure 2 is simply an artefact of urbanization bias.

were able to graphically estimate their land-only values from their July 2005 Tokyo Climate Center Newsletter (Issue No. 1). The China Meteorological Agency's estimate is digitized from Xu et al., 2014[9] (under CC BY-NC-ND).

2.1 Current approaches to measuring urban heat islands

A number of approaches have been taken to identify the extent and magnitude of urban heat islands at individual urban areas[38, 39]. One approach is to traverse a rural-urban area with vehicle-mounted thermometers[40]. Another approach is to temporarily install a number of thermometers at fixed locations throughout the area[41, 42].

The transect approach is relatively quick and straightforward, however it can only give a once-off snapshot in time. Weather station-based studies are typically concerned with annual (or monthly) average temperatures. Therefore, measurements made at one (or even several) times of the day on one day (or several) of the year are only qualitatively of relevance, i.e., they do not indicate how the average annual temperature is affected. However, for the fixed location approach, care is required to ensure that the thermometers remain intact, and adequately exposed for the duration of the study, something which may be difficult in busy urban areas.

Another approach that has become popular with the advent of thermal remote sensing technology is the use of satellite and aircraft imagery[43–46]. These studies, like the vehicle transect studies, typically only provide a once-off snapshot in time. Having said that, with a long enough period of satellite data[47], or repeat studies, it may be possible to study urban heat island evolution. However, remote sensing thermal images are not always good at determining the magnitude of the effect at ground level [48, 49].

Another common approach to estimating urban heat islands is to directly compare thermometer records from weather stations based in an urban area to those in neighbouring rural (or failing that, less urbanized, e.g., suburban) areas. This approach is not as useful for determining the maximum urban heat island effect[50] or its spatial distribution for a given urban area [41]. Since these effects are often very large at their peak, these issues are important[4] for governments and urban/town planners [51]. Areas with large maximum urban heat islands are particularly susceptible to heat waves, which may cause serious social as well as health problems[52]. However, since we are considering the biases introduced to weather station records, in this article, weather station-based studies are actually of more relevance to us.

Unfortunately, a serious problem with most of the weather station-based studies has been the lack of consistency between different studies, making it hard to directly compare them [18, 26, 38, 39, 53]. Indeed, often, while the "rural" stations used for some studies are less urban than the "urban" stations, they are more urbanized than the "urban" stations from other studies [39, 53, 54]. In this case, the calculated urban heat island would only be a **relative** urban heat island, and therefore lead to an underestimation of the actual urban heat island.

In an attempt to overcome this lack of consistency, Stewart & Oke have proposed defining stations as being in different *"thermal climate zones"* [26, 39, 53, 54], rather than sticking rigidly to the limited "urban-rural dichotomy"[53]. Unfortunately, only a few researchers have so far adopted this more flexible approach[55]. As a result, most of the discussion in this article will be limited to "urban"-"rural" comparisons.

[4]See http://www.urbanheatislands.com/ for some discussion.

2.2 Change is more relevant than magnitude

Global temperature estimates are calculated by averaging together the relative changes in temperature over time for individual stations. As a result, it is the *changes* in average temperature, rather than the absolute values of average temperatures which are used in the calculations.

Jones et al., 2008[56] have argued that while urban heat islands in long-established urban areas might be large in some cases, they might not have changed much in recent decades. They based their argument on an analysis of two European cities (London, U.K. and Vienna, Austria). It is a plausible argument. London's heat island was already substantial in the early 19th century[10], so it is possible that its urban heat island development may have slowed. There are, however, a number of serious flaws in Jones et al.'s analysis.

For their London analysis, they explicitly relied on the assumption that the Rothamsted station is a "*truly rural site*" and therefore unaffected by urbanization bias. Parker et al. also made this assumption for constructing their "Central England Temperature" composite dataset[57, 58]. Rothamsted station[5] is on the grounds of an agricultural research station (Rothamsted Research), and so its immediate microclimate is that of a field surrounded by trees. However, those grounds are surrounded by the town of Harpenden, Hertfordshire (current population $\sim$ 30,000). So, the *"truly rural"* Rothamsted station is actually quite urbanized.

For their Vienna analysis, they considered the Hohe Warte station to be urban, but claimed that its record showed "excellent agreement with its rural neighbours", and extrapolated from this claim to the conclusion that the urban heat island in Vienna had not changed in recent decades. However, Böhm, 1998 had found significant urbanization bias for the Hohe Warte record (which, by the way, he regarded as "suburban", and not a city site) over the period 1950-1995[59].

Jones et al. also claimed that their conclusion based on their London and Vienna analyses was in agreement with Gaffin et al., 2008's study of the Central Park station in New York, U.S.A.[60]. However, Gaffin et al. had actually found that urbanization bias was "*responsible for* $\sim 1/3$ *of the total warm-*

[5]Located at approximately 51.82°N, 0.37°W according to e-RA: the electronic Rothamsted Archive.

ing the city has experienced since 1900"[60]. That suggests quite a substantial urbanization bias. In addition, we suspect that Gaffin et al., 2008 underestimated the true magnitude of the New York urban heat island, since their estimate was based on the explicit assumption that the rural/suburban neighbouring station records they were using for comparison had been adequately corrected for urbanization bias, but the neighbouring areas they were using had themselves undergone considerable urbanization[52].

Nonetheless, while the Jones et al., 2008 study may have been seriously flawed, their conclusion, by coincidence, seems to be valid. While Böhm, 1998 had found evidence of some urban warming[59] at some of Vienna's stations, he found that two of the downtown stations showed little change. In other words, as Jones et al., 2008 had argued, the change in urban heat island at a station may be relatively small, if the area was already highly urbanized when the station was set up.

On the other hand, a suburban or even relatively rural area might currently only have a slight urban heat island, but if it only started expanding in the last few decades, then this could introduce a strong warming bias into the station records[61]. Mohsin & Gough, 2010 [62] have suggested this may be happening in the Greater Toronto Area in Canada, since in recent years the increase in urban warming appears to have been greater at suburban stations than in downtown stations.

This leads to a tricky complexity for studying urbanization bias. The station records which are worst affected by urbanization bias are *not necessarily* from the most urbanized stations. Instead, the issue is the amount of urbanization experienced by the station *over the period covered by the record.*

2.3 Evidence for urban heat island effects at individual stations

As described in the previous section, the existence of urban heat islands does not in itself mean that global temperature estimates are seriously affected by urbanization bias. But, there is considerable evidence that urban warming has introduced significant biases into the weather records of a number of urban and semi-urban stations at least.

Studies of urban heat islands are numerous, as a search with a journal search engine, e.g., Google Scholar, for key words such as "urban heat island" will reveal. It is beyond the scope of this article to review all such studies. But, it may be helpful to briefly mention in this section a few representative studies of urbanization bias across the globe.

Much of the early work on urban heat islands has been based in Europe, starting with Howard's 19th century study of London, U.K.[10]. For example, as mentioned in the previous section, Böhm, 1998 noted that even for Vienna, a city that had a population decline in the latter half of the 20th century, some urbanization bias was found[59]. As another example, Moberg & Bergström, 1997 found evidence of urbanization bias in the important long records for the Swedish towns of Uppsala and Stockholm[63].

There have also been a large number of studies investigating urbanization bias in North America. For instance, Goodridge, 1996[64] divided up weather stations for the U.S. state of California into three subsets depending on population size. The more urbanized the subset, the greater the warming trends were. A later study by LaDochy et al., 2007 also found a similar result for the same region[65]. Hinkel & Nelson, 2007 [42] found that even for the small village of Barrow, Alaska (USA), with a population of $\sim 4,500$, there was a substantial urban heat island between the village and the surrounding tundra. Although, the type of urban heat island detected in high-latitude, permafrost areas such as Barrow probably differs from villages in more temperate climates.

Compared to North America and Europe, urban heat island studies for the rest of the world have been less prevalent. But, this does *not* mean that urban heat islands are confined to these areas - often it is just a consequence of the fact that many urban climatologists are based in North America or Europe. In recent decades, other areas have started to receive more attention. Roth, 2007 has reviewed recent studies of urban heat islands in tropical and subtropical areas[66]. In particular, many parts of Asia have undergone considerable urbanization in recent decades. For instance, Kataoka et al., 2009 found strong urban warming trends in a number of large Asian cities[67]. The Middle East has also been affected, e.g., Saadatabadi & Bidokhti, 2011 found substantial bias for an urban station in Tehran, Iran[68].

Unfortunately, studies in the Southern Hemisphere are still relatively infrequent. But, a number of studies have also found examples of urbanization bias there. For example, Coughlan et al., 1990[69] found evidence of substantial urbanization bias in records for large city stations in Australia, while Hughes & Balling, 1996[70] found that much of the apparent warming in South Africa was urbanization bias (during the period 1960-1990 at least).

2.4 Evidence that global temperature estimates are biased by individual station records

In the previous section, it was shown that urbanization has significantly affected a number of weather records across the world. Satellite studies also suggest that substantial urban heat islands are found in urban areas across the globe[44, 46], suggesting that these effects are not just confined to a few countries. However, this, on its own, does not necessarily mean that the *global* temperature estimates are themselves significantly biased. Most of the global temperature estimates use records from several thousand stations. If only a small percentage of those stations are significantly biased, then the overall bias to the global estimates might be quite small - even if individual biases are relatively large.

Unfortunately, most estimates suggest that at least one third to one half of the stations are urbanized to some extent. In addition, the alleged "global warming" trends of Figure 2 are only of the order of a degree Celsius per century. So, even if biases are only of the order of a few tenths of a degree Celsius per century at individual stations, this could still be enough to substantially bias the global temperature estimates.

Because there are so many confounding factors involved in analysis of weather station records, it is not a trivial matter to construct a reliable estimate of the overall urbanization bias in global temperature estimates. For instance, reliable rural stations with long records are often rare in areas which have undergone a lot of urbanization, yet these are the regions most likely to be seriously affected. In addition, weather station records can also be affected by a number of different biases aside from urbanization, such as the siting biases we discuss in Ref. [71]. Depending on the net sign of these biases, this could easily lead to a substantial under- or over-estimation of the true urbanization bias.

As a result, some articles limit themselves to establishing the *plausibility* of urbanization bias, rather than quantifying it. This is often done from a general assessment of the literature. Wood, 1988[33] was one such article, and we will discuss Wigley & Jones' attempted rebuttal[14] of it in Section 4.2, while Idso & Singer, 2009's review[72] includes another.

Other studies attempting to quantify the effects on global temperature estimates have limited themselves to detailed analysis of specific regions. The problem with such regional analyses is that it is possible that the researchers may inadvertently select a region which has a particularly high or low level of bias. Still, if a study shows that urbanization biases are significant for large regions (rather than just a handful of stations), this would provide a strong indication that global analyses are also affected.

Most of these regional studies have been limited to analyses of the U.S.[29–32, 64, 73–81]. This should not be mistaken for implying that urbanization bias is a problem mostly limited to the U.S. On the contrary, as we discuss in Paper III [2], the number of rural stations with long, continuous records used in current global temperature estimates is surprisingly low for regions outside of the U.S. For example, only 8 of the 173 stations in the Global Historical Climatology Network (4.6%) which are identified as rural in terms of both population and night-light intensity and have data for at least 95 of the last 100 years are outside of the U.S.

Indeed, although three of the 30 largest urban agglomerations are currently located in the U.S. (New York, Los Angeles and Chicago)[82], the U.S. is a relatively rural country with an average population density of only 32.2 km^{-2} compared with the world average of 50.6 km^{-2} (as of July 2010)[83]. Instead, it appears that studies of urbanization bias in the U.S. dominate the literature mainly through convenience. Considerable effort has been made in compiling and archiving a large number of rural station records for the contiguous U.S.[32], meaning that direct urban-rural comparisons are considerably easier for the U.S.

In contrast, countries such as China have a severe shortage of long, continuous records for rural regions, and as a result most comparisons are between highly urbanized and moderately urbanized areas[28, 84]. Even still, a number of studies of temperature trends in China indicate considerable urbanization bias in recent decades, e.g., Refs. [27, 28, 84–88] (see Yang et al., 2011 for a review and discussion[88]).

As mentioned in the previous section, Hughes & Balling, 1996 found evidence that gridded estimates for South Africa were biased by urban warming[70]. Englehart & Douglas, 2003[89] found urban stations in Mexico were also substantially affected by urban warming, and they cautioned that almost all of the stations in the Mexican observing network are located in highly urbanized areas.

Fujibe & Ishihara[7, 90–92] have found evidence of substantial urbanization bias in many Japanese station records. Japan is a highly urbanized country - in 2010, Japan had an average population density of 334.9 km^{-2}, nearly 7 times the world average and more than 10 times that of the U.S.[83].

While nearly half of Japan is relatively rural, with a population density of <30 km^{-2}, there are very few long-term weather stations located in those areas. More than half of the long-term stations are located in urban areas with a population density of >1,000 km^{-2}[91]. Fujibe, 2009 attempted to overcome this problem by considering a network of automatic weather stations which were installed in the late 1970s. 9.6% of those stations were located in rural locations, and a further 18.2% of the automatic stations were in lightly urbanized areas with population densities of 30-100 km^{-2}. Stations of all levels of population density showed warming over the 1979-2006 period, but stations located in areas with a population density of >1,000 km^{-2} showed more than 50% extra warming than those with a population density of <30 km^{-2}, and even the stations which were only lightly urbanized (30-100 km^{-2}) showed more than 10% extra warming[91].

Fujibe, 2009's analysis suggests that urbanization bias has significantly overestimated the post-1970s warming trends in Japan. However, as the analysis only began with the implementation of the automatic weather station network in the late 1970s, it does not tell us whether it also underestimated the 1940s-70s cooling which seems to have occurred elsewhere[93], or how Japanese temperatures during the early 20th century warm period compared to the recent warm period.

In a follow-on study, Fujibe & Ishihara, 2010[7] compared a network of 17 of the *least* urbanized Japanese stations with continuous records for the 20th century (as used by the Japan Meteorological Agency) to the Climate Research Unit and Goddard Institute for Space Studies' gridded estimates for the same region. They found that the Climate Research Unit's estimate implied more warming, suggesting urbanization bias. The Goddard Institute for Space Studies' estimate, which includes an urbanization adjustment, in contrast, had a similar long-term trend to Fujibe & Ishihara's network[7]. The 20th century trend for the Japan Meteorological Agency's 17-station network was similar to that of the *global* temperature estimates. On this basis, Fujibe & Ishihara assumed that their network was relatively free of urbanization bias. However, this assumes that the global estimates are themselves relatively free of urbanization bias, i.e., the claim we are disputing in this article. With this in mind, it is worth noting that 10 of the 17 Japan Meteorological Agency stations were in areas with a population density of >1,000 km^{-2}[7], suggesting that even the Fujibe & Ishihara network was itself significantly affected by urbanization bias.

Karl et al., 1988[32] developed a high density, mostly rural, network of stations for the contiguous U.S. (the National Climatic Data Center's United States Historical Climatology Network), which allowed them to carry out a reasonable estimate of urbanization bias for stations within the U.S. They approximated urban development using population size, and developed a population-based adjustment to approximately remove urban bias.

Karl et al. found that urban bias was detectable even for small towns in the U.S., suggesting that the urbanization bias in global temperature estimates could be hard to totally remove. An independent analysis by Balling & Idso, 1989 confirmed this[34]. Although Karl et al. found there was only a slight total urban bias in their mostly rural dataset, much of the warming for the contiguous U.S. in two of the global temperature estimates at the time seemed to be due to urban bias[73, 94]. This suggests that much of the warming in other regions could also be due to urban bias.

Spencer, 2010 presents an analysis on his website[95] where he compared a large number of station pairs (elevation adjusted) on the basis of local population density and average temperature in 2000. He found that stations with a higher population density were warmer on average. If urbanization is assumed to be a function of population density, then this would imply that increasing urbanization is leading to urbanization bias. He found the increase to be greatest for low population densities, i.e., the urban heat island *growth* is greater for a rural station becoming slightly urbanized than for an urban station becoming slightly more urbanized.

2.5 General urban-related bias in global temperature estimates

Rather than estimating urbanization bias by attempting to individually compare urban and rural stations, some groups have compared entire global or regional temperature estimates to other factors[61, 79, 96–102].

Kalnay & Cai, 2003[79] compared U.S. temperature trends calculated from weather station records to temperature trends extrapolated from the National Centers for Environmental Prediction/National Center for Atmospheric Research (or "NCEP-NCAR") reanalysis dataset. This reanalysis dataset did not use surface thermometer records, but rather estimated surface temperatures using weather forecasting models based on weather balloon and satellite atmospheric measurements. For this reason, it should not be affected by urbanization bias. Kalnay & Cai found that the weather station warming trends were significantly greater than the equivalent reanalysis trends (about 44%). They suggested that some or all of this extra warming was due to urbanization bias (or changes in land use).

This study was quite controversial, and led to considerable debate. Trenberth, 2004[103] criticised the reanalysis dataset used by Kalnay & Cai, 2003[79] because it was based on actual physical weather measurements and therefore did not explicitly include the extra warming predicted by anthropogenic global warming theory in between data-points. It is unclear why Trenberth believed theoretical predictions should necessarily be more reliable than experimental measurements. Nonetheless, Cai & Kalnay[104, 105] replied that any warming which occurred as a result of anthropogenic global warming should still be detected by the weather measurements.

Vose et al., 2004[106] disputed Kalnay & Cai's analysis because, when they used the National Climatic Data Center's homogeneity adjusted records for the U.S., they found even more warming than with the unadjusted records Kalnay & Cai had used. This suggested the bias in the National Climatic Data Center's data was even greater than Kalnay & Cai, 2003 had suggested. However, Vose et al. believed their adjustments were reliable, and so argued that the reanalysis dataset had to be at fault instead. In Paper III, we describe problems with the National Climatic Data Center's adjustments [2], so this is not a convincing argument. Cai & Kalnay, 2004[104] acknowledged that the National Climatic Data Center's adjustments for the U.S. led to more warming, but noted that this only increased the probable magnitude of the urbanization/land use bias.

Simmons et al., 2004[107] compared the NCEP-NCAR reanalysis to another reanalysis - the European Centre for Medium-range Weather Forecasts (or ECMWF)'s 40 year Re-Analysis dataset (referred to as the ERA-40 for short). This other reanalysis did not show the same discrepency with weather station-based estimates, and they therefore concluded that the NCEP-NCAR reanalysis (and, hence, Kalnay & Cai's analysis) was unreliable[107]. However, unlike the NCEP-NCAR reanalysis, the ERA-40 reanalysis incorporates weather station measurements, and so cannot be used as a completely independent test of weather station measurements.

Follow-up studies by Kalnay et al.[96–98] have found that urbanization/land use biases are not just confined to the U.S. but are a global problem.

McKitrick et al.[99–101] found significant correlations between temperature trends and various different socio-economic factors, both for the U.S.[99] and the rest of the world[100, 101]. If the temperature trends were only capturing climatic signals then these correlations should not exist. Similarly, de Laat & Maurellis, 2006[102] found that warming trends were greater in industrialized areas than elsewhere.

Both McKitrick et al. and de Laat & Maurellis' studies have been criticised[22, 108–110]. Wickham et al., 2013 criticised the McKitrick et al. studies because they had used national country-averaged population figures for estimating the population of individual grid-boxes, rather than gridded population estimates[22]. This means that their population estimates of some grid-boxes would be underestimated, while other grid-boxes would be overestimated. Wickham et al., 2013 suggested that this crude approximation makes the findings of McKitrick et al. unreliable[22]. However, it appears to us that the crudeness of the approximation should have, if anything, *reduced* the signal-to-noise ratio of McKitrick et al.'s analysis, so the fact that McKitrick et al. were still able to detect significant correlations using nationally-averaged population figures, actually strengthens their conclusions.

The main problem with the McKitrick and de Laat et al. studies is that there are a number of different confounding variables involved, and it is not a trivial matter to separate them. Still, McKitrick et al. have attempted to address these concerns[61, 100, 101, 111, 112]. So, while controversial, their findings may have merit.

Even if their studies are only partially valid, then this suggests that global temperature estimates are significantly contaminated by non-climatic signals such as urbanization bias. This would agree with the suggestions of the studies discussed in the previous section.

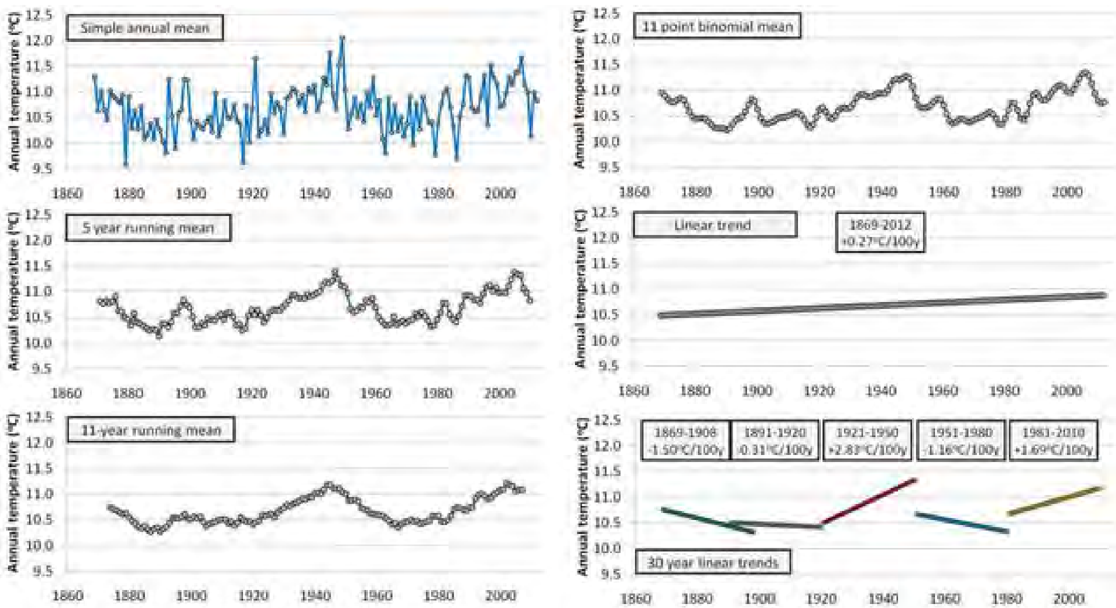


Figure 3: *Several different ways of describing the temperature trends for the rural Valentia Observatory (Ireland) weather station. Annual temperature data is taken from the Global Historical Climatology Network Version 3 unadjusted dataset. Two missing monthly means (June 2004 and July 2012) were added from monthly reports downloaded from the Met Éireann website.*

3 Flaws common to more than one study

3.1 Using linear trends to describe non-linear data

Most of the studies consider linear trends as part of their analyses, i.e., Wigley & Jones, 1988[14]; Jones et al., 1990[15]; Easterling et al., 1997[16]; Peterson et al., 1999[17]; Parker 2006[20]; Efthymiadis & Jones, 2010 [21]; Wickham et al., 2013[22] and the Hansen et al., 1999-2010 studies[3, 11, 12]. In the case of Jones et al., 1990, Efthymiadis & Jones, 2010 and Wickham et al., 2013, it comprises a major part of their analyses.

While linear trend analysis offers a convenient method for describing time series which have linear trends, it can result in misleading, or even invalid conclusions when it is applied to time series that have *non-linear* trends. Temperature records for weather stations frequently show non-linear trends, especially the longer records. For this reason, an over-reliance on linear trend analysis when assessing temperature trends (whether local or global) is unwise.

The problem of using linear trends to describe non-linear trends can be seen from Figure 3, which shows several different methods for describing the temperature trends of the Valentia Observatory station. Valentia Observatory, Ireland has one of the longest and most complete temperature records for a rural station. If we just consider the annual means (top left panel), it is apparent that there is a lot of variability from year to year.

The annual mean already involves a considerable amount of averaging, since it comprises the mean of the 12 monthly averages, the monthly averages are means of the daily averages and the daily averages are themselves estimates of the mean temperature for each 24 hour period (often this is calculated as the simple mean of the maximum and minimum temperatures recorded on a minimum-maximum thermometer). However, because the variability from year-to-year is quite large, it can be difficult to establish what *long-term* trends there are, if any. For this reason, researchers often apply further averaging (or *"smoothing"*) routines to temperature records.

A common smoothing technique is to calculate a *"running mean"* using a fixed number of years (sometimes called a *"boxcar average"*). The five-year and 11-year running means of the Valentia Observatory record are shown in Figure 3. Running means are calculated by replacing the value for a given year with the mean value over the period starting a fixed number of years before the given year and ending that same fixed number of years after the given year. This has the effect of making consecutive years seem quite similar to each other, i.e., it reduces the inter-annual variability. Hence, long-term trends are more apparent.

One problem with running means is that they can *artificially* create the appearance of long-term trends which might not exist. For instance, if a few years had anomalously high (or low) mean temperature, then this would increase (or decrease) the values of the temperatures for several years before and after this anomalous period, creating the impression of a gradual trend over a long period. In order to reduce the magnitude of this statistical artefact, while still reducing the inter-annual variability enough to consider long-term trends, one approach is to apply *"binomial smoothing"*. Like a running-mean, this approach also involves averaging all values over a fixed period. However, in the binomial mean, the weights the neighbouring years contribute to the average are reduced the further away in time they are from the target year. This approach accentuates long-term trends (low-frequency information) without losing all the inter-annual variability (high-frequency information). For our analysis in this paper, we often consider the 11-point binomial means of temperature trends.

From Figure 3, both the running means (5-year and 11-year) and the 11-point binomial mean suggest that there have been a number of "warming" and "cooling" trends at Valentia Observatory, which have each lasted a few decades, since the start of the record in 1869.

Whether these trends are climatic in nature or not, due to the alternation between cooling and warming periods, the temperatures of the last few decades seem neither unusually warm nor unusually cold. Although there has been "warming" since the 1970s, it followed "cooling" since the 1940s. However, if the linear trend is calculated (by linear least squares fitting) over the entire record (1869-2012), it incorrectly suggests there has been a continuous "warming" trend of $+0.27°C$/century. In other words, if a researcher just relied on a linear trend for their analysis, they would fail to notice the actual multi-decadal alternation between cooling and warming, as well as the considerable inter-annual variability.

More worryingly, both the sign and the magnitude of the linear trend depend on both the length of the period and the starting point. This can be seen by comparing the 1869-2012 linear trend for Valentia Observatory (middle right panel) with the various 30-year linear trends in the bottom right panel. In other words, linear trend analysis can provide very inconsistent results when applied to data with non-linear trends.

For this reason, the over-reliance of many of the studies on linear trend analysis may have led to invalid conclusions.

3.2 Assuming rural station records have no non-climatic biases

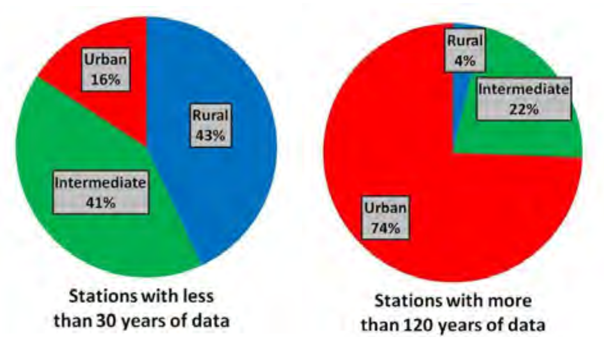


Figure 4: *Breakdown of the degree of urbanization of the station records in the Global Historical Climatology Network (version 3, unadjusted) of the stations with the shortest (left) and the longest (right) records.*

Initially, when considering the urbanization bias problem, one might suppose that a simple solution for estimating global temperature trends would be to construct an estimate using only rural stations. This *"rural sub-setting"* approach has formed the main basis for many of the studies considered here, i.e., Hansen & Lebedeff, 1987[13]; Easterling et al., 1997[16]; Peterson et al., 1999[17]; Wickham et al., 2013[22] and the Hansen et al., 1999-2010 studies[3, 11, 12]. Jones et al., 1990 also used rural sub-setting for their analysis, but they were analysing regional trends, rather than global trends[15]. A similar philosophy was also implicit in the design of the Peterson, 2003[18] as well as the Parker, 2004[19] and 2006[20] studies.

The main problem with this approach is that the available rural station records tend to be of a lower

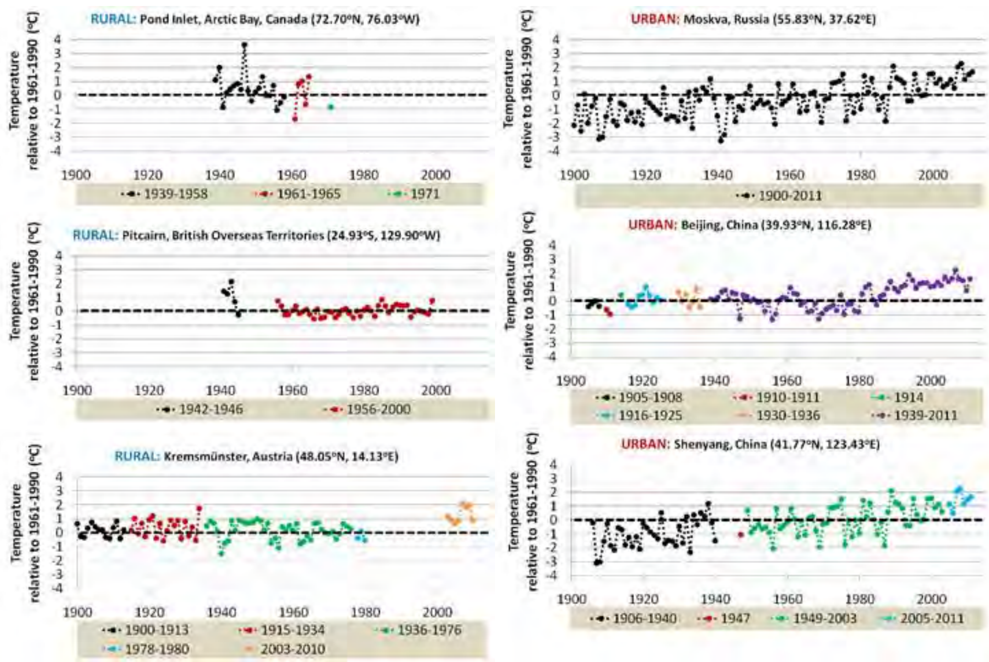


Figure 5: *Annual temperature trends since 1900 for some typical rural (left) and urban (right) stations in the National Climatic Data Center's Global Historical Climatology Network Monthly (unadjusted) Version 3 dataset. Each colour corresponds to a continuous segment of the record.*

quality than their urban counterparts. Rural station records tend to have less data, e.g., only about 4% of the stations in the Global Historical Climatology Network with more than 120 years of data are identified as rural in terms of both population and night-light intensity, while about 74% are identified as urban by both metrics. Correspondingly, only 16% of the stations whose records have less than 30 years of data are urban by both metrics - see Figure 4.

In addition, as we discuss in Paper II[1], rural records are more likely to have large data gaps, which frequently coincide with abrupt non-climatic changes in reported temperatures. The extent of this can be seen by comparing the rural and urban records in Figure 5, which are, in our opinion, fairly representative of the types of records found in the popular Global Historical Climatology Network. While all three of the urban records shown have strong warming trends, it is difficult to identify any common trend in the three rural records. Indeed, there is very little overlap between the records. If the longer rural records include even a few non-climatic biases, this could easily bias the apparent long term trends of global temperature estimates constructed from a rural subset.

As a result of these problems, the rural records available in the current datasets are unfortunately rather limited for calculating global temperature trends for longer than a few decades. However, aside from Wigley & Jones, 1988 who caution that *"(u)nfortunately, rural data series are themselves subject to various non-climatic effects, unrelated to urban warming"*, most of the studies seem to assume that the only difference between rural and urban station records is urbanization bias. This assumption comes in at least three forms:

1. In the case of Hansen & Lebedeff, 1987[13], it appears to have been an implicit assumption, since they did not consider other non-climatic biases in their rural sub-setting experiment.

2. Some studies acknowledge that other non-climatic biases could alter global temperature trends, but assume that the occurrence of these biases is comparable in both rural and urban stations, e.g., Hansen et al., 1999 assume that *"the random component of* [other biases] *tends to average out in large area averages and in calculations of temperature change over long periods"*[11].

3. Other studies agree that non-climatic biases (other than urbanization bias) could theoretically be a problem, but believe that they have successfully removed them by applying *"data homogenization"* techniques.

Amongst those studies which use data homogenization, the homogenization techniques applied vary. Easterling et al., 1997[16] and Peterson et al., 1999[17] used versions of the National Climatic Data Center's Global Historical Climatology Network datasets which had been adjusted using a "step-change" homogenization. Wickham et al., 2013[22] also used a "step-change" homogenization process, and used a weighting procedure which they believed would minimise the impact of non-climatic trend-changes[113].

We argue in Paper III[2] as well as in Ref. [71], that these particular homogenization methods often lead to a "blending" of non-climatic biases between stations, rather than their removal. A consequence of this blending is that the non-climatic biases become spread so uniformly amongst neighbouring stations that they can no longer be identified by comparing stations. If the stations are homogenized in this way, it would seriously hinder any attempts to accurately identify urbanization bias using a rural sub-setting approach.

Implicit in the argument of rural sub-setting studies which use homogenization to account for non-urbanization biases, is the recognition that such biases occur with a different frequency or magnitude for rural stations than for urban stations. Otherwise, it would not be necessary to homogenize the data, since the difference between the unhomogenized urban and rural stations would on average be mostly due to urbanization bias. So, for rural subsetting studies which use homogenization, the argument shifts from assessing the magnitude of urbanization bias in global temperature estimates (the focus of this article) to assessing the robustness of homogenization techniques. We question the reliability of these techniques in Paper III[2].

3.3 Assuming evidence for "global warming" negates the urbanization bias problem

It is often argued that urbanization bias is at most a minor problem for weather station estimates of global temperature trends because there are other indicators of "global warming" which would be unaffected by urbanization bias. For instance, in his rebuttal of de Laat & Maurellis, 2006[102] and McKitrick & Michaels, 2007[100], Schmidt, 2009 claims that *"there is significant independent evidence for warming in the oceans, snow cover, sea ice extent changes, phenological records etc. which are consistent with the land station analyses"*[110]. This common argument misses the point, however. The problem is not in establishing whether or not there have been periods of "global warming", but rather establishing *how much* of the calculated global temperature trends are non-climatic warming trends caused by urbanization bias.

The data sets described above by Schmidt, 2009 do suggest that there has been *some* global warming since the 1970s. *Qualitatively*, this is in agreement with the weather station-based global temperature estimates of Figure 2. However, this does *not* mean that the weather station-based estimates are unaffected by urbanization bias.

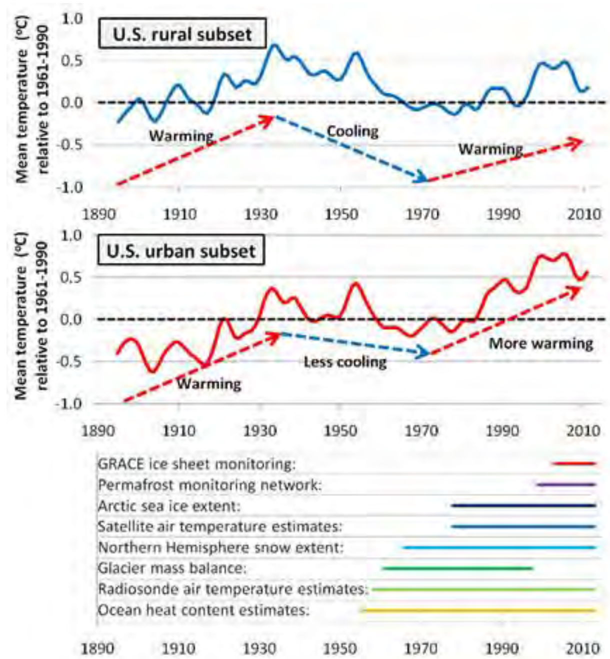


Figure 6: *Gridded mean temperature trends for the subset of most rural stations (top) and most urban stations (middle) in the United States, with 11-point binomial smoothing. The bottom panel shows the time periods covered by the various "global warming indicators". Individual station data is taken from the United States Historical Climatology Network Version 2 unadjusted dataset.*

Figure 6 compares the gridded mean temperature

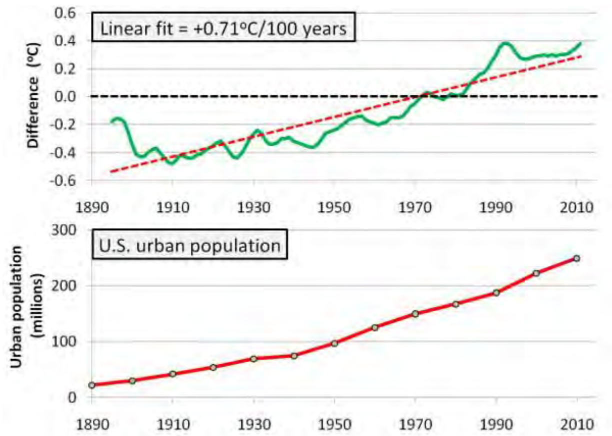


Figure 7: *Comparison of the difference between the urban and rural U.S. temperature trends of Figure 6 (top) and the urban population growth for the U.S. (bottom), as determined from U.S. Census figures (Table 7 of Ref. [114], downloaded from* `http://www.census.gov`*).*

trends of the most rural and most urban stations in the United States Historical Climatology Network (version 2, unadjusted) dataset, which we determined for Paper III[2]. Both subsets agree that there was a warming trend in the U.S. from the 1970s to 2000s. However, in the urban subset, the magnitude of this trend was significantly greater. Both subsets also showed a cooling trend from the 1930s to 1970s. But, in the urban subset, this cooling trend was significantly reduced. This indicates a significant warming bias in the urban subset, relative to the rural subset. From Figure 7, this bias seems to be roughly correlated to the growth in U.S. urban population since the late 19th century. This suggests that the differences between the subsets is urban related, and considering the evidence in Section 2, it seems likely that it is due to urbanization bias in the urban subset.

This difference in trends between the urban and rural U.S. subsets substantially changes the context of the 1970s to 2000s warming. In the urban subset, the 1970s to 2000s warming implies a continuation of a general warming trend since at least the 1890s. This regional warming is consistent with the global warming trends of Figure 2. However, for the rural subset, U.S. temperatures seem to have alternated between periods of warming and periods of cooling. In the rural subset, U.S. temperatures during the 2000s were comparable to those during the 1930s. If urbanization bias is able to introduce such changes to the U.S. temperature trends of Figure 6, then it is plausible that it could have similarly introduced substantial changes into the global temperature trends of Figure 2.

We have heard claims, on numerous occasions, that there are many forms of evidence to corroborate the "global warming" trends of the current weather station-based estimates (i.e., those in Figure 2). With this in mind, we carried out a very careful literature review of the various "global warming indicators" we could find, e.g., Arctic sea ice measurements, northern hemisphere snow extents, ocean heat content estimates. However, remarkably, we were unable to find *any* of these indicators which could be used to conclude that the weather station-based estimates were unaffected by urbanization bias.

As illustrated in Figure 6, most of the datasets used as so-called global warming indicators only have a few decades of data, and so cannot be used to, e.g., compare temperatures in the 2000s to those in the 1930s.

For instance, it is true that the satellite estimates of Arctic sea ice extent suggest a general decrease *"since records began"*[115] (although interestingly not for Antarctic sea ice). However, these records only began in 1978. Similarly, satellite estimates of upper atmospheric air temperatures suggest there has been "global warming", but again these records only began in 1978[116]. The weather balloon estimates of upper atmospheric air temperatures extend further back in time, but only gained reasonable coverage in 1958[117]. Some researchers have tried to estimate climate trends using glacier mass balance measurements, e.g., Dyurgerov & Meier, 2000[118]. However, again, most of the glaciers which have been studied only have a few years of measurements, and before 1957/8 these measurements were only made at a few glaciers[118, 119]. Hence the Dyurgerov & Meier, 2000 study only focused on the 1961-1997 period[118].

The Gravity Recovery And Climate Experiment (GRACE) has allowed some researchers to make detailed measurements of the Antarctic and Greenland ice sheets, e.g., Velicogna, 2009[120]. But, the GRACE satellites were only launched in 2002. Satellite estimates of the northern hemisphere snow cover extent only began in 1966[121] (Rutgers Snow Extent Climate Data Record), while the Global Terrestrial Network for Permafrost, was only established in 1999[122], and the ocean heat content estimates only begin in 1955[123].

This lack of data for the earlier part of the 20th century is understandable since the 20th century coincided with many technological advances. Hence, our ability to monitor the climate system substantially improved over the course of the 20th century. In particular, there was a large increase in climate observation networks around 1957/58, as part of the International Geophysical Year (1957/58)[124]. Similarly, another major increase occurred with the development of satellite technology in the 1960s/1970s. More recently, there have been further improvements in monitoring climate systems during the 1990s and 2000s[6]. However, even though these global warming indicators all agree *qualitatively* with the thermometer-based estimates that there has been "global warming" since the 1970s, they do not tell us what the magnitude of this warming has been. Indeed, it can be seen from Figure 2 of Palmer et al., 2010[123] that, while all of the current ocean heat content estimates suggest a warming trend from the 1970s to 2000s, there is some disagreement over exactly how much.

As another example, a large number of studies of the biological response of different species of plants and animals to annual seasons (*"phenological studies"*) have suggested that the start of spring has been tending to occur earlier in the year in recent decades, e.g., see Ref. [125]. However, again, most of these studies are relatively short and only begin in the 1970s or later, e.g., the Menzel et al., 2006 meta-analysis only considers the 1971-2000 period[126]. Studies which consider longer periods are sometimes more ambiguous, e.g., Kozlov & Berlina, 2002 actually found a decline in the length of the summer season on the Kola peninsula in Russia over the period 1930-1998[127]. In addition, phenological studies do not always provide as straightforward a relationship to global temperatures, as is sometimes assumed. For instance, studies which only consider the first flowering dates of plants can be strongly influenced by changes over the period of the study in either the sample sizes or the sampling frequency[128]. Importantly for our discussion, urbanization bias is known to cause earlier springs in urbanized areas[129].

Moreover, those few indicators which cover longer periods do not necessarily agree with the weather station-based estimates. For instance, Huss et al., 2009 developed a relatively long (94-year) glacial mass balance estimate for four glacial sites in the Swiss Alps, but found that "*Snow and ice melt was stronger in the 1940s than in recent years, in spite of significantly higher air temperatures in the present decade*"[130]. Their air temperature measurements were derived from weather station records. In other words, the mass balance analysis of Huss et al., 2009 suggested that the 1940s were warmer than the weather station records implied.

We did manage to find some indicators which covered a long enough period to assess the long-term trends of the weather station-based estimates - sea surface temperature/marine air temperature estimates; estimates of global sea level trends; studies of glacier lengths; and "multi-proxy" estimates of centennial and millennial temperature trends. It is often suggested that these longer datasets suggest similar global temperature trends to the weather station-based estimates, implying that the weather station-based estimates are reliable, i.e., not particularly affected by urbanization bias. However, a close inspection of these longer datasets reveals that each of them are known to be *also* affected by serious non-climatic biases and/or show substantial differences with the weather station-based estimates when the trends are directly compared.

For instance, the various sea surface temperatures estimates[3, 5, 7, 131] apparently suggest similar trends to the weather station estimates. Indeed, many of the groups using weather station records to construct global temperature estimates also create global "land and sea" temperature estimates by combining their weather station-based "land" estimates with sea surface temperature estimates[3, 5, 7, 131]. As we will discuss in Section 4.8, the Efthymiadis & Jones, 2010[21] study was explicitly based on this. However, these sea surface temperature estimates are known to suffer from a number of serious biases[132–136], particularly for the early 20th century and earlier[5, 131, 137, 138]. It is unclear exactly what corrections need to be applied to them[132]. For instance, compare the 1995 [137] and 2011[138] adjustments proposed by the Hadley Centre group, or consider the discussions of Matthews & Matthews, 2012[135, 136]. Therefore, estimates of long-term temperature trends based on sea surface temperatures need to be treated cautiously.

Some studies have used measurements of glacier lengths as a proxy for global temperatures, e.g., Oerlemans, 2005[139]. Again, these studies suggest that there has been "global warming" since at least the 19th century. However, before the 1950s, most of

[6]Ironically, many of the recent improvements in climate monitoring have arisen out of concern that "global warming" was already occurring, e.g., the founding of the Intergovernmental Panel on Climate Change in 1988 - see http://www.ipcc.ch/docs/UNEP_GC-14_decision_IPCC_1987.pdf.

these measurements were confined to Europe[119], i.e., they do not provide "global" coverage. This is important because many glaciers in Europe are believed to have gone through several periods of glacial advance during the so-called "Little Ice Age"[140]. So, much of the glacial retreat in the 20th century may have involved a reverting to pre-Little Ice Age conditions. Moreover, while global warming could cause glacial retreat, it is not the only cause of glacial retreat[141]. Indeed, even in the absence of any long-term climate change, we should expect glaciers to alternate between periods of glacial advance and glacial retreat[142]. Finally, it is worth noting that even if we ignore these problems, a close inspection of Oerlemans' glacier length-based global temperature estimate actually suggests that the difference between the two 20th century warm periods is less than suggested by the weather station-based estimates (see Figures 3B and S3 of Ref. [139]), i.e., Oerleman's glacier length estimates show noticeably less "global warming" in the 20th century than the weather station-based estimates.

Although most of the available tidal gauges only have data for the past few decades (e.g., since the 1950s), there are enough tidal gauges with relatively long records that a number of researchers have constructed estimates of global sea level changes since the early 20th century or earlier, e.g., Holgate, 2007[143] or Church & White, 2006[144]. Most of these estimates suggest a continuous global sea level rise of a few mm/year since the start of the estimates. This has been used as further evidence of 20th century "global warming", since Global Climate Models predict global warming should cause global sea levels to rise[144].

The first problem with this "global warming indicator" is that sea level rises do not actually prove "global warming". It is true that under global warming, it might be expected that sea levels would rise, e.g., through thermal expansion of the oceans or through extra melting of land ice. However, there are many different factors involved in determining sea levels, and so it is not actually possible to conclusively attribute sea level rises or falls to global warming or cooling. For example, studies such as Chao et al., 2008[145] claim human dam building has led to an underestimate of sea level rises due to global warming, while other studies, such as Wada et al, 2010[146] argue that ground water extraction has led to an overestimate of sea level rises due to global warming.

But, a more challenging problem is that tidal gauges can only be used to measure the *relative* sea level of a location. Stewart, 1989 cautioned that much of the apparent trend in the global sea level estimates constructed from tidal gauges might have nothing to do with climate change (e.g., global warming), but instead be the result of the land to which the gauges are attached moving[147].

For instance, Syvitski et al., 2009 have found that many of the world's largest deltas are subsiding due to local human activity, since they are often densely populated and/or heavily farmed[148]. If the land on which a tidal gauge is located subsides over part of its record, then this would mistakenly create the impression that the local sea levels are rising. If a large number of subsiding gauges are located around the world, this could easily introduce an apparent global sea level rise trend, which is purely an artefact of local subsidence.

Similarly, tectonic activity can introduce substantial biases into tidal gauge records. Quite a few gauges are located in areas which are known to be tectonically active, e.g., western U.S.; New Zealand; the Mediterranean; the Gulf of Mexico. The land on which these gauges are sited may have undergone rises or falls during tectonic events over the course of their records. But, more significantly, many of the gauges are located on land which is near (or on) the boundaries between tectonic plates, e.g., the so-called "Pacific Ring of Fire" which roughly coincides with the Pacific Rim. Even if these gauges are located in regions which have not had many major tectonic events recently, much of the land near these boundaries is gradually rising or falling at rates comparable to the apparent global sea level rise, i.e., a few mm/year.

Several of the tidal gauge regions which might be considered relatively unaffected by tectonic activity are thought to be affected by a phenomenon known as *Post Glacial Rebound* (PGR), e.g., Scandinavia, the British Isles, northeast U.S. During the Last Glacial Maximum ($\sim 20,000 - 25,000$ years ago), many regions which are currently ice-free are thought to have been covered with large glaciers. It is believed that the weight of these glaciers led to deformations in the underlying tectonic plates, and that since they melted, the plates have been gradually readjusting to compensate, again at rates of up to several mm/year. The land in some regions, e.g., areas formerly under glaciers, should be rising (leading to apparent sea level *falls*), while other regions should be falling to compensate (leading to apparent sea level *rises*). Although various *"Glacial Isostatic Adjustment"* (GIA) models have been developed to account for these land changes, they all involve a number of uncertain assumptions, and the pros/cons of the different models are still being debated in the literature[149]. So, it is likely that some of the apparent global sea level rise is due to researchers either failing to properly account for this rebound, or else applying an inappropriate glacial isostatic adjustment.

Partly to overcome these challenging problems associated with the tidal gauge estimates, a series of satellites have been launched since 1992 with instruments which can be used to monitor global sea levels. The current satellite estimates of the global sea level trends suggest that global sea levels have indeed been rising (since at least 1992)[144]. Indeed, some studies suggest that the rate is faster than estimated by the tidal gauges, e.g., Church & White, 2006[144]. However, Mörner, 2004 noted that an early version of the satellite-based global sea level estimates implied that there was essentially no long-term trend, other than a temporary rise during the late 1990s associated with an unusual El Niño year[150].

Nerem et al., 2007 vehemently criticised Mörner, 2004 on the grounds that the version he had used was based on the raw satellite data, and they argued that several adjustments were needed to correct for alleged biases in the raw data[151]. However, whether these adjustments are valid or not, Mörner, 2008, is correct to highlight the fact that most of the alleged trend in the satellite estimates is due to artificial theoretical/semi-empirical adjustments, rather than pure experimental observation[152]. This means that the claims that the satellites are detecting an unusual global sea level rise need to be treated with considerable caution.

Considerable attention has been given to the various proxy-based studies which use temperature proxies such as tree ring measurements and lake sediments to estimate global (or hemispheric) temperature trends on time-scales of centuries and even millennia, e.g., the Mann et al., 2008[153] or Ljungqvist, 2010[154] studies. We consider these studies in detail in a separate paper[155]. In that article, we note that there are still a number of unresolved uncertainties which require careful consideration, before the proxy-based estimates can be considered reliable. However, for the purposes of the current article, it is sufficient to point out just one critical problem - the so-called "divergence problem" - many of the proxies find 20th century temperatures peaked in the early-to-mid 20th century[156], directly contradicting the current weather station estimates.

The fact that there are problems with these longer estimates does not mean that they are without value. Some of these problems may be reduced, or overcome, with further research. But, until their uncertainties can be adequately reduced (particularly for before the mid-20th century), they need to be treated with considerable caution. For this reason, they are inadequate for assessing the extent of urbanization bias in the weather station-based estimates.

3.4 Assuming urbanization bias causes as much "urban cooling" as "urban warming"

As we discussed in Section 2, urbanization bias is predominantly a warming bias. However, the authors of some of the studies appear to have decided that it also leads to major cooling biases[3, 11, 12, 18, 22, 157]. For instance, Peterson, 2003 makes the claim that *"Some urban stations are indeed warmer than nearby rural stations but almost the same number are colder"*. This belief seems to have led these researchers to be less critical of their own analysis when they failed to detect strong urban warming biases - a result that should have been unexpected if they were more familiar with the urban heat island problem.

It is true that under certain circumstances, urban development can sometimes lead to cooler conditions. For instance, in dry, hot desert areas, urban features can sometimes lead to cooler daytime temperatures[45, 158]. But, as we discuss in Paper II[1], these cases of "urban cooling" tend to be in the minority, and the main tendency of urban development seems to be towards urban warming. This can be seen from Figure 7, or indeed from the literature review in Section 2. In addition, it is apparent from the fact that there is such interest in trying to modify urban planning and development to deliberately counteract urban *heat* islands[158–161], e.g., see http://www.urbanheatislands.com/. If urbanization bias were predominantly a cooling bias, then there should be no interest in trying to mitigate the magnitude of urban heat islands.

3.5 Relying on a single urbanization threshold

Although it seems reasonable to suggest that the magnitude of urban heat islands should generally increase with increasing urbanization, the exact relationship seems to vary depending on the type of urban development, location within the urban area, and the underlying climate of the area[23–26, 45]. In addition, urban development itself takes many different forms, depending on the culture, history, types of urban activity in the region, etc.

This means that it is difficult to establish a single "urbanization metric" that can universally identify how urbanized a particular area is, let alone establish the extent to which that area is affected by urbanization bias.

Several different urbanization metrics have been used for estimating how urbanized particular stations are, but often different metrics provide different estimates. Many early studies of urbanization bias relied on local population size as an approximation of the degree of urbanization, e.g., Karl et al., 1988[32]. However, with advances in satellite technology, a number of gridded datasets have been developed for estimating urbanization using night-light brightness[162], vegetation and land cover[75], Impervious Surface Area (ISA)[45], or combinations of several metrics, such as the "Moderate Resolution Imaging Spectroradiometer" (MODIS) datasets[163].

There is often a considerable degree of overlap between different urbanization metrics, e.g., Imhoff et al., 2010 found similar urban boundaries when using the 25% ISA threshold and the MODIS "Urban Built-Up land" map[45]. However, so far, all of the current metrics have their limitations. For instance, Imhoff et al., 1997 calibrated night-light brightnesses to U.S. energy use and developed a dataset which was very successful at identifying urban boundaries for the U.S.[162]. But, as we discuss in Paper II[1], the U.S. has a particularly high per capita electricity usage, and so this U.S. calibrated dataset seriously underestimates the degree of urbanization of other countries.

Unfortunately, most of the ten sets of studies only used one metric for distinguishing urban and rural stations, and just used a single threshold value of that metric for describing all stations: Hansen & Lebedeff, 1987 defined a station as being "urban" if it had an associated 1970s population greater than 100,000. Otherwise, it was considered "rural"[13]. Easterling et al., 1997 used a similar approach, but used a population threshold of 50,000[16]. Peterson, 2003[18] and Wickham et al., 2013[22] used different metrics (night-light brightness and MODIS, respectively), but again just used a single threshold value to distinguish between urban and rural stations. Parker 2004 & 2006[19, 20] considered a few threshold values for his windy-calm sub-setting experiments, but for each experiment, only allowed one value.

The Hansen et al. studies considered two different metrics, but only used one at a time - Hansen et al., 1999 used a population-based metric[11]; Hansen et al., 2010 used a night-light brightness metric[3]; while Hansen et al., 2001 used a night-light brightness metric for their U.S. stations and a population-based metric for the rest of their stations[12]. Peterson et al., 1999 was the only one of the studies which used more than one metric simultaneously for evaluating their stations, but again they only allowed two possibilities for a station - if a station had a low night-light brightness *and* a low associated population, then it was considered "rural", otherwise it was not[17]. Hansen et al., 1999 did allow three values for their stations - if a station had an associated population less than 10,000 it was considered "rural"; if it was greater than 50,000 it was considered "urban"; but if it had an intermediate population it was considered a "small town" station[11]. However, in their subsequent studies, Hansen et al. combined their "small town" and "urban" categories, i.e., they only considered two categories[3, 12].

There are several problems with just using a single metric and threshold value for assessing the urbanization of stations.

Firstly, it does not allow for the fact that urbanization development is generally a progressive and continuous process. The magnitude of the urbanization bias at a station should in general increase over time, as the area becomes increasingly urbanized. But, this process would vary over time, and so the extent of bias in a given record depends not just on the current urban heat island, but also on how that urban heat island expanded over the course of the record[34, 61], as we discussed in Section 2.2.

Some studies which use a population-based metric have used population *growth* as a metric, rather than the populations at a fixed time, e.g., Karl et al., 1988[32] or Hausfather et al., 2013[164]. However, this requires a dataset which can compare populations over time, and so is only possible for regions which have carried out regular censuses for a long enough period, e.g., the U.S.

Since most of the satellite-based metrics are only based on an analysis over a relatively short period, most researchers using these metrics have instead taken the approach of assuming that, if a station is currently rural it always was and is unaffected by urbanization bias, but if it is currently urban, and its record is sufficiently long, then it is probably affected to some extent by urbanization bias[22]. This is an understandable approximation, but it should be recognised that it is a very crude one. As we discussed in Section 2.2, the history of urban development at a station is more relevant for studying urbanization bias than the size of the current urban heat island. Also, it is important to remember that a station which is currently rural may well have been moved from an earlier urban location.

A second problem with using a single threshold is that values which are too strict will reduce the number of "rural" stations so much that the analysis will lose significance, while using a threshold which is too lax will result in many stations which are suffering from urbanization bias being treated as "rural". A third problem is that the threshold may be inaccurate under certain circumstances. We mentioned above that some metrics, such as night-brightness, have different relationships to actual urbanization, depending on the characteristics of individual nations. We also mentioned that all satellite-based metrics require that the station co-ordinates assigned to each station are accurate, but that this is not always the case. Static population-based metrics (as opposed to population growth-based) often only provide crude estimates of urbanization bias, because the strength of an urban heat island depends on the location in the urban area, e.g., a station near the centre of a small town might be affected by a larger urban heat island than a station on the outskirts of a large town.

We appreciate the motivation for using a single metric and threshold for studying urbanization bias - it is much easier, and requires a lot less effort. However, it seriously limits the detection ability of such studies. We recommend that future studies use more flexible approaches. Stewart & Oke, 2012[39]'s recommendation that researchers use a variety of thermal climate zones for describing the urbanization of stations would certainly allow a more nuanced identification of the different degrees of bias at individual stations, although it would probably require a detailed inspection of each station being studied[7], and so might require too much work for a global temperature analysis using several thousand stations. A simpler, but still useful, approach might be to use a scale which allows different levels of urbanization from very rural to heavily urbanized. Imhoff et al., 2010 took this approach and defined areas as having one of five different levels of urbanization[45].

In much of our analysis in this series of papers, we used the same metrics used by Peterson et al., 1999, i.e., the night-brightness and population values associated with each station in the Global Historical Climatology Network dataset. However, we combine these metrics to provide three possible values for each station - a station is "rural" if it is identified as rural by both metrics, "urban" if it is identified as urban by both metrics, and otherwise "intermediate". We believe this offers a more nuanced approach than that taken by the ten sets of studies described here. But, we suspect further refinement of the urbanization identification process would yield more reliable estimates of the extent of urbanization bias in current global temperature estimates.

4 Reassessment of the studies claiming urbanization bias is small or negligible

4.1 Hansen & Lebedeff, 1987

As well as constructing their standard global temperature estimate, Hansen & Lebedeff, 1987 [13] also constructed another estimate based on a "rural" subset of stations. This subset consisted of all their stations which were not associated with a city with a 1970s population greater than 100,000.

Their estimated temperature difference between 1880-1885 and 1980-1985 was reduced from $0.7°C$ in the full set to $0.6°C$ in the rural subset[13]. They concluded that the difference of $0.1°C$ was due to urbanization bias, and guessed that the remaining bias (stations with 1970s populations less than 100,000) would not be any greater. Hence, they estimated an upper bound of $0.2°C$/century of their estimated global warming, i.e., less than 30%, was due to urban bias.

Their estimate of the urbanization bias is actually quite substantial relative to the conclusions of the other studies being reassessed in this article - a bias

[7] Ideally, this would be done by on-site inspection. But, reasonable estimates may be possible using satellite imagery software, such as Google Earth, provided the station co-ordinates are accurate.

of $0.2°C$/century in a trend of $0.7°C$/century corresponds to a bias of $\sim 28.6\%$. Moreover, presumably it would be greater now following the increase in worldwide urbanization since the early 1980s (see Figure 2). Nonetheless, Hansen & Lebedeff, 1987 used it to conclude that the urbanization bias in their estimates was small[13]. Indeed, Hansen used an update of this estimate in his high-profile testimony to a 1988 U.S. Senate Committee to conclude that his calculated global warming was not natural, but anthropogenic, due to increasing atmospheric carbon dioxide (CO_2) concentrations[165].

Since the stations and station data used by Hansen & Lebedeff, 1987 are not publicly available, it is difficult to properly assess how robust Hansen & Lebedeff's analysis was. However, Karl & Jones, 1989[73] were provided with such information for their analysis of the U.S., and their study offers some insight. As we will discuss in Sections 4.2 and 4.3, Karl & Jones, 1989 were attempting to estimate the magnitude of urbanization bias in then-current global temperature estimates. They compared Karl et al., 1988's mostly-rural high density U.S. station network (which had been adjusted to account for urbanization bias) to the U.S. trends of both the Jones estimate [166, 167] and the Hansen & Lebedeff estimate.

Karl & Jones' analysis suggested that Hansen & Lebedeff's estimates left a substantial $0.3 - 0.4°C$ of urban bias for the period 1901-84[8] for the U.S.[73]. Hansen & Lebedeff were unable to detect this bias in their sub-setting experiments[13]. So, this suggests the sub-setting experiments carried out by Hansen & Lebedeff, 1987 were inadequate.

Balling & Idso also developed population-adjusted temperature trend estimates for the U.S.[34]. When they calculated the 1920-1984 temperature trends for the eastern portion of the U.S. (box 16 of Hansen & Lebedeff's analysis), they got a 64-year cooling trend of $-0.39°C$, however Hansen & Lebedeff only produced a cooling trend of $-0.02°C$ for that period. This suggested a warming bias of $+0.37°C$/64 years ($+0.58°C$/century)[34].

Christy & Goodridge, 1995 [77] compared their own archive of 112 relatively-long (covering at least 1910-1989) stations for the western U.S. state of California to Hansen & Lebedeff's California records. They found that Hansen & Lebedeff only used 6 stations of comparable length (5 of these stations were also in Christy & Goodridge's archive), and the trends for each of those stations all showed more warming than at least half of the stations in Christy & Goodridge's archive. Three of Hansen & Lebedeff's long California station records (i.e., half) showed more warming than the 90th percentile of Christy & Goodridge's archive. In other words, the stations selected by Hansen & Lebedeff showed more warming than average (for California at least). This suggests that, even if the most urbanized stations were successfully removed in Hansen & Lebedeff, 1987's sub-setting experiments, many of the remaining stations could have still been biased warm.

[8] In contrast, they estimated the bias in the Jones estimate to only be $\sim 0.1°C$ over the period 1901-84 - see Section 4.3.

4.2 Wigley & Jones, 1988

In the 1980s, Jones et al. published an early version of the Climate Research Unit's current global temperature estimates, which was based on what was known as the "Jones dataset"[166–170]. By making site-by-site comparisons between pairs of stations, they believed[166–170] that they had successfully identified and corrected for spurious, non-climatic temperature changes (due to station moves, changes in instrumentation, etc.) in the Jones dataset.

They also believed they had detected those artificial trends which were due to urban warming, and to have explicitly removed them from their dataset. They identified 41 stations which were already showing strong urban warming by 1984: 38 in the Northern Hemisphere and 3 in the Southern Hemisphere. Presumably these effects would be stronger and more widespread now.

Wood, 1988 expressed concern that this approach had been insufficient[33]. This elicited a rather heated response from Wigley & Jones of the Climate Research Unit [14], who claimed that *"The arguments presented by Wood and his criticisms of the methods used by Jones et al. are largely fallacious and are generally based on misconceptions and unwarranted assumptions"*. But, it appears that they seriously misread Wood, 1988 when making their attempted rebuttal. As will be discussed below, many of Wood's concerns were valid (and still are).

With this in mind, it is worth systematically comparing the alleged "errors" Wigley & Jones' claimed to have identified in Wood's article, with the claims Wood had actually made. Wigley & Jones claimed to have found 9 errors in Wood's analysis:

Alleged "errors" in Wood's analysis

1. **Wigley & Jones:** *"Wood states that, in the*

Jones et al. (1985) work, the Urban warming was clearly identified at **only** *38 stations out of 2666. This is a serious distortion of the facts. Station records were examined for non-climatic trends (of any type, not necessarily just urban warming trends) only after examination for other types of error. Many records which may have had urban warming trends were removed for other reasons during these earlier error detection stages"* [Emphasis added in bold.]

However, Wood had actually made a different statement:

Wood: *"Urban warming was clearly identified at 38 stations out of 2666 for the Northern Hemisphere... Additional stations that experienced urban warming may have been removed from the data set for other reasons (such as uncorrectable changes in instrumentation or observation times), although this cannot be determined from the published station details."*

2. **Wigley & Jones:** *"Wood suggests the possibility that* **all** *neighbouring stations used for comparison may have had similar urban warming trends. This is unlikely."* [Emphasis added in bold.]

At no point, did Wood suggest that *all* neighbouring stations were biased:

Wood: *"One possibility is that for a significant number of locations, neighbouring stations also may have experienced urban warming, so that station comparisons obscured some or all of the warming bias. This could be particularly the case when urban stations are compared with other urban stations or with suburban or smaller city stations that may, nonetheless, experience significant warming bias themselves."*

3. **Wigley & Jones:** *"Wood implies that urban warming is correlated with population growth. While this is true qualitatively, the correlation between the rate of warming and the rate of population growth is not strong. Karl et al. (1988) have shown that, for cities with populations below 100 000, population accounts for only up to 4% of the daily mean temperature difference between urban and neighbouring rural sites."*

This is a puzzling statement, as Karl et al., 1988[32] had reached the opposite conclusion: *"Urbanization has influenced the climate records of even small towns in the United States."*

4. **Wigley & Jones:** *"Wood implies that all sites with a location label identifiable with an urban centre are, in fact, representative of a centrally-located site. This is wrong. Many such sites are, in fact, in locations peripheral to the urban centres from which they take their names and are located in regions which may have undergone only minor changes in their environments."*

But, Wood had **not** implied that. Instead, he had noted the valid point that:

Wood: *"Many land stations are located in* **or near** *areas that have become increasingly urbanized during the twentieth century, and thus the temperature data, if uncorrected, will reflect a gradual warming associated with urbanization."* [Emphasis added in bold.]

5. **Wigley & Jones:** *"Wood notes (following Kukla et al., 1986) that even peripheral sites may show urban warming effects. This is correct, but irrelevant. In our analyses, such sites were most probably eliminated in the early stages since none remained in the final testing for spurious trends. For example, the Puerto Rico site mentioned by Wood and examined by Duchon (1986) did not survive our various tests. In contrast to Duchon, however, we interpreted the apparent trend as a step change of 0.8°C around 1970 and corrected the record accordingly."*

Wigley & Jones appear to have been confused about why Wood mentioned San Juan (*"the Puerto Rico site"*). Wood was discussing the problem (still unresolved[53, 80, 171]) of whether airport weather stations should be regarded as "rural" or "urban". This is an important issue since many of the current weather stations (both rural and urban) are located at airports. With that in mind, referred to Duchon's study of San Juan airport[172]:

Wood: *"Another potential problem mentioned by Kukla et al. (1986) is that stations located (or relocated) at airports, once thought to be relatively free of urban warming, may have experienced increased urban warming in recent decades due to growth in and around airports. A recent analysis of San Juan, Puerto Rico international airport station suggests a substantial urban warming of about 0.8°C per decade since 1956, presumably resulting from runway and terminal facilities expansion plus adjacent residential and commercial development (Duchon, 1986). This suggests the need to carefully examine airport stations for previously undetected urban warming."*

6. **Wigley & Jones:** *"Wood suggests that the detection threshold used by Jones et al., in the last analysis stage may have been too high. He notes that the 1881-1980 trend difference in the 38 relevant pairs was 0.89°C. This number is correct, but it is irrelevant and bears no direct relationship to the threshold used by Jones et al."*

The value 0.89°C was indeed irrelevant to Wood's discussion of the Jones et al. threshold, but it was not mentioned there:

Wood: *"While the actual detection threshold used by Jones et al. (1986a, 1985) cannot be determined from the published station details, the 38 stations identified as exhibiting urban warming collectively increased by 1.2°C over the 1881-1980 period, more than double the mean for all stations, and more than three times the 0.33°C warming measured at 38 neighbouring stations thought to be free of warming bias. Thus stations exhibiting a lesser but still significant degree of urban warming relative to neighbouring stations may not have been identified and removed."*

Instead, he mentioned it (or rather the slightly different value of 0.87°C) in the next point, where it **was** relevant:

Wood: *Actually, the 0.87°C average warming for the period 1881-1980 (1.20°C less 0.33°C) due to urban heat island effects for the 38 stations translates to about 0.009°C warming per year or 0.09°C per decade. This is towards the lower end of urban warming rates identified in a variety of independent research studies."*

7. **Wigley & Jones:** *"Wood states that, of the 34 city sites used in the Kukla et al., (1986) analysis, 23 appear in the Jones et al. (1985) work, but only three of these were eliminated 'on the basis of urban warming'. This is a misleading statement. In fact, of Kukla et al.'s 34 city sites, only ten were used without correction. In other words, our analyses showed that the annual mean data for these ten sites showed no appreciable urban warming effect. This result is not incompatible with Kukla et al.'s work. Of the remaining thirteen, eight were corrected and five were eliminated from the data set. Of these five, three were eliminated solely because of spurious warming trends. Spurious warming trends were not the sole basis for eliminating stations, only the last criterion used."*

However, Wood was not disputing the notion that Jones et al. might have removed additional stations which had urban warming in their earlier stages. He was merely pointing out that of 23 stations identified by another research group as "urban" (Kukla et al.[31]), Jones et al. had only explicitly identified three of them as "urban", and that there was not enough published information to explain why:

Wood: *"Of the 34 urban station locations studied by Kukla et al., 11 were not in the Jones et al. data set and of the 23 that were, only three (Denver, CO; Oklahoma City, OK; Tucson, AZ) were identified and removed by Jones et al., on the basis of urban warming. A definitive comparison requires analysis of the detailed station data (not available in the published works), and presumes that urban stations listed by Kukla et al. and Jones et al. under the same name are in fact the same stations."*

8. **Wigley & Jones:** *"Wood notes that some stations lacked neighbouring sites for intercomparisons. This is certainly true; it is noted in the original paper by Jones et al. (1985). However, just because Jones et al. (1985) do not list a comparison station does not mean that no intercomparisons were made. Indeed for places like the Peoples Republic of China, extensive intercomparisons were made. This work was still in progress when Jones et al. (1985) was written."*

Wood was merely evaluating the Jones et al. analysis on the basis of published data. It is hardly an "error" on Woods behalf that he could not find inter-comparisons for the Peoples Republic of China (and other stations), if Jones et al. had not made them public. Indeed, the still unpublished Chinese inter-comparisons would be of importance today, because as will be discussed in Section 4.3, genuinely rural Chinese stations are very rare and most "rural - urban" inter-comparison studies in China are actually "strongly urban - moderately urban" inter-comparisons.

9. **Wigley & Jones:** *"Wood notes that some of our U.S. station comparisons involved pairs that were some distance from each other. This is only true for the earlier parts of the record. Even then, comparisons were always made between stations with well-correlated inter-annual variations, so the distance criticism is a red herring."*

Wigley & Jones may have personally felt this was "*a red herring*", but we still agree with Wood's suggestion that *"(w)hether such stations are really suitable for the detection of urban warming"* at the very least *"warrants attention"*.

Was Wood wrong?

It is unfortunate that Wigley & Jones were unable to address Wood's actual concerns. However, it must be acknowledged that, subsequently, Jones et al., 1989[94] did discuss in more detail some of the issues raised in point #7, i.e., the differences between Kukla et al., 1986[31] and the Jones et al., 1986[166] U.S. component.

Jones et al., 1989 also revealed a previously unpublished detail, i.e., how stations were identified as having *"urban warming"* in the Jones dataset. They had looked for: *"stations that exhibited the 'classic'* *urban warming effect: a steady, quasi-linear rise in temperature in comparison to neighbouring stations."* Stations which did not show this particular behaviour were assumed to be unaffected by urbanization bias.

This perhaps explains why their approach failed to detect many urban heat islands outside of North America; only 10 of the 41 stations identifed by Jones et al. as having urbanization bias were outside North America. 1046 out of their 2666 Northern Hemisphere stations (39%) were located in North America[166]. This meant that there was a very high density of stations there, and hence a lot of stations would have several nearby stations that could be used for comparison.

In contrast, the rest of the world had a much lower density, meaning that stations often would only be compared to distant "neighbours". If we consider the record for Moscow, Russia (Figure 8), we find several large step changes, which probably are non-climatic in between gaps in the early part of the record. But, after 1880, we find a steady, quasi-linear rise in temperature. By itself, this would appear to be a perfect example of Jones et al., 1989's definition of an urban warming effect, and indeed Lokoshchenko & Isaev, 2003 noted a substantial urban heat island for Moscow city[173]. However, because Jones et al. did not have any nearby stations with a long enough record, they had to use two stations nearly 1000km away for their comparisons - Riga (c. 850km away) and Arkangelsk (c. 980km away)[170].

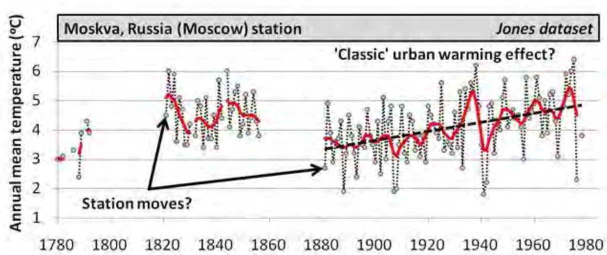


Figure 8: *Annual 1781-1984 mean temperatures for the Moscow, Russia station. Thick red line corresponds to 11-point binomial smoothed average. Data taken from the Climate Research Unit's 1991 dataset: http://cdiac.ornl.gov/ftp/ndp020/.*

Compared to Riga (which showed no significant trend), the Moscow station *would* have shown a *" 'classic' urban warming effect"*. But, compared to Arkhangelsk (which showed warming in the early 20th century, then cooling until about 1970), the difference would **not** be a simple, *"steady, quasi-linear rise in temperature"* - see Figure 9. In other words, the Moscow station did not meet Jones et al.'s criteria for identifying stations biased by urban warming. The station was therefore included in the Jones dataset, *without any urban warming correction*, and its steady, quasi-linear rise in temperature was implicitly assumed to be "climatic".

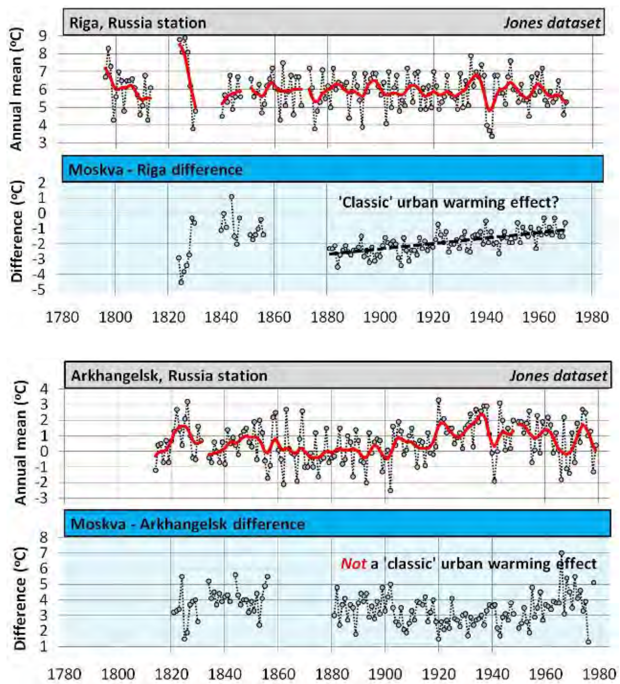


Figure 9: *Annual temperatures for the two stations Jones et al. used for assessing the Moskva (Moscow, Russia) station and the corresponding difference series. Thick red lines correspond to 11-point binomial smoothed averages. Data adapted from the Climate Research Unit's 1991 dataset: http://cdiac.ornl.gov/ftp/ndp020/*

Their method appeared to work relatively well for the U.S. since, according to Karl & Jones, 1989, the urban bias remaining in the Climate Research Unit estimate for the US was only of the order of 0.1°C over the period 1901-84[73]. In comparison, the Goddard Institute for Space Studies' then approach[13] apparently left 0.3-0.4°C of urban bias for the U.S (see Section 4.1). However, since Jones et al. were only able to identify 10 stations as being affected by "urban warming" outside of North America, it is likely that their detection method was too weak for the rest of the world. In Section 4.3, a later assessment of the urbanization bias in the Jones dataset is considered.

It is worth noting that, for their current global temperature estimates, the Climate Research Unit have *not* carried out any further explicit urbanization adjustments other than the ones Wood, 1988 was evaluating[33]. Instead, the Climate Research Unit now *assume* that the data providers they receive the rest of their station data from have already removed any non-climatic biases from the data[174]. It should be clear from the discussion throughout this paper that this is an unwise assumption. In addition, urbanization has actually accelerated since the 1980s (Figure 2), so even if their adjustments in the early 1980s[166–170] were as comprehensive as Wigley & Jones, 1988 claimed[14], they would be seriously outdated by now.

4.3 Jones et al., 1990

Jones et al., 1989[94] compared the U.S. subset of the Jones dataset (Section 4.2) to Karl et al., 1988[32]'s mostly rural U.S. dataset (mentioned in Section 2.4), which had been adjusted to remove urbanization bias (using city populations as a metric for urbanization). They found that the Jones version showed $0.15°C$ more warming over the period 1901-84 than Karl et al.'s most reliable subset, although only $0.08°C$ more warming than a less reliable subset of Karl et al.'s.

Jones et al., 1989 therefore decided that the urbanization bias in the U.S. subset of the Jones dataset was about $0.1°C$ over the period 1901-84. They guessed that the bias for the rest of the world might not be any bigger than this[94], but recommended that more work be carried out.

Karl & Jones, 1989[73] were more cautious. They calculated a similar estimate for the U.S. subset of the Jones dataset. However, for the U.S., this accounted for much of the "warming trend". Indeed, the U.S. appeared to show a "cooling trend" since the 1930s.

They also noted that urban heat islands were substantial in many parts of the world, outside the U.S. This suggested the possibility that much of the $0.4°C$ warming trend for 1901-84 for the rest of the world in the Jones dataset could also be biased by urban warming[73].

Jones et al., 1990[15] therefore attempted to carry out estimates of the urban bias for three regions outside the U.S.: western U.S.S.R., eastern Australia and eastern China. They attempted to select mostly rural networks for each of those regions and compared them to both the Jones dataset and Vinnikov et al., 1990[175, 176]'s dataset (a predecessor of the Lugina et al., 2006 dataset[6]).

All networks showed a cooling trend for western U.S.S.R. over the 1930-87 period. However, the Jones subset showed less cooling ($\sim$ -$0.1°C$) than the rural and Vinnikov subsets ($\sim$ -$0.2°C$).

For eastern China, they constructed two networks - one highly urbanized and one moderately urbanized (which they considered "rural"). The networks only covered the period 1954-83. For the rural network, peak temperatures occurred in the 1960s, but yielded a warming linear trend of $0.23°C$. This was considerably less than the highly urbanized network's linear trend of $0.39°C$, however was greater than the linear trends for the other two datasets, Jones ($0.19°C$) and Vinnikov et al. ($0.13°C$).

The Australian subsets showed the most warming of any of the subsets (including the U.S. subsets of Jones et al., 1989[73, 94]). However, while the Jones subset showed more warming ($0.60°C$ over the period 1930-88) than the rural subset ($0.56°C$ over the period 1930-88), it was not by much. The Vinnikov et al. subset was comparable to the rural subset, although calculated over a slightly different period ($0.55°C$ over the period 1930-87).

Because the differences in the linear trends of the "rural" subsets and the Jones and Vinnikov et al. equivalents were small for the periods considered, Jones et al., 1990 concluded that the Jones and Vinnikov et al. datasets were not overly affected by urban warming for those regions[15]. They then extrapolated that conclusion to assume that their hemispheric estimates were not overly affected by urban warming either.

There are at least three major flaws in the Jones et al., 1990 analysis.

First, was their extensive reliance on linear trends of their various subsets in their comparison. We discussed in Section 3.1 how using linear trends to describing data with strong non-linear trends can provide very misleading results. From Figure 10, we can see that all of these subsets were dominated by *non-linear* trends. This is also evident from the fact that the only linear trends which were statistically significant were the Australian subsets and their urban Chinese subset. Hence, it is difficult to see that "linear" trend values have any relevance for such analysis.

Second, the particular subsets and periods Jones et al., 1990 considered (with the possible exception of Australia) did *not* show the strong warming trends of Jones et al. and Vinnikov et al.'s hemispheric estimates that were under contention. This can be seen by comparing the subset trends of Figure 10 to the

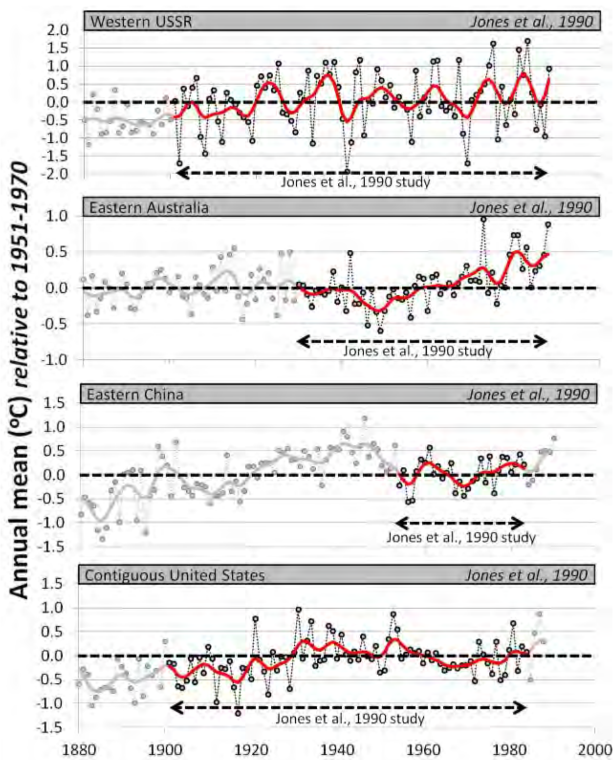


Figure 10: *Temperature trends from "the Jones dataset" for the four regions considered by Jones et al., 1990[15]. The periods considered by Jones et al., 1990 for each of the regions are highlighted. Thick lines correspond to 11-point binomial smoothed averages. Data adapted from the Climate Research Unit's 1991 dataset:* `http://cdiac.ornl.gov/ftp/ndp020/`.

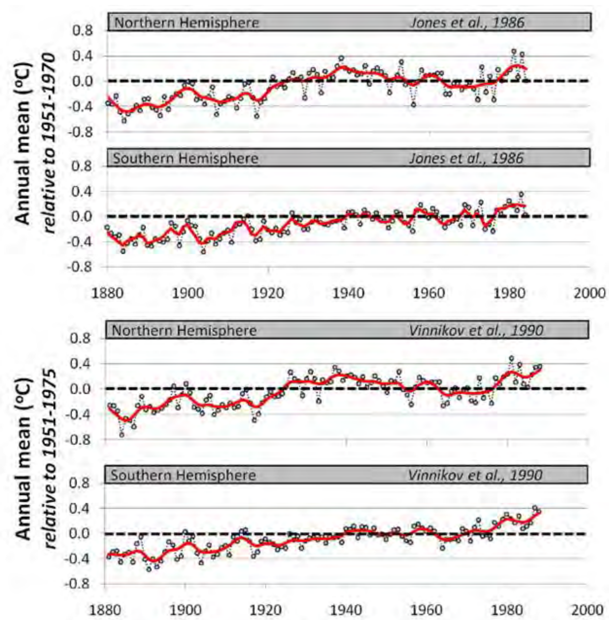


Figure 11: *Northern and southern hemispheric temperature trend estimates of Jones et al., 1986[166, 168] and Vinnikov et al., 1990[176]. Thick lines correspond to 11-point binomial smoothed averages. Jones et al., estimates downloaded from* `http://cdiac.ornl.gov/ftp/ndp003/`. *Vinnikov et al. estimates transcribed from Vinnikov et al., 1990[176].*

hemispheric trends in Figure 11.

Failing to detect unusual "urban" warming in subsets which do not themselves show much warming tells us nothing about how much the global and hemispheric estimates (which do show strong warming trends) are biased by urban warming. Hence, Jones et al., 1990's conclusion that the U.S. urban bias of $\sim 0.15°C$ over the period 1901-84 represents an upper bound for their hemispheric estimates is unwarranted. Their belief that the overall bias is even less, i.e., $< 0.05°C$, appears to be based on wishful thinking, rather than any scientific basis.

Third, later research has suggested some problems (or at least uncertainties) with the subsets used by Jones et al., 1990:

Wang et al., 1990 [84] were unclear how Jones et al., 1990 had concluded the eastern China data were unaffected by urbanization bias, as they had found the exact opposite. They noted that genuinely rural stations with useful, long records were very rare in China. As a result, Jones et al.'s "rural" Chinese stations were not truly rural, and so were also likely to be affected by urbanization bias. Even still, Wang et al. found substantial urban biases between the least urbanized stations and the most urbanized stations.

It is possible that the confusion between Wang et al., 1990 and Jones et al., 1990, who both appear to have used similar (possibly identical[9]) datasets, is due to Jones et al., 1990's reliance on linear trends for their comparison. The Jones and Vinnikov et al. Chinese subsets showed smaller linear trends over the period 1954-1983 than the rural and urban subsets from both studies[15, 84]. Jones et al., 1990 appear to have interpreted this as meaning the Jones & Vinnikov et al. datasets were not affected by urban bias. However, as we mentioned above, linear trends are somewhat arbitrary for non-linear datasets such as these. Indeed, only one of the linear trend values cal-

[9]See the Climate Audit blog for some discussion of the datasets.

culated by Jones et al., 1990 for the Chinese network was statistically significant (the "urban" subset), confirming that such comparisons are irrelevant.

Li et al.[177, 178] recently reached a similar conclusion to Jones et al., 1990 for China. However, most studies now acknowledge that a substantial portion of the recent warming in China is urban-related, including a recent study with the same lead author as Jones et al., 1990, i.e., Jones et al., 2008 [56]. See the introduction to Yang et al., 2011 [88] for a recent review.

Rural-urban comparisons are also difficult in Australia, since most of the urban areas are coastal, and most of the long station records are from urban stations[69]. Coughlan et al., 1990 found that stations in Australia's largest urban areas showed very strong urban warming[69], so it was a point of concern.

Hence, Hughes & Balling, 1994[179] decided to construct an alternative rural composite station network, although their study failed to pass peer review. Their "Hughes dataset" only showed about half the warming of Jones et al., 1990's "rural" subset, suggesting that *all* of the Australian subsets considered by Jones et al., 1990 were substantially affected by urban bias.

Kamél, 2004[180] considered the Climate Research Unit's temperature trends for Russia (former U.S.S.R.), although his study also failed to pass peer review. He found that the Climate Research Unit substantially overestimated the warming there, possibly indicating urban bias remained[180]. Kamél had considered a region to the east of the Jones et al., 1990 subset (southern Siberia), so it is plausible that Jones et al., 1990 had coincidentally chosen a region of Russia which was relatively biased. But, in either case, their conclusion that their *global* estimates were relatively unbiased[15] was invalid.

Finally, we note that the conclusions reached by Karl et al., 1988[32], Karl & Jones, 1989[73] and Wang et al., 1990[84] all differ strongly from those of Jones et al., 1989[94] and Jones et al., 1990[15]. This is surprising, because all of these studies were published at around the same time, and there is a significant overlap between the authors of the studies. This suggests that the authors of the Jones et al., 1990 study were not unanimous in their conclusions.

4.4 Easterling et al., 1997

Easterling et al., 1997[16] was not predominantly concerned with the urbanization bias problem. Instead, Easterling et al., were assessing long term global changes in a temperature variable known as the "diurnal temperature range". However, one of their findings has been used to suggest that urbanization bias is small, and so it is worth reviewing.

Before improvements in automation, the total number of temperature measurements that could be made at a weather station in any day was very limited. As a result, observers would typically just use a minimum/maximum thermometer, then check and reset it once (or possibly a few times) a day. Hence, the "daily temperature averages" in station records often were simply the mean of the maximum (T_{max}) and minimum (T_{min}) temperature reached in that 24 hour period, i.e.,

$$T_{avg} = \frac{T_{max} + T_{min}}{2} \tag{1}$$

This single variable does not provide any information about the temperature variability throughout the day. For this reason, some researchers also study the diurnal temperature range (DTR), defined as,

$$DTR = T_{max} - T_{min} \tag{2}$$

This variable together with T_{avg} provides a rough description of the entire daily temperature description, constructed from just two measurements, i.e., T_{min} and T_{max}. In a sense, they offer a crude analogue for the mean and standard deviation of a large number of daily temperature measurements.

Unfortunately, many of the data sources for monthly or even daily temperatures just report T_{avg}. However, for a few thousand stations, the National Climatic Data Center were able to collect monthly averages of T_{max} and T_{min} as well as T_{avg} for their Global Historical Climatology Network datasets.

Easterling et al., 1997[16] decided to use these monthly averages to construct estimates of the changes in DTR, i.e.,

$$DTR_{mon} = [T_{max}]_{mon} - [T_{min}]_{mon} \tag{3}$$

where the subscript $_{mon}$ corresponds to the monthly average. It should be recognised that this is **not** strictly the same as the average monthly DTR, $[DTR]_{mon}$, i.e., the monthly average of the daily DTR values. But, for the purposes of discussion, it will be assumed that they are comparable.

In a similar earlier analysis, Karl et al., 1993[181] had noted a general global decrease in DTR. While they considered a number of different possible explanations for this, including urbanization, they suggested that anthropogenic global warming from increasing atmospheric carbon dioxide (CO_2) concentrations might be a major factor.

Their argument was that anthropogenic global warming would cause T_{min} to increase more than T_{max}. This would lead to a decrease in DTR. However, there are actually quite a few different mechanisms which could lead to changes in DTR, including,

1. Changes in local climate, e.g.,
   (a) Urbanization bias[30, 76, 181–185]
   (b) Land use[76, 79, 181, 182]
   (c) Cloud cover[181, 186]
2. Changes in measurement procedure, e.g.,
   (a) Instruments[181, 187, 188]
   (b) Microclimate and site location[189]
   (c) Time of observation[10]
3. Changes in global climate, e.g.,
   (a) Anthropogenic global warming[16, 181, 190]
   (b) Natural global warming or cooling[181]

Urbanization bias often affects T_{min} more than T_{max}[182], leading to a decrease in DTR similar to Karl et al.'s proposed anthropogenic global warming mechanism. Gallo et al., 1996 argued that urbanization and other land-use changes had significantly affected the DTR of many weather station records[182].

For this reason, as part of Easterling et al., 1997's study they carried out a rural subsetting experiment[16]. Both their complete set and their rural subsets showed similar decreases in DTR, leading them to conclude that the decrease in DTR was not due to urbanization bias. This was then extrapolated to suggest that global T_{avg} estimates were *also* unaffected by urbanization bias[16, 190]. As we will discuss below, we do not agree that the second claim automatically follows from the first.

In a follow-up study, Vose et al., 2005[190] noted the apparent decrease in DTR had slowed down since Easterling et al., 1997. Rohde et al., 2013a have also confirmed this[8]. Other studies of DTR have also suggested that Easterling et al., 1997's "global" trends were not always apparent in individual regions. For example, Europe has apparently shown an increase in DTR in recent decades [191], as has Mexico [192].

Hence, it is worth re-assessing these trends using the National Climatic Data Center's latest version of the Global Historical Climatology Network monthly temperatures, i.e., version 3. We adopt a similar approach to Vose et al., 2005[190], i.e., using the "Common Anomaly Method" developed by Jones et al.[166–168, 170], with some minor modifications. Vose et al., 2005 only required that each station have at least 8 months of data in a given year to be considered, and a station had to have at least 20 of those years' worth of data to be included. For the analysis presented here, stations were required to have a full 12 months of data for each year to be considered, but stations only needed to have at least 15 years of data for the common anomaly period (1961-1990, as in the Vose et al., 2005 study).

[10] Although we have not found any studies which explicitly consider the effects of changing time of observation on DTR, it is well known that different times of observation can alter T_{min} and T_{max}. Hence, it could also alter DTR.

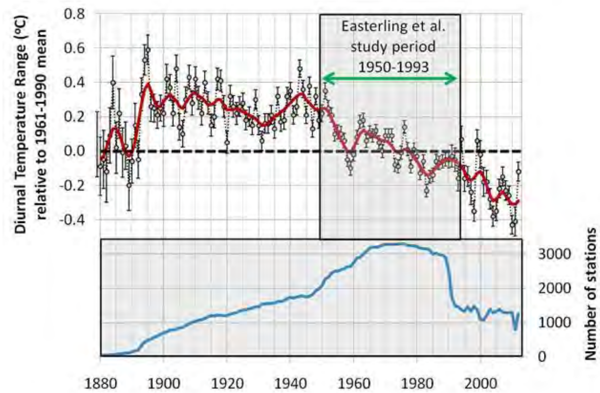


Figure 12: *Changes in the globally averaged gridded diurnal temperature ranges (DTR_{mon}) relative to the 1960-1991 average, calculated from the National Climatic Data Center's monthly Global Historical Climatology Network T_{max}/T_{min} datasets. Error bars correspond to twice the standard error of the gridded means, and the red thick line corresponds to the 11-point binomial smoothed mean.*

Following Vose et al., the anomalies for all the available stations in a given 5° × 5° grid were averaged together for each year. These annual grid averages were weighted by the cosine of the latitude of the middle of the grid, and averaged together to yield a gridded global mean DTR for that year. The trend in this value over the 1880-2012 period is plotted in Figure 12.

The 1950-1993 trend is similar to that reported by Easterling et al., 1997[16], although there appears to have been considerable variability over the longer (albeit more data-sparse) 1880-2012 period. However, from Table 3, it can be seen that, rather than stations being uniformly distributed throughout the globe, most of the stations in the datasets are lo-

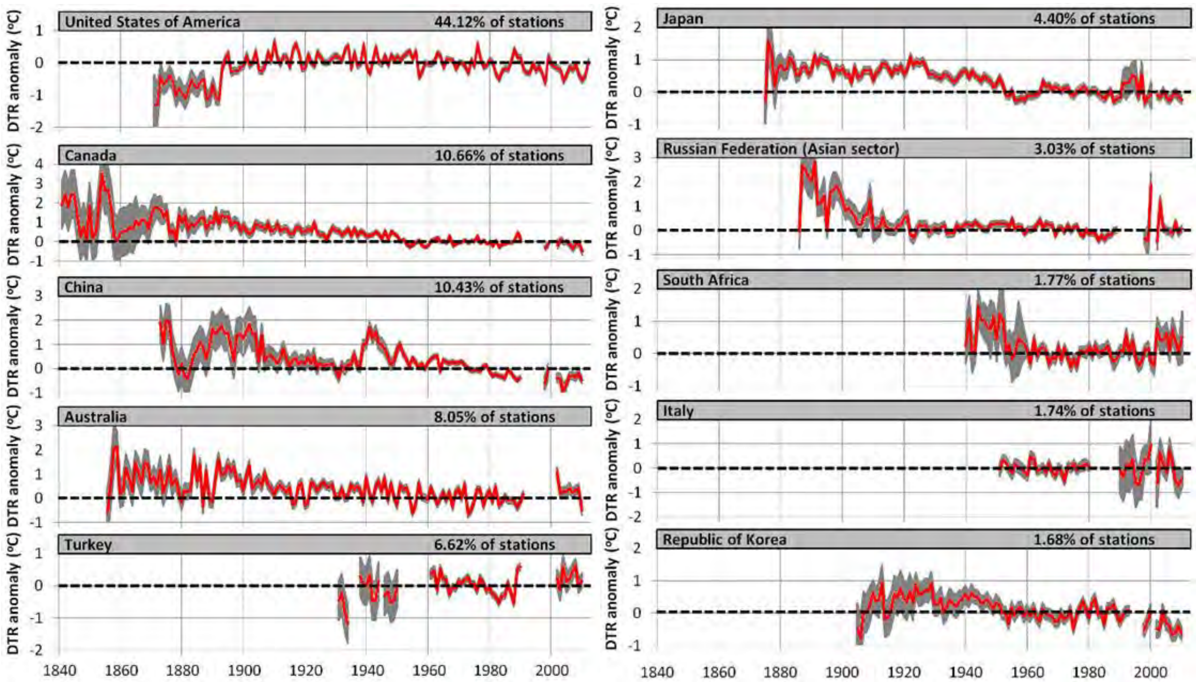


Figure 13: *Changes in the gridded mean diurnal temperature ranges (DTR_{mon}) relative to the 1960-1991 average, for the ten countries with the most number of stations in the National Climatic Data Center's monthly Global Historical Climatology Network T_{max}/T_{min} datasets. Gray bands correspond to twice the standard error of the gridded means. The stations for these 10 countries account for $\sim 92.5\%$ of the stations in the Global Historical Climatology Network with at least 15 complete years of data in the 1961-1990 anomaly period.*

cated in just a few countries. Hence, it is instructive to consider the gridded averages for each of the main countries separately.

Figure 13 shows the gridded mean trends for the top 10 countries in Table 3. Although these 10 countries make up less than 40% of the global land area, they comprise 92.5% of the stations in the Global Historical Climatology Network T_{min} and T_{max} monthly datasets with at least 15 complete years of data in the 1961-1990 anomaly period.

A striking feature of the different country averages is the lack of consistency in trends, both between countries and over time. If the changes in DTR are as uneven and regionally variable as suggested by Figure 13, then it suggests that most of the trends in DTR are a result of regional variability and/or local changes in observation practice, rather than global climate change. It seems that in our above list of proposed mechanisms for DTR changes, mechanisms of Types 1 and/or 2 contribute more than those of Type 3.

Not mentioned in our above list is the possibility that there are errors in the National Climatic Data Center's monthly T_{min} and T_{max} Global Historical Climatology Network datasets. Figure 14 shows the country average trends for Poland (the country ranked 11th in Table 3). The remarkable increase in T_{min} of $\sim 10°C$ and decrease in T_{max} of $\sim 5°C$ for 2002 onwards are too great to be genuine. The fact that they coincide with a change in data source suggests that the explanation is probably some sort of clerical error[11]. We note that the 1961-1990 mean values of T_{max} and T_{min} are very high and low respectively, for a mid-latitude European country such as Poland, which suggests that the data source for the earlier period of the Polish records is unreliable.

So, were Easterling et al. correct in concluding that

[11] The changes occur for all of the Polish stations with data for those years. The National Climatic Data Center provide "adjusted" and "unadjusted" versions of their monthly T_{min} and T_{max} Global Historical Climatology Network datasets. However, the Polish records for both versions are identical, so it is unclear that any "adjustments" were actually carried out.

	Country	Stations	%	Total
1	U.S.A.	1573	44.12%	44.12%
2	Canada	380	10.66%	54.78%
3	China	372	10.43%	65.22%
4	Australia	287	8.05%	73.27%
5	Turkey	236	6.62%	79.89%
6	Japan	157	4.40%	84.29%
7	Russian Fed. (Asia part)	108	3.03%	87.32%
8	South Africa	63	1.77%	89.09%
9	Italy	62	1.74%	90.83%
10	Rep. of Korea	60	1.68%	92.51%
11	Poland	53	1.49%	94.00%
12	Russian Fed. (Europe part)	43	1.21%	95.20%
13	Kazakhstan	22	0.62%	95.82%
14	Sudan	15	0.42%	96.24%
15	Ukraine	15	0.42%	96.66%

Table 3: *Countries with the most stations with at least 15 complete years of monthly T_{max} and T_{min} data during the 1961-1990 period in the National Climatic Data Center's monthly Global Historical Climatology Network datasets. In total, the National Climatic Data Center has 3565 stations from 63 countries which meet this requirement.*

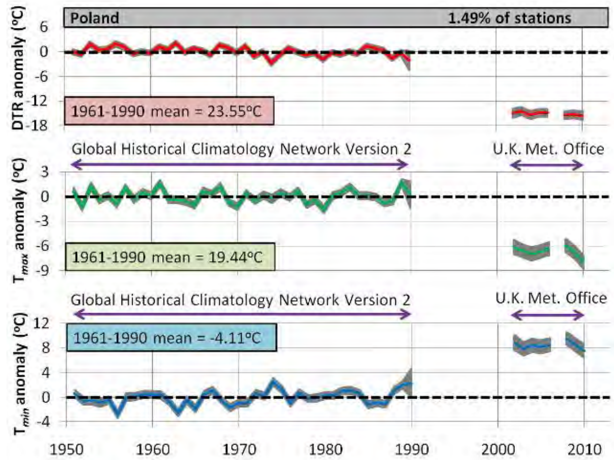


Figure 14: *Changes in the gridded mean T_{max}, T_{min} and DTR anomalies relative to the 1960-1991 average, for Poland in the National Climatic Data Center's monthly Global Historical Climatology Network T_{max}/T_{min} datasets. Gray bands correspond to twice the standard error of the gridded means. Labels at the top of the middle panel indicate the National Climatic Data Center's sources for the data.*

the globally-averaged trend in DTR was not due to urbanization bias? Perhaps. However, neither was it predominantly due to global climate change. Unfortunately, it seems that the quality of the Global Historical Climatology Network monthly T_{min} and T_{max} datasets is currently too irregular and unreliable to draw any meaningful conclusions about urbanization bias from the simple sub-setting experiments of Easterling et al., 1997.

4.5 Peterson et al., 1999

Peterson et al., 1999 [17] carried out a rural sub-setting analysis on the National Climatic Data Center's Global Historical Climatology Network monthly temperature dataset. They used version 2 of their homogeneity adjusted dataset, and estimated the urbanization of stations using two metrics - estimates of the population associated with the area and a satellite-based measure of the night-light intensity of the area during the period 1994-95[17].

Less than a third of their stations met the requirements of having a dark night-light intensity and an associated population less than 10,000. However, when they calculated their gridded global temperature estimate (for 1880-1998) from just those stations, they obtained a similar result to their estimate from the complete set of Global Historical Climatology Network stations. On this basis, they concluded that their complete global temperature estimate was essentially unaffected by urbanization bias[17].

Initially, this might appear a reasonable conclusion. However, a closer inspection of the data they used suggests it is unwarranted. Gray, 1999 (rejected)[193] considered the gridded temperature trends of a dataset similar to that used by Peterson et al. He suggested that the temperature trends of Peterson et al.'s rural subset were overly dominated by anomalously strong warming trends from stations in the former U.S.S.R. He hypothesised that, if these were excluded, Peterson et al. would have detected a substantial difference between the rural and full subsets. Gray did not test his hypothesis (and Gray, 1999 failed to pass peer review). But, it highlights the importance of considering the data from which Peterson et al.'s estimates were constructed - both the full estimate and the rural subset.

While the *unadjusted* Global Historical Climatology Network dataset contains records for 7280 stations (2,290 of them meeting Peterson et al.'s "rural" requirements), the number of stations in the homogeneity adjusted dataset was significantly reduced - 4771 stations, 1401 of which were rural in terms of both population and night-lights[12].

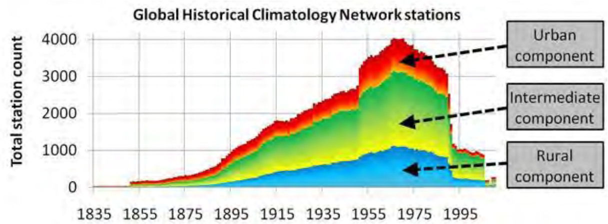


Figure 15: *Total number of stations with data for a given year in Version 2 of the National Climatic Data Center's Global Historical Climatology Network* homogeneity adjusted *monthly temperature dataset. The urbanization of the stations is determined by two metrics, i.e., associated population and night-light intensities. "Urban component" refers to stations considered urban by both metrics; "Rural component" refers to stations considered rural by both metrics; "Intermediate component" refers to all other stations.*

Moreover, many of those stations have relatively short records, meaning that they are only of limited value for assessing long term temperature trends. In particular, it can be seen from Figure 15 that the number of stations (rural or otherwise) is dramatically reduced after 1990, and before the 1950s. Although Peterson et al. predicted that the number of rural stations with post-1990 records in the Global Historical Climatology Network would improve "*with the creation of the Global Climate Observing System Surface Network*" [17], more than a decade later, there still does not appear to have been much improvement. For this reason, a large fraction of the Global Historical Climatology Network station records (rural or otherwise) only have a few decades data, and mostly during the period of roughly 1950-1990.

Figure 16 illustrates the locations of homogeneity adjusted Global Historical Climatology Network version 2 stations which have data for at least 75% of the 1880-1998 period considered by Peterson et al., 1999[17] from two subsets - the stations identified as rural in terms of both population and night-lights and the stations identified as urban by both terms (population >100,000 and high night-light intensities). It can be seen that outside of the contiguous

[12]Hence, if a grid box had no homogeneity adjusted records, Peterson et al. used unadjusted records for that grid box[17].

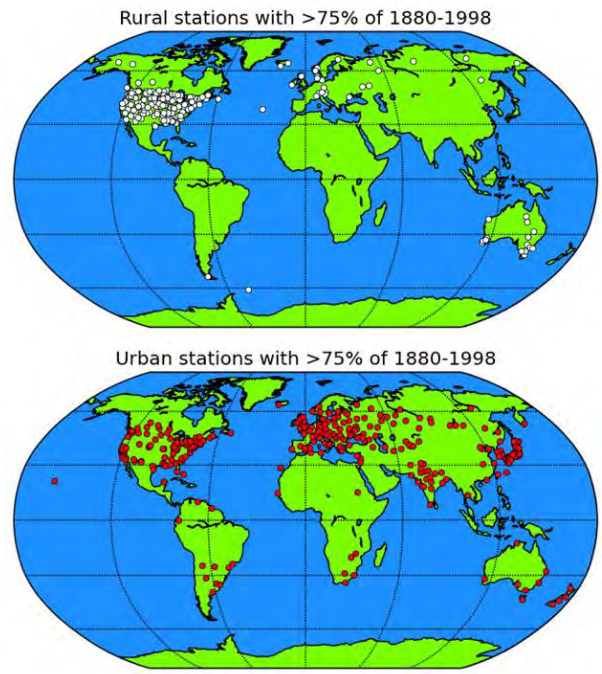


Figure 16: *Locations of the stations used by Peterson et al., 1999[17] which are rural (top) or urban (bottom) in terms of both population and night-light brightness and which have data for at least 75% of the 1880-1998 period.*

U.S. and possibly Australia, there are very few rural stations meeting those relatively basic requirements. In contrast, the urbanized stations have a considerably higher and more uniform station distribution.

This leads to three major concerns about the reliability of global temperature estimates based on the homogeneity adjusted Global Historical Climatology Network version 2 stations. First, aside from the contiguous U.S. (and to a lesser extent Australia), most of the 1401 rural Global Historical Climatology Network station records cover less than 75% of the 1880-1998 period. This means that most of the stations least likely to be affected by urbanization bias cannot be used for *directly* comparing the various warming and cooling trends since the late 19th century.

Second, the region which has (by far) the best coverage of long, rural records in the dataset, i.e., the contiguous U.S., is one which shows considerably more 1940s-1970s cooling and less 1980s-2000s warming than the "global" temperature estimates. This can be seen by comparing the mean temperature trends of the rural U.S. (Figure 6) and the globe

(Figure 2). The contiguous U.S. is only a small percentage of the global land mass (∼4%), and so one might argue that its trends are different due to regional variability[12]. But, Figure 16 suggests that the long-term trends of the rest of the world are more likely to be dominated by urban stations. If urbanization bias is substantial outside the contiguous U.S., then this could also explain the apparent difference between the U.S. and "global" temperature trends.

Finally, it raises serious questions over the robustness of the Global Historical Climatology Network's homogeneity adjustments. We discuss these adjustments in detail in Paper III[2], but a brief discussion is relevant here. The homogeneity adjustments used for version 2 of the Global Historical Climatology Network involved comparing each station's trends to the trends of neighbouring stations.

If a station's record showed a step-change relative to neighbouring stations in a given year, then the record was adjusted to remove that step change. Significantly, trend-changes were not considered, even though many non-climatic biases, including urbanization bias, involve trend changes rather than step changes. Some researchers have argued that step-change adjustments could remove some urbanization bias, anyway[194]. But, that should only occur if stations with urbanization bias are rare. From Figure 16, it can be seen that outside of the contiguous U.S., the opposite is the case. Indeed, it is likely that, if non-urbanization biased stations are rare (as is apparently the case here), such adjustments could actually *introduce* urbanization bias into the records of rural records - a process known as "urban blending". This is of particular concern, since the homogeneity adjustments of the rural stations used by Peterson et al., 1999[17] appear to have been carried out using all Global Historical Climatology Network stations, including the urban ones.

For all these reasons, Peterson et al., 1999's rural subsetting experiment was not a reliable approach to estimating the magnitude of urbanization bias on global temperature estimates.

4.6 Peterson, 2003

Peterson, 2003[18] criticised previous urbanization bias studies which had not accounted for other non-climatic biases in weather records. He decided to attempt to adjust his data to account for these biases before carrying out his own assessment of the magnitude of the urbanization bias. He chose 289 stations, which were grouped into approximately 40 different clusters[13]. Each cluster contained between 4 and 18 stations that were relatively close to each other. The clusters were reasonably evenly distributed across the contiguous U.S. (i.e., all of the U.S. except Alaska and Hawaii). The stations in these clusters were identified as "urban" or "rural" depending on satellite-based estimates of average night-light intensity.

He developed various adjustments for latitude, elevation, the time of day at which thermometers were reset ("time of observation") and the types of thermometer used. He ignored differences in station micro-climate, e.g. the presence or absence of nearby trees/buildings/pavements. However, he did remove 2 out of his 289 stations from his analysis for being rooftop stations.

He calculated a $0.31°C$ difference between his urban and rural stations, but claimed that when he applied his adjustments, most of this difference disappeared. He therefore concluded that the urban stations in his analysis had a negligible urban heat island, and that their apparent heat islands were instead due to urban stations having the following characteristics:

- Having a smaller fraction of stations at which thermometers were reset in the morning (37% of stations instead of 53% of rural stations).
- Being at lower altitudes (his rural stations were located on average 20m higher than his urban stations).
- Having a different ratio of thermometer systems (e.g., 13.6% of his urban stations used hygrothermometers, compared to 7.1% of rural stations)

He calculated that these differences were slightly counteracted by his urban stations being on average 0.02° further north than his rural stations.

He had a hunch that his analysis was unable to detect a significant urbanization bias because of a guess of his that a lot of urban weather stations might be located in city parks. He did not actually test whether his hunch was accurate or not, but he pointed out

[13] Peterson provided an enumerated list of the stations and the clusters that he used to McIntyre, who posted them on his Climate Audit website. However, the stations for Clusters 3, 19 and 30 were not listed on McIntyre's website. It is unclear whether this was an oversight of either McIntyre or Peterson, or whether Peterson dropped those clusters from his analysis. But, the 37 clusters included 288 stations, which is close to the figure of 289 which Peterson reported[18], so the following discussion assumes the latter, i.e., that there were only 37 clusters in Peterson's final analysis.

that parks and green space areas in cities are known to partially mitigate urban heat islands - the "park cool island" effect[195], schematically illustrated in Figure 1.

Peterson noted that Spronken-Smith & Oke, 1998 had found night-time cooling in parks could be similar to that in rural areas[195]. From this he appears to have concluded that park cool islands effectively counteract urban heat islands. This is surprising since Spronken-Smith & Oke were only arguing that "*[p]arks form depressions (or cool pools) in the warm urban landscape*" and in each of their examples, the parks, while cooler than their surroundings, still had urban heat islands[195]. Gaffin et al., 2008 specifically tested Peterson's hypothesis by evaluating the strength of the urban heat island in New York City (NY, U.S.)'s Central Park[60]. They found that Central Park had a substantial urban heat island, despite itself being a green area, contradicting Peterson's hypothesis.

Peterson also thought that urbanization processes should frequently lead to "urban cooling" - a theory also used by Hansen et al. to justify the Goddard Institute for Space Studies' urbanization adjustments, which we discuss in Paper II[1]. To explain why he could only find articles discussing urban heat islands, and not his hypothesised urban cool islands, he suggested that the scientific literature was biased[18]. However, as we mentioned in Section 3.4, urbanization bias is predominantly a warming bias. A simpler explanation would be if Peterson's analysis was flawed. Hence, it is worth reassessing his analysis.

Peterson had chosen a particular three year period (January 1989 to December 1991) for his analysis, apparently to "*avoid the confounding influence of the Automated Surface Observing System (ASOS) deployment, which started in 1992.*" This is an unusual justification, because there does not appear to have been a particularly noticeable drop in station changes in 1992 (see Figure 17). In addition, it was also a period which included the Mount Pinatubo volcanic eruption in July 1991. It has been argued that this eruption significantly influenced global temperatures (see e.g., Ref. [196] and references therein) for several years afterwards. If this theory is valid, then it may have unnecessarily introduced a confounding factor into Peterson's analysis.

McIntyre, 2007[197] carried out for his website a useful qualitative approach to assessing Peterson's analysis. On request, Peterson had provided a list of the stations he had used, as well as whether he

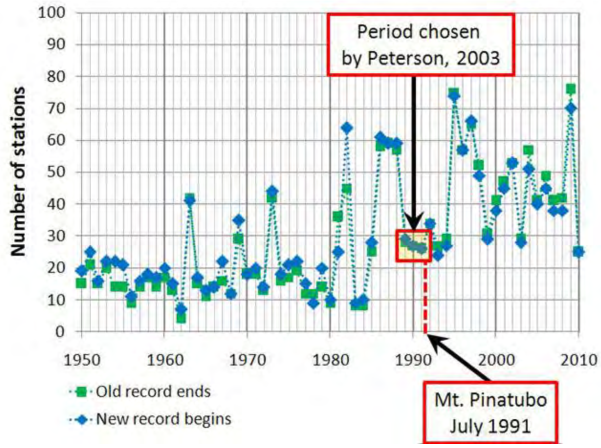


Figure 17: *Years in which the stations used by Peterson, 2003[18] underwent enough of a change that, officially, an old record was ended and a new record begun for that station. Based on metadata from the NOAA/NWS Cooperative Observer Network, using the station identifications listed on Climate Audit.*

regarded them as "rural" or "urban". McIntyre obtained temperature data for (most of) those stations from the National Climatic Data Center's Global Historical Climatology Network Daily dataset, and by simply averaging together trends from the stations in each subset, he was able to construct two temperature trend estimates[197].

There was a substantial growing divergence between the temperatures of the two subsets, suggesting urbanization bias[197]. Peterson, 2003 had proposed that the apparent difference between urban and rural stations were due to differences in location, as well as slightly different frequencies in the types of instruments and observation times used. These would be once-off differences, which would imply the current urban-rural difference had been relatively constant over time. However, McIntyre's analysis suggested a continually growing divergence between urban and rural stations over the entire 20th century, which would be more indicative of growing urban heat islands.

McIntyre's analysis was merely qualitative, since he had simply averaged together all stations in each subset. Hence, it is worth repeating his analysis using a gridded approach. We identified 283 of Peterson's 289 stations as having a Global Historical Climatology Network Daily record. Of those stations, Global Historical Climatology Network Daily records from each subset were converted to annual temperature

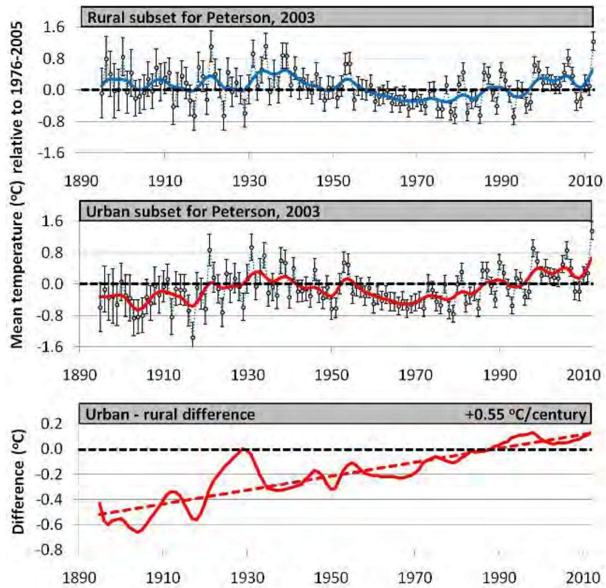


Figure 18: *Comparison of annual deviation from 1976-2005 mean temperatures for Peterson, 2003[18]'s rural and urban subsets, and the difference between them. Data calculated from the National Climatic Data Center's Global Historical Climatology Network Daily dataset, similar to the approach of McIntyre, 2007[197] - records for 283 of Peterson's 289 stations were available. Thick lines correspond to 11-point binomial running means. Bottom panel illustrates the growing divergence between the two subsets.*

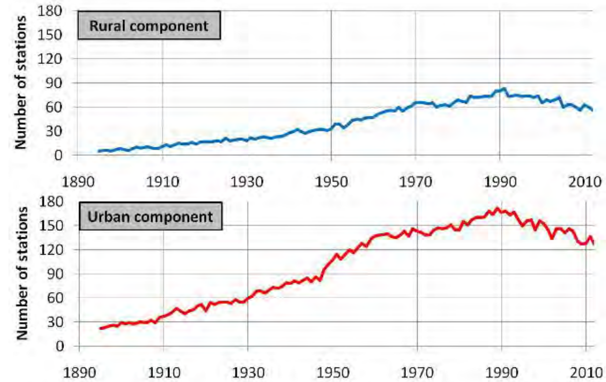


Figure 19: *Station numbers available for the Peterson, 2003 subsets for each year.*

deviations from the 1976-2005 mean for each station. The 1976-2005 period was chosen as this was the 30 year period with the greatest number of available stations - see Figure 19. These anomalies were then gridded into $5° \times 5°$ grid-boxes, before averaging together to yield a single estimate. The results are shown in Figure 18 and seem similar to McIntyre's non-gridded analysis, confirming his qualitative analysis.

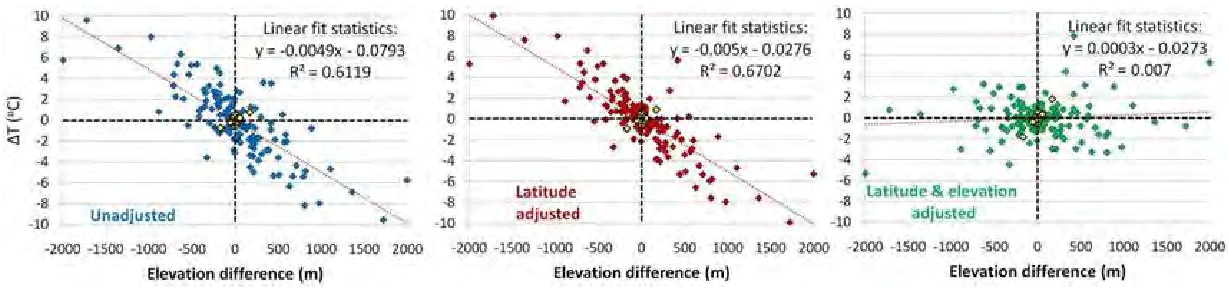
Gallo, 2005[198] attempted to overcome some of the problems of non-climatic biases that Peterson had expressed concern over, by using data from the National Climatic Data Center's United States Climate Reference Network dataset. Although only set up recently (starting in 2003), these Climate Reference Network stations are sited in rural locations, record hourly measurements and all use the same instrumentation. For this reason, the relationship between individual records is unlikely to be overly affected by (1) urbanization bias, (2) instrument bias or (3) time of observation bias. Hence, Gallo, 2005 used five Climate Reference Network station pairs to evaluate Peterson, 2003's latitude and elevation adjustments which Peterson had applied to his stations which may have been affected by those biases.

Gallo found that in four out of the five pairs, Peterson's latitude and elevation adjustments actually *increased* the temperature difference between the pairs. After applying Peterson's latitude adjustments, Gallo found that if he assumed the only remaining difference between the Climate Reference Network station pairs was due to elevation (as Peterson, 2003 implied), the apparent "lapse rate" (i.e., decrease in temperature with elevation) varied from -30.3 $°C\ km^{-1}$ to +83.1 $°C\ km^{-1}$[198].

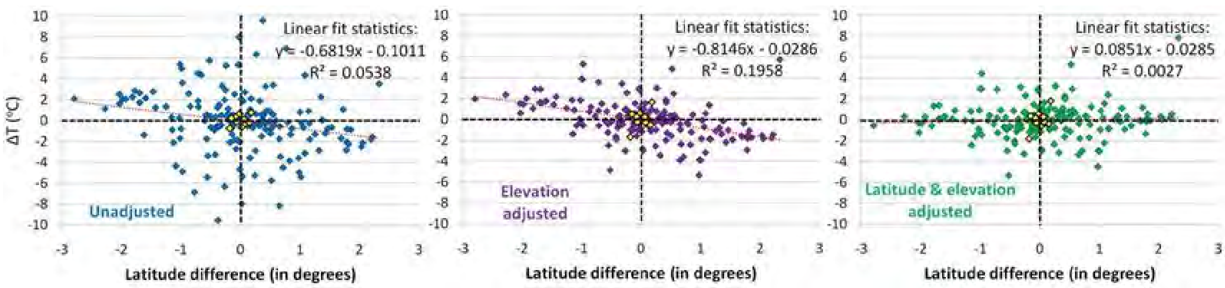
Such values were clearly unrealistic, and suggested to Gallo that other factors, such as the microclimate of the station strongly influenced the station temperatures. He suggested that Peterson, 2003's use of a constant lapse rate was inappropriate for adjusting the temperatures of weather stations at ground level[198].

Peterson & Owen, 2005 disputed Gallo's conclusions on the basis that he had only used five station pairs[199]. Gallo had limited his analysis to five pairs of neighbouring Climate Reference Network station pairs which had been specifically selected for inter-station comparisons. However, if we relax this requirement to all nearby Climate Reference Network station pairs, we can extend Gallo, 2005's analysis to overcome Peterson & Owen's criticism.

By 2011, 218 Climate Reference Network stations had been set up by the National Climatic Data Center throughout the U.S. (including Hawaii and Alaska). 179 of the stations in the contiguous U.S. had a full year of data for 2011 and a nearest neighbouring Cli-



(a) *Temperature - elevation relationship*



(b) *Temperature - latitude relationship*

Figure 20: *Average temperature relationships between United States Climate Reference Network weather stations and their nearest neighbours in 2011, before and after applying Peterson, 2003's latitude and elevation adjustments. Each point corresponds to the mean temperature difference between a station and its neighbour, for 2011. Yellow points correspond to the 5 station pairs used by Gallo, 2005[198].*

mate Reference Network station which also had a full year of data for 2011. We used these station pairs to re-assess Gallo, 2005's conclusions. The results of the analysis are shown in Figure 20.

Daily average temperatures were calculated for each Climate Reference Network station and its nearest neighbour by summing the maximum and minimum recorded temperatures over the 24 hour period (starting at midnight) and then dividing the sum by two. The difference between the average temperatures recorded at the two stations for that day was then determined, and monthly differences were calculated by averaging together each of these daily differences for a given month. These monthly differences were then averaged to yield a yearly average.

Nearest neighbours were at most 2.79° latitude away, and the differences in elevation were all less than 2km. Distances from nearest neighbours varied from 1.4 to 348.5km[14]. Station locations were extracted from metadata in the National Climatic Data Center's station update reports on their public ftp website, and elevations were then calculated using the GPS Visualizer website, which uses the U.S. Geological Survey's National Elevation Dataset for locations in the U.S.

From Figure 20, it appears that both Peterson[18] and Gallo[198] were at least partially correct. For the unadjusted stations, there appears to be a relatively strong linear relationship between temperature and elevation as Peterson had claimed (Figure 20a), although the temperature-latitude relationship appears quite weak, at best (Figure 20b).

The application of Peterson's adjustments appears to remove most of this relationship, and in this sense appears successful. However, the adjusted differences are still quite substantial. If Peterson's assumption that the only important differences between stations were latitude, elevation, instruments used, time of ob-

[14] Note that Gallo did comparisons between specific station pairs. In our analysis, each station was instead compared to its nearest neighbour in turn. When the nearest neighbour for that neighbour was calculated, this was not necessarily the same as the first station. For example, if three stations, A, B and C are in a row, B might be A's nearest neighbour, but B's nearest neighbour might be C. While 179 stations were analysed, only 129 stations were used as nearest neighbours (89 were used once, 34 were used twice and 6 were used three times).

servation, urbanization or the possibility of a station being on a roof, then the temperature differences of the adjusted data should be small, i.e., the *"residuals"* remaining after adjustment should all be close to zero. It can be seen from the two plots on the right hand side of Figure 20 that this is not the case.

This suggests that Gallo was correct in claiming that there are other important differences which need to be considered, such as station micro-climate[198]. Gallo's five station pairs all had relatively small elevation differences (the yellow points in Figure 20). It may be that elevation adjustments are important when the elevation differences are substantial. But, other factors than those Peterson considered also appear to be at least as important.

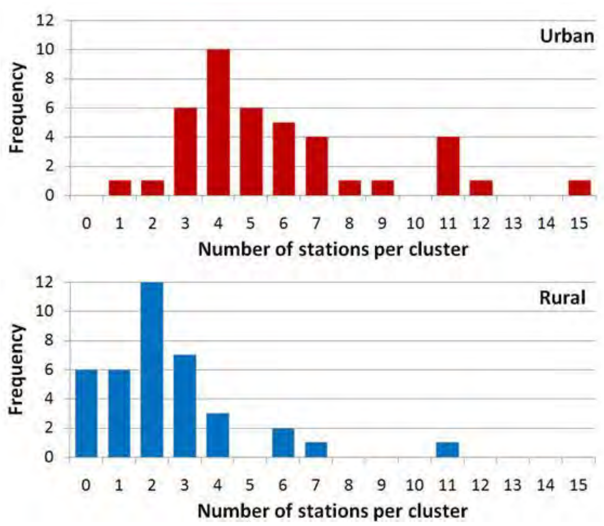


Figure 21: *Number of rural and urban stations included in Peterson, 2003[18]'s rural-urban comparison clusters. Rural-urban identification taken from Climate Audit.*

A serious difficulty with Peterson's analysis was his unusual selection of "urban" and "rural" stations in each cluster. Many clusters consisted of mostly urban stations, and six of the clusters had no rural stations in them - see Figure 21. Although Peterson dropped the clusters with no rural stations from his analysis[18], more than half of the remaining clusters only had one or two rural stations. Peterson calculated his rural/urban difference by subtracting the average of the "urban" stations from the average of the "rural" stations in each cluster. But, if the "rural" average for a cluster depended on just one or two stations, the analysis was highly dependent on those stations being representative of the non-urbanized climate of the region.

Peterson & Owen, 2005[199] later revisited this issue, and found that if different urbanization thresholds (which yielded more rural stations per cluster) were used, a substantial urban heat island could be detected. For this reason, they conceded that there had been some urbanization bias which Peterson, 2003 had failed to detect due to the use of an inappropriate threshold. They still believed that the effect of urbanization bias on U.S. temperature trends was very small. But, this appears to be based on an analysis of the homogenized version of the U.S. Historical Climatology Network, which as we show in Paper III[2], is contaminated by urban blending.

Summary of the flaws with Peterson, 2003

There were a number of flaws in Peterson, 2003's analysis. Some of these might not have been critical, e.g., the somewhat arbitrary period of analysis. But, there does seem to have been significant urbanization bias in his data (see Figure 18). The fact that Peterson was unable to detect this bias after making various non-urbanization adjustments[18] suggests that either his adjustments were problematic, his detection method was inadequate, or both.

4.7 Parker, 2004; Parker, 2006

Making the explicit assumption that urban heat islands *"are largely absent in windy weather"*[20], Parker [19, 20] proposed a new approach to quantifying the extent of urbanization bias in global temperature estimates. Parker created two different global temperature estimates from the same set of 265 stations. He constructed a "windy" subset based on station data for days associated with relatively high wind speeds in the vicinity of the station, and a "calm" subset based on daily data associated with relatively low wind speeds.

There was very little difference between the two subsets, and both subsets were similar to the Climate Research Unit's global temperature estimate. This led Parker to conclude that the urbanization bias in global temperature estimates is very small.

But, are urban heat islands largely absent in windy weather? Parker made this assumption from his interpretation of Johnson et al., 1991 [200]. However, it appears that Parker's interpretation was derived from a rather cursory reading, as Johnson et al., 1991 did not make any such claim. Rather, they observed that urban heat islands tend to reach their **maximum** "*a few hours after sunset on calm, cloudless summer nights*" (for most mid-latitude studies at least).

Johnson et al. suggested that this was because the difference between urban and rural locations was greatest (and therefore the urban heat island was at its maximum) when the rural locations were cooling rapidly (which often happens on cloudless nights), but the urban locations were cooling slowly.

As windy weather should tend to dissipate sensible heat from urban surfaces, it was suggested that the rate of urban cooling was lower on calm nights than windy nights. But, this is not the same as Parker's assumption that urban heat islands are "*largely absent in windy weather*". For instance, while Morris et al., 2001[201] found that Melbourne's night-time urban heat island was greater on calm, cloudless conditions, they also found that "*...even under conditions of strong winds and 8 octas of cloud cover, Melbourne exhibits* [an urban heat island]".

There certainly appears to be a link between urban heat islands and wind speed. Indeed, not only does wind speed influence urban heat islands, but urban heat islands may themselves influence wind speed and direction[202–205] (hence the existence of the term *"country breeze"*[202]). However, as Stewart, 2000 points out, there are other variables involved. In particular, cloud cover (which Parker found difficult to estimate) may often play a larger role than wind speed [40]. Moreover, as discussed in Section 2.2, it is not the maximum size of urban heat islands which is relevant to global temperature estimates, but rather the changes it introduces over time to average annual temperatures.

So, it appears that the fundamental assumption that formed the basis of Parker's study was based on an inaccurate interpretation of Johnson et al., 1991[200]. However, is it possible that his analysis could coincidentally work?

One way to test this is to see if there was evidence of urbanization bias in his data. Parker was unable to detect any urbanization bias in his selection of stations, using his windy/calm sub-setting approach. Therefore, if it transpires that his selection of stations *was* actually affected by urbanization bias, then this would indicate that his detection method was unreliable, in which case his conclusion would be invalid.

Monthly mean temperature records were available from the National Climatic Data Center's Global Historical Climatology Network Monthly (unadjusted) Version 3 dataset for 253 of the 265 stations used by Parker. The National Climatic Data Center also

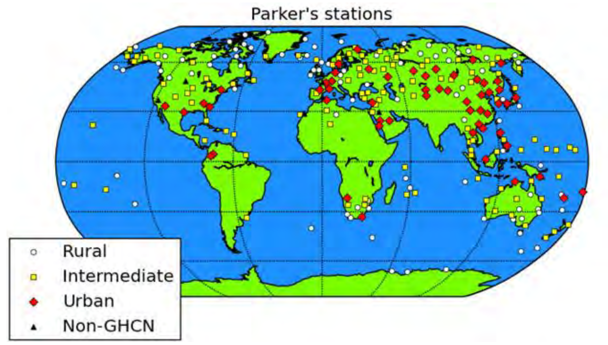


Figure 22: *Location of Parker's stations used in Figure 23, divided into subsets based on degree of urbanization. The locations of the 12 stations used by Parker which did not have a Global Historical Climatology Network record ("Non-GHCN") are also shown. Data taken from U.K. Meteorological Office's Hadley Centre.*

provide metadata for their Global Historical Climatology Network stations, suggesting how urbanized the station currently is.

Hence, using the National Climatic Data Center's metadata, it is possible to group Parker's stations into different subsets, based on their degree of urbanization:

Rural - 90 of Parker's stations (35.6%) which were associated with a low population ($< 10,000$) and night-light brightness.

Urban - 56 of Parker's stations (22.1%) which were associated with a high population ($> 100,000$) and night-light brightness.

Intermediate The remaining 107 of Parker's stations (42.3%) with Global Historical Climatology Network records.

The locations of the three subsets are illustrated in Figure 22. All three subsets are quite small. However, they each have a *fairly* similar distribution across the globe. For this reason, they should each provide similar estimates of global temperature trends, provided that the urbanization bias is as negligible as Parker had claimed.

For each of the subsets, the annual temperature trends for the stations were determined from the Global Historical Climatology Network Monthly (unadjusted) Version 3 dataset. These trends were then rescaled to the deviations from their 1961-1990 mean temperature, and binned into $5^\circ \times 5^\circ$ grid-boxes. The

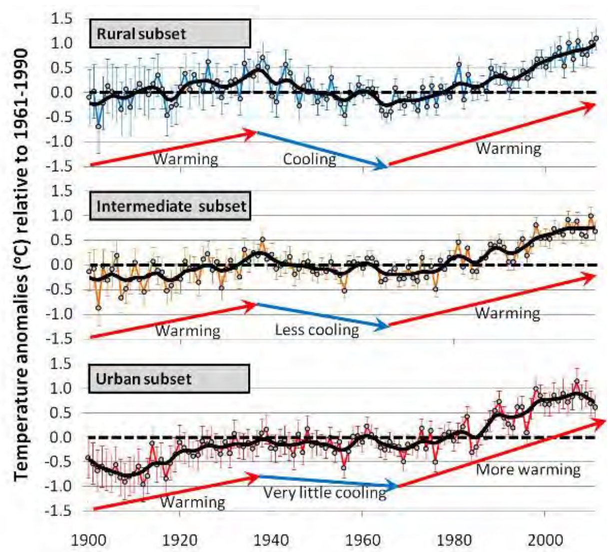


Figure 23: *Mean annual gridded temperature trends, relative to 1961-1990, of the Rural, Intermediate and Urban subsets of the 253 stations used by Parker that are also in the National Climatic Data Center's Global Historical Climatology Network Monthly (unadjusted) Version 3 dataset. Solid black lines correspond to 11-point binomial smoothed trends. Error bars correspond to twice the standard error of the gridded mean for each annual value.*

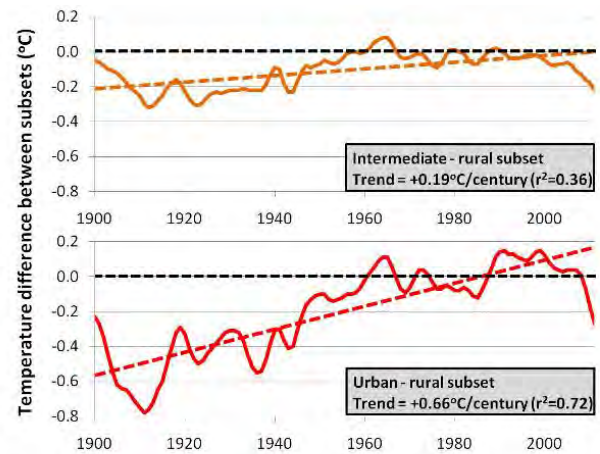


Figure 24: *Difference between the 11-point binomial smoothed trends of the Intermediate and Rural subsets (top) and the Urban and Rural subsets (bottom) from Figure 23.*

rescaled trends in each grid-box were averaged together, and the average grid-box trends were then averaged together (weighting by the cosine of the latitude of the middle of each grid-box) to yield a single global temperature estimate for each subset.

The mean annual gridded temperature trends from 1900 to 2011 of all three subsets are shown in Figure 23. Unfortunately, the total number of stations used by Parker is quite small, and when divided into three subsets, this number obviously is further reduced. Hence, the error bars for individual annual temperature anomalies is quite large. Nonetheless, there are noticeable differences between the long-term trends of the subsets. The rural subset shows the least warming of the three and the urban subset shows the most warming of the three, i.e., what would be expected if the stations are affected by urbanization bias. This is even more apparent in Figure 24. This indicates that the stations *are* affected by urbanization bias, despite Parker's claim.

If Parker's windy/calm sub-setting approach had been successful in detecting urbanization bias in his selected stations, then it should have been able to detect the urbanization bias which is apparent from Figures 23 and 24. This suggests to us that his approach is *not* successful. This agrees with the recent findings of McKitrick, 2013[61]. Therefore, Parker's claim that the effects of urbanization bias on global temperature estimates are negligible is invalid.

As an aside, some readers might wonder if we could use the rural subset of Figure 23 as a reliable estimate of the true global temperature trends since 1900. We would advise against this. It is reasonable to assume that the extent of urbanization bias is substantially reduced in the rural subset - although probably not eradicated, since urban heat islands can occur for even modestly urbanized stations, e.g., Hinkel et al., 2007[42]. However, the subset only contains 90 stations, so is quite a small sample size. In addition, most of the station records are quite short and contain other non-climatic biases aside from urbanization bias as well as a lot of data gaps. This can be seen by examining the six station records in Figure 5 which are all from the Parker subsets (three from the rural subset and three from the urban subset). Instead, we use the three subsets merely to illustrate that the more urban subsets show a warming bias relative to the less urban subsets (Figure 24), which indicates that urbanization bias is a problem for the stations Parker considered, yet his method for identifying this bias did not manage to detect it.

4.8 Efthymiadis & Jones, 2010

Efthymiadis & Jones, 2010 attempted to determine an upper bound for urbanization bias in the land-based global surface temperature trends by directly comparing the gridded mean temperature trends of various coastal 5° × 5° grid boxes ("land temperatures") with the gridded mean Sea Surface Temperature (SST) trends of the neighbouring grid boxes[21] ("marine temperatures").

They analysed 8 regions (44 gridboxes) from the northern extratropics, 5 regions (20 gridboxes) from the tropics/subtropics and 5 regions (20 gridboxes) from the southern extratropics. For each of these regions they calculated the linear trends for both the land temperature trends and the marine temperature trends over three periods: 1951-2009, 1951-1979 and 1979-2009. Since urbanization bias should not directly affect the marine temperature trends, they assumed that the difference between these values should give an upper bound for urbanization bias.

Although most of the regions showed greater warming linear trends for the coastal grid boxes than the marine grid boxes, three of the five regions in the southern extratropics showed much greater warming linear trends for the marine grid boxes than the coastal grid boxes. Therefore, when they averaged together the linear trends for their three latitudinal zones to generate their "global" estimates, the net difference between the land and marine linear trends was only 0.02°C/decade (0.2°C/century) for the 1951-2009 period. This was only 14% of the +0.14°C/decade linear trend for their coastal land "global" temperatures. Moreover, since ocean temperature trends tend to be lower in magnitude than land temperature trends (due to the large heat capacity of the oceans), they argued that this 0.02°C/decade extra warming for land was to be expected. On this basis, they concluded that urbanization bias was at most a small problem.

There are several flaws in their analysis. First, as we discussed in Section 3.1, using linear trends as the basis for an analysis can give misleading results when your data contains non-linear trends, as is the case here.

Second, as we discussed in Section 3.3, there are known biases and problems with the sea surface temperature data. Although various attempts have been made to correct for these biases, it is unclear exactly what adjustments are required, e.g., see Kennedy et al., 2011 for a summary[138]. With this in mind, it is worth noting that several groups have actually used the land-based temperature estimates to develop and/or justify their sea surface temperature adjustments, e.g.,

> *"In the* a posteriori *approach, exemplified by Jones et al. (1986c), adjustments are made so that the hemispheric means of* [the Marine Air Temperature and Sea Surface Temperature data]*are in accord with the near-propinquitous land-based data on decadal and longer time scales."* - Farmer et al., 1989, p5[206]

In those cases where the various adjustments (and/or lack of adjustments) of the sea surface temperature data are being justified by assuming the land-based data is more reliable, it is circular logic to *then* use the adjusted/unadjusted sea surface temperature data to claim the land-based data is reliable (i.e., unaffected by urbanization bias).

Third, as discussed earlier (particularly Section 3.3), the challenge in resolving the urbanization bias problem is *not* in establishing whether or not there has been "global warming", but in establishing how urbanization bias has affected the long-term global temperature trends. E.g., how does the recent warm period compare with the early 20th century warm period in Figure 6? Since the Efythmiadis & Jones, 2010 analysis only considered the 1951-2009 period, they were unable to directly compare the recent warm period to pre-1951 temperatures.

Nonetheless, even if we disregard the above flaws, a close inspection of their results reveals that their main conclusion is unjustified by their data. For their analysis, they considered 18 regions (84 grid boxes) divided into three latitudinal zones, i.e., "northern extratropics", "tropics/subtropics" and "southern extratropics". Although technically the simple mean of the 1951-2009 linear trends for these three zones gives a relatively small "upper bound" for urbanization bias of 14%, this seems to be a result of the unusual weighting they applied to obtain their "global average".

For two of their three zones (northern extratropics and tropics/subtropics), the average 1951-2009 linear trends was 50% greater for the land temperatures than the marine temperatures. That is, the average land trends were 0.15°C/decade while the average marine trends were only 0.10°C/decade. This would imply an "upper bound" of 0.05°C/decade (0.5°C/century), which is considerably more substantial than their conclusion, i.e., urbanization bias could account for up to 1/3 of the 1951-2009 linear trends.

However, for the southern extratropics zone, three of the five regions they considered ("West Australia", "East Australia" and "North Chile") yielded strongly negative land-marine trends: -0.11°C/decade, -0.05°C/decade and -0.16°C/decade respectively. Although these regions only accounted for 1/6 of the regions and 12 out of the 84 grid boxes they considered, because they only considered five regions for the southern extratropics zone, this meant that the zonal "average" land-marine trend was -0.04°C/decade. Since they applied equal weighting to all three of their latitudinal zones, this meant that for their "global" average, the net land-marine trend was only 0.02°C/decade or 14% of the land trends.

Indeed, if they had removed the two Australian regions and the North Chile region from their analysis, the remaining two regions ("South Africa" and "Argentina-Brazil") both had positive land-marine trends over the 1951-2009 period: +0.04°C/decade and +0.08°C/decade respectively.

In other words, Efthymiadis & Jones, 2010's conclusion that "global" surface temperature trends were relatively unaffected by urbanization bias was essentially a consequence of just three regions: West Australia, East Australia and North Chile. We note that the average population density for these two countries is below the world average (54 persons/km^2 in 2013) - in 2013, Australia had an average population density of 3 persons/km^2 and Chile had an average population density of 24 persons/km^2. So, although urbanization bias may also be a problem for these countries (e.g., see our discussion of Australia in Section 4.3), it is unwise to extrapolate these results to applying to "global" averages - which Efthymiadis & Jones effectively did.

4.9 Wickham et al., 2013

Recently, Wickham et al., 2013[22] carried out a subsetting study and concluded that urbanization bias has a negligible effect on global temperature estimates.

The Wickham et al. study was based on a far larger dataset than previous studies, i.e., the new Berkeley Earth Surface Temperature dataset. This comprised 39,028 stations. Using "Moderate Resolution Imaging Spectroradiometer" (MODIS) satellite estimates[163], they identified 16,132 (41.3%) of these stations as rural[15]. For our analysis in this section, we downloaded lists of the station identifications used for Wickham et al., 2013 from the Climate Audit website.

The Berkeley Earth group archived their first official dataset in February 2012. But, the available rural/urban station IDs refer to the October 2011 "Preliminary dataset" used by Wickham et al., and hence the following analysis refers to the earlier version. Wickham et al.'s study comprised two parts.

In the first part, the linear trends for all the stations were calculated. Then, histograms of the trends were plotted for both the complete set and the rural subset. A broad distribution of trends was obtained, particularly for the stations with the shortest records. However, less than a third of the stations had negative trends, and the median trend was positive. The distributions were similar for both the complete set and the rural subset.

For these reasons, Wickham et al. argued that on average there has been "global warming" and that this is apparent in the rural stations as well as the non-rural stations. They then concluded that this "global warming" was not due to urbanization bias, and that "*the effect of urban heating on the global trends is nearly negligible*".

This is flawed logic. As we discussed in Section 3.3, the issue is *not* whether or not there has been "global warming", but establishing by how much warming trends have been overestimated and cooling trends underestimated by urbanization bias. Indeed, Wickham et al., noted that between a quarter and a third of their stations showed cooling trends, contradicting the popular notion of almost continuous global warming since at least the late 19th century[35].

While Wickham et al. conceded that their linear trend analysis was "*a very crude way to look at global temperature change*", they do not appear to have realised just how crude, and inappropriate it was. As mentioned in Section 3.1, linear trends are of dubious value when describing non-linear data. This can be illustrated by the following thought experiment.

Let us suppose that global temperature trends for the last few centuries were exactly described by a sine wave, with a period of several decades (Figure 25). By definition, such a periodic function would have no long term trend, but it would go through multi-decadal periods of "global warming" and "global cool-

[15] They used the term "very-rural", but did not offer any justification for the "very-" prefix, so the conventional "rural"/"urban" nomenclature will be retained here.

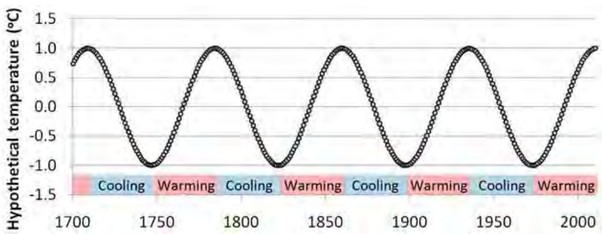


Figure 25: *Hypothetical "global temperature" changes from 1701 to 2010, used for comparison to the temperature data used by Wickham et al.[22].*

ing".

The bottom panel of Figure 26 illustrates the linear trend histograms which Wickham et al.'s stations yield if all of the available annual temperatures in their records are replaced with the corresponding hypothetical "temperatures" of Figure 25 for that year. The distribution for these hypothetical temperatures *also* shows a majority of "warming" trends, like the distribution calculated from the actual temperatures (top panel), i.e., the ones reported by Wickham et al.[22]. In fact, from Table 4, it can be seen that the mean and median linear trends of the hypothetical temperatures are actually greater, and would therefore by Wickham et al.'s logic indicate more "global warming". However, by definition, there is **no** long-term "global warming" trend for our hypothetical temperatures. In other words, Wickham et al.'s trend analysis did **not** show that the "global warming" of recent decades was unusual or unprecedented.

The second part of the Wickham et al. study was a rural sub-setting experiment, similar to the Hansen & Lebedeff, 1987[13] and Peterson et al., 1999[17] experiments described in Sections 4.1 and 4.5. Like the other studies, Wickham et al., 2013 also found little difference between their rural subset and their complete set. On this basis, they concluded that their estimates were not significantly affected by urbanization.

The main difference between the Wickham et al., 2013 sub-setting experiment, and the previous experiments was that they had used a much larger selection of stations, particularly for the post-1970s period, i.e., the new Berkeley Earth dataset. However, even though this dataset has an impressive total number of stations (nearly 40,000), the number of stations is still dramatically reduced for the early 20th century (Figure 27). Just like the Global Historical Climatology Network discussed in Section 4.5, the problem is worse for the rural stations than the urban stations,

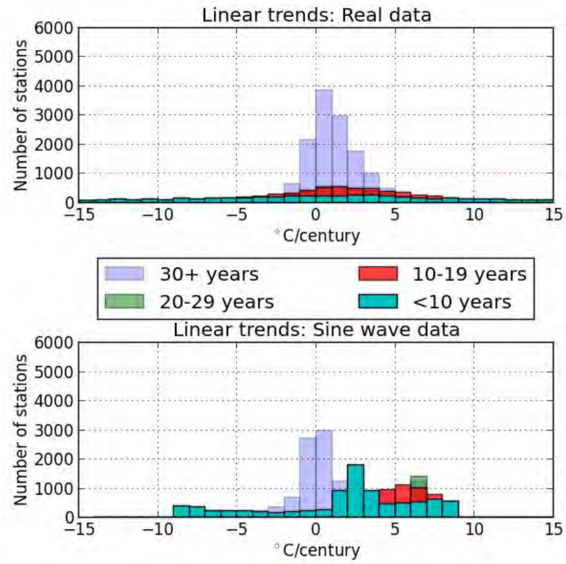


Figure 26: *Histograms of "linear trends" for different subsets of the Berkeley Earth station records, either (top) using actual temperature data, or (bottom) substituting annual values with the equivalent "temperature" from the hypothetical temperature curve of Figure 25.*

and rural stations comprise less than 15% of the stations for most of the 19th century (Figure 28).

Another problem which the Berkeley Earth dataset shares with the Global Historical Climatology Network is that the spatial distribution of rural stations becomes increasingly uneven for the earlier periods. It can be seen from Figure 29 that, in terms of stations with relatively long and complete records, the Berkeley Earth dataset is not a whole lot better than the dataset used by the earlier Peterson et al., 1999 study discussed in Section 4.5. It is true that the

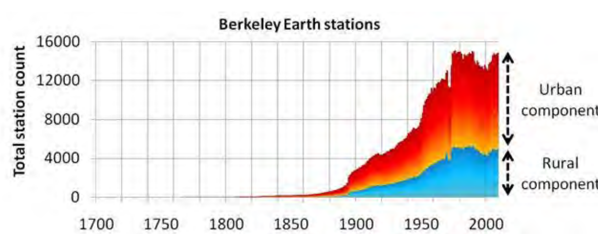


Figure 27: *Total number of Berkeley Earth stations available for a given year. The relative fraction of these totals which is urban or rural are also indicated.*

		Actual		Hypothetical	
Subset	Stations	Median	Mean ± σ	Median	Mean ± σ
All	31353	1.02	0.35 ± 16.33	2.35	2.05 ± 3.92
Rural	12935	1.09	0.35 ± 16.95	2.43	2.17 ± 3.97
Urban	18191	0.98	0.34 ± 15.24	2.24	1.98 ± 3.85
< 10 years	7331	-0.44	-2.69 ± 32.52	2.67	1.86 ± 4.59
10-20 years	6263	1.64	1.29 ± 8.02	4.81	3.07 ± 4.58
20-30 years	4074	1.55	1.74 ± 4.02	4.71	3.00 ± 4.31
> 30 years	13685	0.94	1.14 ± 1.82	0.62	1.41 ± 2.75

Table 4: *Median and mean values of "linear trends" for different subsets of the Berkeley Earth station records, either using actual temperature data, or substituting annual values with the equivalent "temperature" from the hypothetical temperature curve of Figure 25. Note that the standard deviations of the means are genuinely that large - this can be partially seen by considering the histograms of Figure 26, but the complete spread of the non-hypothetical histogram is not shown, as the histograms are truncated to the range, -15 . . . 15, following the approach of Wickham et al.[22]. See supplementary information for complete spread.*

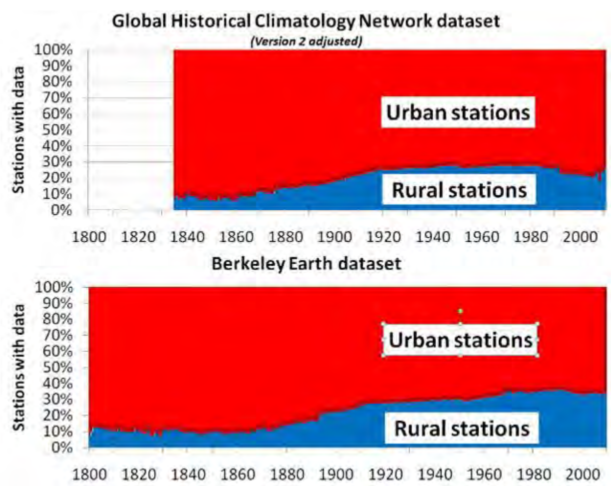


Figure 28: *Percentage of stations with data for a given year which were identified as rural, for two datasets: the Berkeley Earth Surface Temperature dataset used by Wickham et al.[22] and Version 2 of the National Climatic Data Center's homogeneity adjusted Global Historical Climatology Network dataset as described in Section 4.5.*

number of rural stations in Figure 29 is greater than in Figure 16. But, as in Figure 16, outside of the U.S. (and, perhaps, Europe), the density of rural stations is very low, while the urban stations are relatively well-distributed.

Wickham et al. used the homogenization and averaging algorithms described by Rohde et al., 2013b[113] before making their rural sub-setting comparisons. Rohde et al.'s algorithm involves several techniques which are different from those used by the other studies re-assessed in this article. Their algorithm allows the use of all station records with at least 2 months of data. Hence, they use a lot of relatively short records, and <42% of their stations have more than 30 years of data (Table 4). They also use a more complex spatial averaging process (related to Kriging) than the simple gridding approach adopted in this article, and by others[131].

To homogenize their data, Rohde et al.[113] first implement a technique they refer to as *"the scalpel"*. This splits a station record into two separate records whenever a step change is identified. Their step change identification procedure is based on internal changes in a record, and does not consider neighbouring stations, unlike those used by the National Climatic Data Center (either the one discussed in Section 4.5 or the more recent Menne & Williams, 2009 algorithm[194]).

Rohde et al., 2013b also implement a weighting algorithm which in theory could account for some trend-change biases. Stations whose record trends differ strongly from neighbouring stations are given a low weight, and so have a low contribution to their total temperature estimates. Hence, Wickham et al. optimistically suggest that "*the influence of sites with anomalous trends, such as urban heat island effects, should be reduced by the averaging procedure even when sites with spurious warming are part of the dataset being considered*"[22]. However, they do not appear to recognise that, in heavily urbanized areas, **or** areas with a low number of rural stations, the rural trends may be the ones regarded as *"anomalous"* by Rohde et al.'s approach. From Figures 28

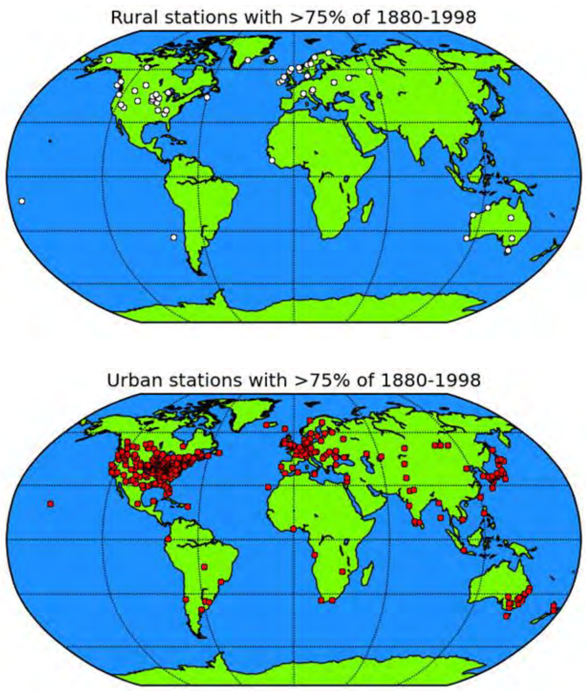


Figure 29: *Locations of rural (top) and urban (bottom) stations in the Berkeley Earth dataset, which have data for at least 75% of the 1880-1998 period, for comparison with the Global Historical Climatology Network (adjusted, version 2) dataset used for Figure 16.*

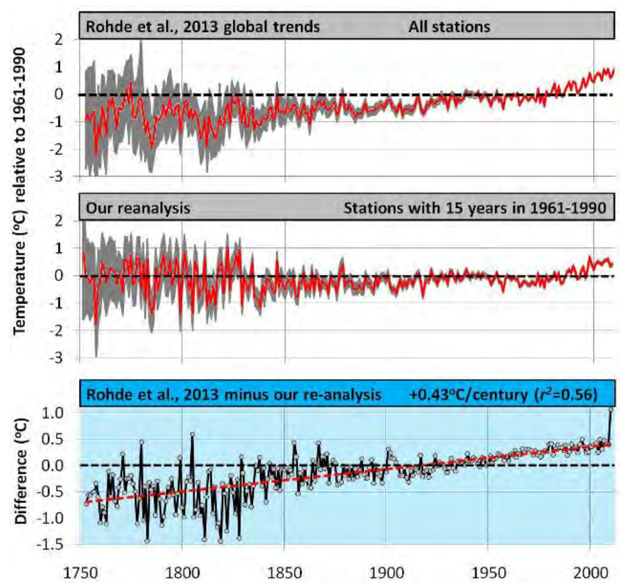


Figure 30: *Comparison between Rohde et al., 2013a[8]'s global temperature estimate (top) using the Berkeley Earth dataset and our alternative estimate (middle) described in the text. The bottom panel shows the annual difference between the two estimates. The grey bands in the top and middle panels correspond to twice the standard error of the means.*

and 29, it can be seen that this may be a frequent problem for the Berkeley Earth dataset.

Figure 30 illustrates the difference between Rohde et al.'s approach and the approach adopted in this article, i.e., simple gridded averaging of non-homogenized station deviations from their 1961-1990 mean temperature, for all stations with at least 30 years of data, and 15 years of data in the 1961-1990 period. The differences are striking. Instead of Rohde et al.'s relatively smooth and continuous global warming from 1800 to present, our reanalysis shows considerable multi-decadal variability.

It can be seen from the bottom panel of Figure 30 that the Rohde et al., 2013b approach to averaging the Berkeley Earth data introduces a warming trend of about $+0.43°C$/century, relative to our reanalysis. Our reanalysis is a relatively simple approach, similar to that used by the other groups in Table 1. So, if our approach is unreliable, then this would suggest that all of the other global temperature estimates in Table 1 are similarly unreliable. This may well be the case. But, if so, then it suggests that determining global temperature trends is highly dependent on the statistical sampling approach taken. This would mean *all* of the estimates in Table 1 would need to be treated cautiously, until the effects of the various statistical approaches have been carefully studied - regardless of the urbanization bias problem. We note that Rohde et al., 2013a argue that their averaging method is comparable to the other methods, and so they do not appear to be making this claim[8].

Although the rate of global warming since the mid-20th century is noticeably reduced in our reanalysis, it is still sufficiently high to make the recent decades seem the warmest since 1880, i.e., when the other estimates in Figure 2 begin. However, interestingly, our reanalysis suggests that temperatures in the 18th and early-19th centuries were comparable to recent decades. The global temperatures of recent decades do not appear particularly unusual in this context.

Unfortunately, this does not tell us anything quantitative about the extent of urbanization bias in the estimates (either in our reanalysis or in the original Rohde et al. analysis). Still, by comparing the spatial and temporal distribution of the rural and urban stations, we can draw some relevant conclusions.

Figure 29 illustrates that rural stations are in the minority. In particular, before the late 19th century, the vast majority of the stations with available data are currently urban. It seems reasonable to assume that many of the stations that are currently urban are considerably more urbanized than they were in the 19th century. This suggests that many of the stations with data for the 19th century have been subject to urban development since the start of their records, and are therefore likely to be affected by urbanization bias. For this reason, we argue that it should be assumed that comparisons between 19th century temperatures and 20th century temperatures are at least partially affected by urbanization bias.

Further, on the basis of Figure 29, we suggest that the density of rural stations with relatively long and complete records is actually too low for reliably estimating *global* temperature trends for more than a few decades. This means that the 20th century temperature trends of both estimates are probably dominated by urban stations. Hence, we argue that we should also assume that comparisons between early 20th century and late 20th/early 21st century temperatures are partially affected by urbanization bias too.

While there does not appear to be enough rural data for a long-term global temperature estimate, Figure 29 suggests that it may be possible to construct reasonable estimates *regional* 20th century temperature trends for the United States, Europe and Australia by only using rural stations with relatively long and complete records. We have not yet analysed the Australian rural records in detail, but in Figure 6 we presented the rural 20th century temperature trends for the contiguous United States and in Figure 3 we presented the temperature record for one of the longest rural European records, i.e., Valentia Observatory (Ireland). Neither of those figures agree with the almost continuous "global warming" trends implied by the current global temperature estimates (Figure 2). Instead, as we discussed in Section 3, they indicate a multi-decadal alternation between periods of warming and periods of cooling. This suggests to us that the true long-term global temperature trends would look markedly different from the current estimates, if a large enough distribution of rural stations with long, relatively complete records could be obtained.

For these reasons, we conclude that the Wickham et al., 2013 study was unable to reliably estimate the true extent of urbanization bias in their global temperature estimates.

4.10 Hansen et al., 1999-2010 studies

As mentioned earlier, the Goddard Institute of Space Studies is currently the only group that explicitly attempts to correct their global temperature estimates for urbanization bias (Table 1). Their urbanization bias adjustments were introduced by Hansen et al., 1999[11], and some subsequent modifications to these initial adjustments were described in the follow-on papers, Hansen et al., 2001[12] and Hansen et al., 2010[3].

The net effect of their urbanization adjustments on the overall trends of their global temperature estimates is quite small ($\sim 0.1°C$/century). As a result, the urbanization bias adjusted Goddard Institute of Space Studies global temperature estimate is remarkably similar to the estimates without an explicit urbanization bias adjustment, i.e., the estimates in Figure 2. This can be seen from Hansen et al., 2010's Figure 11, for instance[3].

The three Hansen et al. studies argue that the similarity between their urbanization adjusted estimates and estimates without urbanization adjustments indicates that the effects of urbanization bias on global temperature estimates are small or negligible[3, 11, 12]. But, as we discuss in Paper II[1], there are a number of serious problems with the urbanization adjustments applied by the Goddard Institute of Space Studies. We find that their adjustments are seriously flawed, unreliable and inadequate. Hence, we cannot rely on the small net magnitude of their adjustments as an accurate estimate of the actual urbanization bias in current global temperature estimates.

A detailed discussion of the Goddard Institute of Space Studies' urbanization adjustments, and their reliability (or lack thereof) is beyond the scope of this article. Instead, we assess their adjustments separately in Paper II[1]. Nonetheless, the three Hansen et al. papers describing their adjustments[3, 11, 12] also use other arguments to conclude that the net effect of urbanization bias on global temperature estimates is small. Hence, it is worth briefly considering these other arguments and assessing their validity here.

Both Hansen et al., 1999 and Hansen et al., 2010 claim that the net bias introduced by unaccounted for urban effects must be small, because there is other evidence of "global warming" over the past century, such as studies of glacier lengths and borehole temperatures[3, 11]. However, as we discussed in Section 3.3, this is a logical fallacy, since the urbanization bias problem is not over whether or not there have been periods of "global warming" and "global cooling", but rather establishing to what extent urban heat islands have introduced artificial warming biases into weather station-based global temperature estimates.

Hansen et al., 1999[11] also carry out a sub-setting experiment similar to the Peterson et al., 1999 study[17] discussed in Section 4.5. Stations with associated populations greater than 50,000 were considered "urban stations", those with associated populations less than 10,000 were considered "rural stations" and those with intermediate associated populations were considered "small-town stations".

Using these definitions, they constructed four different global temperature estimates, distinguishing weather stations on their associated populations. In the first estimate, only rural stations were used; in the second estimate, rural and small-town stations were used; in the third estimate, all stations were used; in the final estimate, all stations were used, but they applied their urbanization adjustments to the urban stations. All four estimates were similar, and on this basis they concluded that the urban influence on their estimates was small. However, most of the flaws of the Peterson et al., 1999 study that we discussed in Section 4.5 also applied to this study. The Hansen et al., 1999 sub-setting experiments also had an additional flaw in that they only used one metric for identifying urban stations, i.e., the associated population size. In contrast, the Peterson et al., 1999 study had at least used two metrics for identifying urban stations (associated population size *and* night-light brightness), offering a stricter detection method.

Hansen et al., 1999 also carried out a third test of the effect of urbanization bias on their estimates by considering in detail the *regional* effects on the contiguous U.S. - a region with one of the highest densities of stations, as we discussed in Section 2.4. Although they seem to have reached the conclusion from this test that these effects were small, their actual results seem to us to suggest the opposite. For instance, they found that urbanization bias changed the relative warmth of the years 1934 and 1998. In their rural subset, 1934 was the 20th century's hottest year in the contiguous U.S., and the 1920s-1930s seem to have been generally warmer than the late 20th century. However, when they used all stations (i.e., rural, small town and urban), this increased the apparent warmth of the 1980s-1990s, making the two warm periods seem comparable, and 1998 a close contender to 1934 for the hottest year. When they applied their urbanization adjustments to the urban stations, this gave an intermediate result. In other words, as we noted in Section 3.3, urbanization bias for the U.S. is substantial enough to alter the relative warmth of the early and late 20th century warm periods.

Another finding of the Hansen et al., 1999 U.S. case study, which appears to us to contradict their conclusion, is their analysis of the 1950-1998 linear trends of their four subsets. For their rural subset, the 1950-1998 linear trend for the contiguous U.S. was a cooling one, while for their unadjusted subset containing all stations, the trend was a warming one. Their subset in which they applied urbanization adjustments to the urban stations was again intermediate between the two, with almost no trend. In other words, urbanization bias alters the 1950-1998 linear trends for the U.S. so much that the sign of the trends changes. We find it hard to reconcile these findings with their claim that *"(t)he temperature curve, based on rural stations, is not affected much by addition of small-town or urban data"*.

Hansen et al., 2001 introduced a number of modifications to their global temperature analysis[12]. Several of these modifications were confined to the contiguous U.S. component of their analysis. In particular, they switched to using the homogeneity-adjusted version of the U.S. Historical Climatology Network (version 1). As we mentioned in Section 2.5, and discuss in Paper III[2], one of the main effects of these homogeneity adjustments is to introduce a warming trend into U.S. temperature trends, thereby increasing the apparent warmth of the late 20th century for the U.S.

Hansen et al., 2001 also carried out a study of the effect of urbanization bias on regional temperature trends for the contiguous U.S. From this study, they reached the conclusion that globally, *"the effect is modest in magnitude"*. However, again, this seems to be contradicted by their actual findings. If they used the unadjusted version of the U.S. Historical Climatology Network, the long-term linear trend for the contiguous U.S. was $+0.16°C$/century. When they used the homogeneity-adjusted version, this trend nearly trebled to $+0.46°C$/century. However, when they carried out their urbanization adjustments this reduced the trend to $+0.32°C$/century. In other words, their estimate of urbanization bias for the contiguous U.S. was $+0.14°C$/century, which is either 30.4% of the trend for the homogeneity-adjusted data or 87.5% of the trend for the unadjusted data. This hardly seems *"modest in magnitude"*.

5 Conclusions

A number of studies[13–20, 22] have claimed that urbanization bias has only had a small or negligible effect on global temperature estimates derived from weather station records. However, in this article, each of those studies was systematically reassessed and found to be flawed.

Although determining the exact magnitude of the bias was beyond the scope of this article, it does seem to have led to a substantial underestimation of post-1940s global cooling and a substantial overestimation of post-1970s global warming. This suggests that the simple description of almost continuous global warming since the late 19th century which has previously been suggested[3–9] is wrong. Instead, temperatures appear to have alternated between multi-decadal periods of global warming and global cooling.

As we discussed in Section 3.3, the 1980s-2000s global warming seems to have been a genuine (albeit overstated) phenomenon. However, when we take into account the urbanization bias problem, it is plausible that it was just as warm in the 1930s/1940s. In that case, the recent warm period would be neither unprecedented nor unusual. We note that this would contradict the current climate models which assume that increasing atmospheric carbon dioxide concentrations have caused an unusual "anthropogenic global warming" over the last few decades[36].

As far as we have been able to determine, the ten sets we revisited in this article are the entire basis in the current literature for the claim that urbanization bias only has a small or negligible effect on the global temperature estimates derived from weather station records. We believe that we have shown that in each of the cases, the basis for this claim was unjustified. We did not manage to determine the exact magnitude of the bias, but it appears to be substantial.

We hope that we have adequately highlighted some of the challenges inherent in solving the urbanization bias problem, so that future researchers will be better able to tackle it. We suggest that it is now time for researchers to stop trying to disprove and/or dismiss the existence of urbanization bias in the current estimates, but instead return to trying to quantify and/or remove the bias.

Regardless, the findings of this article show that the decision of most of the groups constructing global temperature estimates from weather station records[4–9] not to attempt to explicitly correct for urbanization bias was unjustified. In Paper II, we show that the adjustments devised by the only group which does attempt this, i.e., the Goddard Institute for Space Studies[3] are flawed and inadequate[1]. So, probably a more rigorous approach is required. However, as we discuss in Paper III[2], it may be that more information on the stations used than is currently archived needs to be compiled first.

Acknowledgements

No funding was received for this research.

We would like to thank Don Zieman for some useful discussions during the writing of this article. We would also like to thank several readers of version 0.1 who between them provided helpful feedback, spotted a few minor typos/mistakes and suggested some useful additional references.

References

[1] R. Connolly and M. Connolly. "Urbanization bias II. An assessment of the NASA GISS urbanization adjustment method". 31 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/31.

[2] R. Connolly and M. Connolly. "Urbanization bias III. Estimating the extent of bias in the Historical Climatology Network datasets". 34 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/34.

[3] J. Hansen et al. "Global surface temperature change". *Rev. Geophys.* 48 (2010), RG4004. DOI: 10.1029/2010RG000345.

[4] P. D. Jones et al. "Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010". *J. Geophys. Res.* 117 (2012), p. D05127. DOI: 10.1029/2011JD017139.

[5] T. M. Smith et al. "Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006)". *J. Clim.* 21 (2008), pp. 2283–2296. DOI: 10.1175/2007JCLI2100.1.

[6] K. M. Lugina et al. "Monthly surface air temperature time series area-averaged over the 30-degree latitudinal belts of the globe, 1881-2005". In: *Trends: A compendium of data on global change.* Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, U.S.A. DOI: 10.3334/CDIAC/cli.003.

[7] F. Fujibe and K. Ishihara. "Possible urban bias in gridded climate temperature data over the Japan area". *SOLA* 6 (2010), pp. 61–64. DOI: 10.2151/sola.2010-016.

[8] R. Rohde et al. "A new estimate of the average Earth surface land temperature spanning 1753 to 2011". *Geoinfor. Geostat.* 1 (2012). DOI: 10.4172/gigs.1000101.

[9] W.-H. Xu et al. "Overview of global monthly surface temperature data in the past century and preliminary integration". *Adv. Clim. Change Res.* 5 (2014), pp. 111–117. DOI: 10.1016/j.accre.2014.11.003.

[10] G. Mills. "Luke Howard and The Climate of London". *Weather* 63 (2008), pp. 153–157. DOI: 10.1002/wea.195.

[11] J. Hansen et al. "GISS analysis of surface temperature change". *J. Geophys. Res. D* 104 (1999), pp. 30997–31022. DOI: 10.1029/1999JD900835.

[12] J. Hansen et al. “A closer look at United States and global surface temperature change”. *J. Geophys. Res. D* 106 (2001), pp. 23947–23963. DOI: 10.1029/2001JD000354.

[13] J. Hansen and S. Lebedeff. “Global trends of measured surface air temperature”. *J. Geophys. Res. D* 92 (1987), pp. 13345–13372. DOI: 10.1029/JD092iD11p13345.

[14] T. M. L. Wigley and P. D. Jones. “Do large-area-average temperature series have an urban warming bias?” *Clim. Change* 12 (1988), pp. 313–319. DOI: 10.1007/BF00139435.

[15] P. D. Jones et al. “Assessment of urbanization effects in time series of surface air temperature over land”. *Nature* 347 (1990), pp. 169–172. DOI: 10.1038/347169a0.

[16] D. R. Easterling et al. “Maximum and minimum temperature trends for the globe”. *Science* 277 (1997), pp. 364–367. DOI: 10.1126/science.277.5324.364.

[17] T. C. Peterson et al. “Global rural temperature trends”. *Geophys. Res. Lett.* 26 (1999), pp. 329–332. DOI: 10.1029/1998GL900322.

[18] T. C. Peterson. “Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found”. *J. Clim.* 16 (2003), pp. 2941–2959. DOI: 10.1175/1520-0442(2003)016<2941:AOUVRI>2.0.CO;2.

[19] D. E. Parker. “Large-scale warming is not urban”. *Nature* 432 (2004), p. 290. DOI: 10.1038/432290a.

[20] D. E. Parker. “A demonstration that large-scale warming is not urban”. *J. Clim.* 19 (2006).

[21] D. A. Efthymiadis and P. D. Jones. “Assessment of maximum possible urbanization influences on land temperature data by comparison pf land and marine data around coasts”. *Atmosphere* 1 (2010), pp. 51–61. DOI: 10.3390/atmos1010051.

[22] C. Wickham et al. “Influence of urban heating on the global temperature land average using rural sites identified from MODIS classifications”. *Geoinfor. Geostat.* 1 (2013). DOI: 10.4172/gigs.1000104.

[23] T. R. Oke. “The energetic basis of the urban heat island”. *Quat. J. Roy. Meteor. Soc.* 108 (1982), pp. 1–24. DOI: 10.1002/qj.49710845502.

[24] T. R. Oke et al. “Simulation of surface urban heat islands under ‘ideal’ conditions at night: Part 2. Diagnosis of causation”. *Boundary-Layer Meteor.* 56 (1991), pp. 339–358. DOI: 10.1007/BF00119211.

[25] A. J. Arnfield. “Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island”. *Int. J. Clim.* 23 (2003), pp. 1–26. DOI: 10.1002/joc.859.

[26] T. R. Oke. “Initial guidance to obtain representative meteorological observations at urban sites”. I. O. M. Report 81. World Meteorological Organization, Geneva. 2006. URL: http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-81/IOM-81-UrbanMetObs.pdf.

[27] Y.-J. Yang et al. “Impacts of urbanization and station-relocation on surface air temperature series in Anhui Province, China”. *Pure Appl. Geophys.* 170 (2013), pp. 1969–1983. DOI: 10.1007/s00024-012-0619-9.

[28] G. Ren et al. “Urbanization effects on observed surface air temperature trends in North China”. *J. Clim.* 21 (2008), pp. 1333–1348. DOI: 10.1175/2007JCLI1348.1.

[29] J. M. Mitchell. “On the causes of instrumentally observed secular temperature trends”. *J. Meteor.* 10 (1953), pp. 244–261. DOI: 10.1175/1520-0469(1953)010<0244:OTCOIO>2.0.CO;2.

[30] D. R. Cayan and A. V. Douglas. “Urban influences on surface temperatures in the southwestern United States during recent decades”. *J. Clim. Appl. Meteor.* 23 (1984), pp. 1520–1530. DOI: 10.1175/1520-0450(1984)023<1520:UIOSTI>2.0.CO;2.

[31] G. Kukla, J. Gavin, and T. R. Karl. “Urban warming”. *J. Clim. Appl. Meteor.* 25 (1986), pp. 1265–1270. DOI: 10.1175/1520-0450(1986)025<1265:UW>2.0.CO;2.

[32] T. R. Karl, H. F. Diaz, and G. Kukla. “Urbanization: Its detection and effect in the United States climate record”. *J. Clim.* 1 (1988), pp. 1099–1123. DOI: 10.1175/1520-0442(1988)001<1099:UIDAEI>2.0.CO;2.

[33] F. B. Jr. Wood. “Comment: On the need for validation of the Jones et al. temperature trends with respect to urban warming”. *Clim. Change* 12 (1988), pp. 297–312. DOI: 10.1007/BF00139434.

[34] R. C. Jr. Balling and S. B. Idso. “Historical temperature trends in the United States and the effect of urban population growth”. *J. Geophys. Res.* 94D (1989), pp. 3359–3363.

[35] S. Solomon et al. *Climate change 2007: The physical science basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press. Cambridge. New York., 2007, 1056pp. URL: http://www.ipcc.ch/.

[36] G. C. Hegerl et al. “7. Understanding and attributing climate change”. In: *Climate change 2007: The physical science basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Ed. by S. Solomon et al. Cambridge University Press. Cambridge. New York., 2007, 1056pp. URL: http://www.ipcc.ch/.

[37] N. Stern. “The economics of climate change: The Stern review”. In: Cambridge University Press. Cambridge. New York., 2007, 712pp. ISBN: 9780521700801.

[38] I. D. Stewart. “A systematic review and scientific critique of methodology in modern urban heat island literature”. *Int. J. Clim.* 31 (2011), pp. 200–217. DOI: 10.1002/joc.2141.

[39] I. D. Stewart and T. R. Oke. “Local climate zones for urban temperature studies”. *Bull. Amer. Meteor. Soc.* 93 (2012), pp. 1879–1900. DOI: 10.1175/BAMS-D-11-00019.1.

[40] I. D. Stewart. “Influence of meteorological conditions on the intensity and form of the urban heat island effect in Regina”. *Can. Geograph.* 44 (2000), pp. 271–285. DOI: 10.1111/j.1541-0064.2000.tb00709.x.

[41] D. Kolokotsa, A. Psomas, and E. Karapidakis. “Urban heat island in southern Europe: The case study of Hania, Crete”. *Sol. Energy* 83 (2009), pp. 1871–1883. DOI: 10.1016/j.solener.2009.06.018.

[42] K. M. Hinkel and F. E. Nelson. “Anthropogenic heat island at Barrow, Alaska, during winter: 2001-2005”. *J. Geophys. Res.* 112 (2007), p. D06118. DOI: 10.1029/2006JD007837.

[43] J. A. Voogt and T. R. Oke. “Thermal remote sensing of urban climates”. *Rem. Sens. Environ.* 86 (2003), pp. 370–384. DOI: 10.1016/50034-4257(03)00079-8.

[44] P. Zhang et al. “Characterizing urban heat islands of global settlements using MODIS and nighttime light products”. *Can. J. Rem. Sens.* 36 (2010), pp. 185–196. DOI: 10.5589/m10-039.

[45] M. L. Imhoff et al. “Remote sensing of the urban heat island effect across biomes in the continental USA”. *Rem. Sens. Environ.* 114 (2010), pp. 504–513. DOI: 10.1016/j.rse.2009.10.008.

[46] S. Peng et al. “Surface Urban Heat Island across 419 global big cities”. *Environ. Sci. Technol.* 46 (2012), pp. 696–703. DOI: 10.1021/es2030438.

[47] D. R. Streutker. “Satellite-measure growth of the urban heat island of Houston, Texas”. *Rem. Sens. Environ.* 85 (2003), pp. 282–289. DOI: 10.1016/50034-4257(03)00007-5.

[48] Y. Goldreich. “Ground and top of canopy layer urban heat island partitioning on an airborne image”. *Rem. Sens. Environ.* 104 (2006), pp. 247–255. DOI: 10.1016/j.rse.2005.08.018.

[49] D. A. Hartz et al. “Linking satellite images and hand-held infrared thermography to observed neighborhood climate conditions”. *Rem. Sens. Environ.* 104 (2006), pp. 190–200. DOI: 10.1016/j.rse.2005.12.019.

[50] E. Jauregui. “Heat island development in Mexico City”. *Atm. Environ.* 31 (1997), pp. 3821–3831. DOI: 10.1016/S1352-2310(97)00136-2.

[51] D. Sharma and A. Bharat. “Urban heat island effect - Causes, impacts, methods of measurement and mitigation options”. *Inst. Town Planners India J.* 6 (2009), pp. 69–77. URL: http://itpi.org.in/content/apr_09.aspx.

[52] C. Rosenzweig et al. "Characterizing the urban heat island in current and future climates in New Jersey". *Env. Hazards* 6 (2005), pp. 51–62. DOI: 10.1016/j.hazards.2004.12.001.

[53] I. D. Stewart. "Landscape representation and the urban-rural dichotomy in empirical urban heat island literature". *Acta Clim. Chloro.* 40-41 (2007), pp. 111–121. URL: http://www2.sci.u-szeged.hu/eghajlattan/akta07/index.html.

[54] I. D. Stewart and T. R. Oke. "Newly developed "Thermal Climate Zones" for defining and measuring heat island magnitude in the canopy layer". Reprints, T. R. Oke Symposium and Eight Symposium on Urban Environment. January 11-15, 2009. Phoenix, AZ (U.S.A.) URL: http://ams.confex.com/ams/89annual/techprogram/paper_150476.htm.

[55] I. C. Nduka and A. I. Abdulhamed. "Classifying urban climate field sites by "Thermal Climate Zones" The case of Onitsha metropolis". *Res. J. Environ. Earth Sci.* 3 (2011), pp. 75–80. URL: http://maxwellsci.com/jp/abstract.php?jid=RJEES&no=89&abs=02.

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

[57] D. Parker, T. P. Legg, and C. K. Folland. "A new daily central England temperature series, 1772-1991". *Int. J. Clim.* 12 (1992), pp. 317–342. DOI: 10.1002/joc.3370120402.

[58] D. Parker and B. Horton. "Uncertainties in central England temperature 1878-2003 and some improvements to the maximum and minimum series". *Int. J. Clim.* 25 (2003), pp. 1173–1188. DOI: 10.1002/joc.1190.

[59] R. Böhm. "Urban bias in temperature time series - A case study for the city of Vienna, Austria". *Clim. Change* 38 (1998), pp. 113–128. DOI: 10.1023/A:1005338514333.

[60] S. R. Gaffin et al. "Variations in New York City's urban heat island strength over time and space". *Theor. Appl. Clim.* 94 (2008), pp. 1–11. DOI: 10.1007/s00704-007-0368-3.

[61] R. McKitrick. "Encompassing tests of socioeconomic signals in surface climate data". *Clim. Change* in press (2013). DOI: 10.1007/s10584-013-0793-5.

[62] T. Mohsin and W. A. Gough. "Trend analysis of long-term temperature time series in the Greater Toronto Area (GTA)". *Theor. Appl. Clim.* 101 (2010), pp. 311–327. DOI: 10.1007/s00704-009-0214-x.

[63] A. Moberg and H. Bergström. "Homogenization of Swedish temperature data. Part III: The long temperature records from Uppsala and Stockholm". *Int. J. Clim.* 17 (1997), pp. 667–699. DOI: 10.1002/(SICI)1097-0088(19970615)17:7<667::AID-JOC115>3.0.CO;2-J.

[64] J. D. Goodridge. "Comments on "Regional simulations of greenhouse warming including natural variability"". *Bull. Amer. Meteor. Soc.* 77 (1996), pp. 1588–1589. URL: http://www.scribd.com/doc/67524224/PDF-1996-Goodridge-Comments-on-Regional-Simulations-of-Greenhouse-Warming-Including-Natural-Variability.

[65] S. LaDochy, R. Medina, and W. Patzert. "Recent California climate variability: spatial and temporal patterns in temperature trends". *Clim. Res.* 33 (2007), pp. 159–169. DOI: 10.3354/cr033159.

[66] M. Roth. "Review of urban climate research in (sub)tropical regions". *Int. J. Clim.* 27 (2007), pp. 1859–1873. DOI: 10.1002/joc.1591.

[67] K. Kataoka et al. "Urban warming trends in several large Asian cities over the last 100 years". *Sci. Tot. Environ.* 407 (2009), pp. 3112–3119. DOI: 10.1016/j.scitotenv.2008.09.015.

[68] A. R. Saadatabadi and A. A. Bidokhti. "Urbanization effects on local climate in Tehran megapolis". *Res. J. Environ. Sci.* 5 (2011), pp. 1–21. DOI: 10.3923/rjes.2011.1.21.

[69] M. J. Coughlan, R. Tapp, and W. R. Kininmonth. "Trends in Australian temperature records". In: *Observed climate variations and change: Contributions in support of Section 7 of the 1990 IPCC Scientific Assessment.* Ed. by D. E. Parker. IPCC/WMO/UNEP, pp. 1–28.

[70] W. S. Hughes and R. C. Jr. Balling. "Urban influences on South African temperature trends". *Int. J. Clim.* 16 (1996), pp. 935–940. DOI: 10.1002/(SICI)1097-0088(199608)16:8<935::AID-JOC64>3.0.CO;2-V.

[71] R. Connolly and M. Connolly. "Has poor station quality biased U.S. temperature trend estimates?" 11 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/11.

[72] C. Idso and S. F. Singer. *Climate change reconsidered: 2009 report of the Nongovernmental Panel on Climate Change (NIPCC).* 2009. URL: http://nipccreport.org/.

[73] T. R. Karl and P. D. Jones. "Urban bias in area-averaged surface air temperature trends". *Bull. Amer. Meteor. Soc.* 70 (1989), pp. 265–270. DOI: 10.1175/1520-0477(1989)070<0265:UBIAAS>2.0.CO;2.

[74] D. L. Epperson et al. "Estimating the urban bias of surface shelter temperatures using upper-air and satellite data. Part II: Estimation of the urban bias". *J. Appl. Meteor.* 34 (1995), pp. 358–370. DOI: 10.1175/1520-0450-34.2.358.

[75] K. P. Gallo et al. "The use of NOAA AVHRR data for assessment of the urban heat island effect". *J. Appl. Meteor.* 32 (1993), pp. 899–908. DOI: 10.1175/1520-0450(1993)032<0899:TUONAD>2.0.CO;2.

[76] K. P. Gallo et al. "Temperature trends of the U.S. Historical Climatology Network based on satellite-designated land use/land cover". *J. Clim.* 12 (1999), pp. 1344–1348. DOI: 10.1175/1520-0442(1999)012<1344:TTOTUS>2.0.CO;2.

[77] J. R. Christy and J. D. Goodridge. "Precision global temperatures from satellites and urban warming effects of non-satellite data". *Atm. Environ.* 29 (1995), pp. 1957–1961. DOI: 10.1016/1352-2310(94)00240-L.

[78] S. A. Changnon. "A rare long record of deep soil temperatures defines temporal temperature changes and an urban heat island". *Clim. Change* 42 (1999), pp. 531–538. DOI: 10.1023/A:1005453217967.

[79] E. Kalnay and M. Cai. "Impact of urbanization and land-use change on climate". *Nature* 423 (2003), pp. 528–531. DOI: 10.1038/nature01675.

[80] B. Jr. Stone. "Urban and rural temperature trends in proximity to large US cities: 1951-2000". *Int. J. Clim.* 27 (2007), pp. 1801–1807. DOI: 10.1002/joc.1555.

[81] Z. Hausfather et al. "Assessing the urban heat island signal in the U.S. Historical Climatology Network monthly temperature data". Presentation to "19th Conference on Applied Climatology/36th Annual Meeting of the American Association of State Climatologists and Practical Solutions for a Warming World: AMS Conference on Climate Adaptations". 2011. URL: http://ams.confex.com/ams/19Applied/webprogram/Paper190465.html.

[82] United Nations. "World urbanization prospects: The 2009 revision". *Population Division* (2010). URL: http://esa.un.org/unpd/wup/index.htm.

[83] United Nations. "World population prospects: The 2010 revision". *Population Division* (2010). URL: http://esa.un.org/unpd/wpp/index.htm.

[84] W.-C. Wang, Z. Zeng, and T. R. Karl. "Urban heat islands in China". *Geophys. Res. Lett.* 17 (1990), pp. 2377–2380. DOI: 10.1029/GL017i013p02377.

[85] Y. Ren and G. Ren. "A remote-sensing method of selecting reference stations for urbanization effect on surface air temperature trends". *J. Clim.* 24 (2011), pp. 3179–3189. DOI: 10.1175/2010JCLI3658.1.

[86] Y. Li et al. "Urbanization impact on temperature change in China with emphasis on land cover change and human activity". *J. Clim.* 26 (2013), pp. 8765–8780. DOI: 10.1175/JCLI-D-12-00698.1.

[87] F. Wang and Q. S. Ge. "Estimation of urbanization bias in observed surface temperature change in China from 1980 to 2009 using satellite land-use data". *Chin. Sci. Bull.* 57 (2012), pp. 1708–1715. DOI: 10.1007/s11434-012-4999-0.

[88] X. Yang, Y. Hou, and B. Chen. "Observed surface warming induced by urbanization in east China". *J. Geophys. Res.* 116 (2011), p. D1413. DOI: 10.1029/2010JD015452.

[89] P. J. Englehart and A. V. Douglas. "Urbanization and seasonal temperature trends: Observational evidence from a data-sparse part of North America". *Int. J. Clim.* 23 (2003), pp. 1253–1263. DOI: 10.1002/joc.935.

[90] F. Fujibe. "Temperature rising trends at Japanese cities during the last hundred years and their relationships with population, population increasing rates and daily temperature ranges". *Pap. Meteor. Geophys.* 46 (1995), pp. 35–55. DOI: 10.2467/mripapers.46.35.

[91] F. Fujibe. "Detection of urban warming in recent temperature trends in Japan". *Int. J. Clim.* 29 (2009), pp. 1811–1822. DOI: 10.1002/joc.1822.

[92] F. Fujibe. "Urban warming in Japanese cities and its relation to climate change monitoring". *Int. J. Clim.* 31 (2011), pp. 162–173. DOI: 10.1002/joc.2142.

[93] E. M. Agee. "Present climatic cooling and a proposed causative mechanism". *Bull. Amer. Meteor. Soc.* 61 (1980), pp. 1356–1367. DOI: 10.1175/1520-0477(1980)061<1356:PCCAAP>2.0.CO;2.

[94] P. D. Jones et al. "The effect of urban warming on the Northern Hemisphere temperature average". *J. Clim.* 2 (1989), pp. 285–290. DOI: 10.1175/1520-0442(1989)002<0285:TEOUWO>2.0.CO;2.

[95] R. Spencer. *The global average urban heat island effect in 2000 estimated from station temperatures and population density data*. 2010. URL: http://www.drroyspencer.com/2010/03/the-global-average-urban-heat-island-effect-in-2000-estimated-from-station-temperatures-and-population-density-data/.

[96] Y.-K. Lim et al. "Observational evidence of sensitivity of surface climate changes to land types and urbanization". *Geophys. Res. Lett.* 32 (2005), p. L22712. DOI: 10-1029/2005GL024267.

[97] E. Kalnay et al. "Estimation of the impact of land-surface forcings on temperature trends in eastern United States". *J. Geophys. Res.* 111 (2006), p. D06106. DOI: 10.1029/2005JD006555.

[98] Y.-K. Lim et al. "Impact of vegetation types on surface temperature change". *J. Appl. Meteor. Clim.* 47 (2008), pp. 411–424. DOI: 10.1175/2007JAMC1494.1.

[99] R. McKitrick and P. J. Michaels. "A test of corrections for extraneous signals in gridded surface temperature data". *Clim. Res.* 26 (2004), pp. 159–173. DOI: 10.3354/cr026159. URL: http://www.uoguelph.ca/~rmckitri/research/MM04.Correction.pdf.

[100] R. McKitrick and P. J. Michaels. "Quantifying the influence of anthropogenic surface processes and inhomogeneities on gridded global climate data". *J. Geophys. Res.* 112 (2007), p. D24509. DOI: 10.1029/2007JD008465.

[101] R. McKitrick and N. Nierenberg. "Socioeconomic patterns in climate data". *J. Econ. Soc. Meas.* 35 (2011), pp. 149–175. DOI: 10.3233/JEM-2010-0336.

[102] A. T. J. de Laat and A. N. Maurellis. "Further evidence for influence of surface processes on lower tropospheric and surface temperature trends". *Int. J. Clim.* 26 (2006), pp. 897–913. DOI: 10.1002/joc.1292.

[103] K. E. Trenberth. "Rural land-use change and climate". *Nature* 427 (2004), p. 213. DOI: 10.1038/427213a.

[104] M. Cai and E. Kalnay. "Reply". *Nature* 427 (2004), p. 214. DOI: 10.1038/427214a.

[105] M. Cai and E. Kalnay. "Can reanalysis have anthropogenic climate trends without model forcing?" *J. Clim.* 18 (2005), pp. 1844–1849. DOI: 10.1175/JCLI3347.1.

[106] R. S. Vose et al. "Impact of land-use change on climate". *Nature* 427 (2004), pp. 213–214. DOI: 10.1038/427213b.

[107] A. J. Simmons et al. "Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperatures". *J. Geophys. Res.* 109 (2004), p. D24115. DOI: 10.1029/2004JD005306.

[108] R. E. Benestad. "Are temperature trends affected by economic activity? Comment on McKitrick & Michaels (2004)". *Clim. Res.* 27 (2004), pp. 171–173. DOI: 10.3354/cr027171.

[109] K. E. Trenberth et al. "3. Observations: Surface and atmospheric climate change". In: *Climate change 2007: The physical science basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by S. Solomon et al. Cambridge University Press. Cambridge. New York., 2007, 1056pp. URL: http://www.ipcc.ch/.

[110] G. A. Schmidt. "Spurious correlations between recent warming and indices of local economic activity". *Int. J. Clim.* 29 (2009), pp. 2041–2048. DOI: 10.1002/joc.1831.

[111] R. McKitrick and P. J. Michaels. "Are temperature trends affected by economic activity? Reply to Benestad (2004)". *Clim. Res.* 27 (2004), pp. 175–176. DOI: 10.3354/cr027175.

[112] A. T. J. de Laat. "IPCC and attribution of 20th century lower tropospheric temperature changes". *Rejected by Int. J. Clim.* (2008). URL: http://www.knmi.nl/~laatdej/2008_IJOC.pdf.

[113] R. Rohde et al. "Berkeley Earth temperature averaging process". *Geoinfor. Geostat.* 1 (2013). DOI: 10.4172/gigs.1000103.

[114] U.S. Census Bureau. "United States Summary: 2010; Population and housing unit counts. CPH-2-1. Issued September 2012". *Census of population and housing, 2010* (2012). URL: http://www.census.gov/prod/www/decennial.html.

[115] J. C. Comiso et al. "Accelerated decline in the Arctic sea ice cover". *Geophys. Res. Lett.* 35 (2008), p. L01703. DOI: 10.1029/2007GL031972.

[116] J. R. Christy et al. "What do observational datasets say about modeled tropospheric temperature trends since 1979?" *Rem. Sens.* 2 (2010), pp. 2148–2169. DOI: 10.3390/rs2092148.

[117] P.W. Thorne et al. "Revisiting radiosonde upper air temperatures from 1958 to 2002". *J. Geophys. Res.* 110 (2005), p. D18105. DOI: 10.1029/2004JD005753.

[118] M. B. Dyurgerov and M. F. Meier. "Twentieth century climate change: Evidence from small glaciers". *Proc. Natl. Acad. Sci.* 97 (2000), pp. 1406–1411. DOI: 10.1073/pnas.97.4.1406.

[119] United Nations Environment Programme and World Glacier Monitoring Service. *Global glacier changes: facts and figures*. 2008. ISBN: 978-92-807-2898-9. URL: http://www.grid.unep.ch/glaciers/.

[120] I. Velcogna. "Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE". *Geophys. Res. Lett.* 36 (2009), p. L19503. DOI: 10.1029/2009GL040222.

[121] D. A. Robinson, K. F. Dewey, and R. R. Heim. "Global snow cover monitoring: an update". *Bull. Amer. Meteor. Soc.* 74 (1993), pp. 1689–1696. DOI: 10.1175/1520-0477(1993)074<1689:GSCMAU>2.0.CO;2.

[122] V. E. Romanovsky, S. L. Smith, and H. H. Christiansen. "Permafrost thermal state in the Polar Northern Hemisphere during the International Polar Year 2007-2009: a synthesis". *Permafrost Periglac. Process* 21 (2010), pp. 106–116. DOI: 10.1002/ppp.689.

[123] M. Palmer and Co-Authors. "Future observations for monitoring global ocean heat content". In: *Proceedings of Ocean Obs'09: Sustained ocaen observations and information for society. Venice, Italy, 21-25 Spetember 2009*. Vol. 2. ESA Publication WPP-306, 2010. DOI: 10.5270/OceanObs09.cwp.68.

[124] F. L. Korsmo. "The genesis of the International Geophysical Year". *Phys. Today* 60 (2007), pp. 38–43. DOI: 10.1063/1.2761801.

[125] H. W. Linderholm. "Growing season changes in the last century". *Ag. Forest Meteor.* 137 (2006), pp. 1–14. DOI: 10.1016/j.agrformet.2006.03.006.

[126] A. Menzel et al. "European phenological response to climate change matches the warming pattern". *Glob. Change Biol.* 12 (2006), pp. 1969–1976. DOI: 10.1111/j.1365-2486.2006.01193.x.

[127] M. V. Kozlov and N. G. Berlina. "Decline in length of the summer season on the Kola Peninsula, Russia". *Clim. Change* 54 (2002), pp. 387–398. DOI: 10.1023/A:1016175101383.

[128] A. J. Miller-Rushing, D. W. Inouye, and R. B. Primack. "How well do first flowering dates measure plant responses to climate change? The effects of population size and sampling frequency". *J. Ecol.* 96 (2008), pp. 1289–1296. DOI: 10.1111/J.1365-2745.2008.01436.X.

[129] J.-H. Jeong et al. "Impact of urban warming on earlier spring flowering in Korea". *Int. J. Climat.* 31 (2011), pp. 1488–1497. DOI: 10.1002/joc.2178.

[130] M. Huss, M. Funk, and A. Ohmura. "Strong Alpine glacier melt in the 1940s due to enhanced solar radiation". *Geophys. Res. Lett.* 36 (2009), p. L23501. DOI: 10.1029/2009GL040789.

[131] P. Brohan et al. "Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850". *J. Geophys. Res.* 111 (2006), p. D12106. DOI: 10.1029/2005JD006548.

[132] T. P. Barnett. "Long-term trends in surface temperature over the oceans". *Mon. Wea. Rev.* 112 (1984), pp. 303–312. DOI: 10.1175/1520-0493(1984)112<0303:LTTIST>2.0.CO;2.

[133] P. D. Jones and T. M. L. Wigley. "Estimation of global temperature trends: what's important and what isn't". *Clim. Change* 100 (2010), pp. 59–69. DOI: 10.1007/s10584-010-9836-3.

[134] A. Bernaerts. "How useful are Atlantic sea-surface temperature taken during World War II". Proceedings Oceanology International 98, Brighton, UK, 1013 March 1998, Oceanology International. 1998. URL: http://www.warchangesclimate.com/English/Atlantic_SST_1998.pdf.

[135] J. B. R. Matthews. "Comparing historical and modern methods of Sea Surface Temperature measurement - Part 1: Review of methods, field comparisons and dataset adjustments". *Ocean Sci. Discuss.* 9 (2012), pp. 2951–2974. DOI: 10.5194/osd-9-2951-2012.

[136] J. B. R. Matthews and J. B. Matthews. "Comparing historical and modern methods of Sea Surface Temperature measurement - Part 2: Field comparison in the Central Tropical Pacific". *Ocean Sci. Discuss.* 9 (2012), pp. 2975–3019. DOI: 10.5194/osd-9-2975-2012.

[137] C. K. Folland and D. E. Parker. "Correction of instrumental biases in historical sea surface temperature data". *Quart. J. Roy. Meteor. Soc. B* 121 (1995), pp. 319–367. DOI: 10.1002/qj.49712152206.

[138] J. J. Kennedy et al. "Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization". *J. Geophys. Res.* 116 (2011), p. D14104. DOI: 10.1029/2010JD015220.

[139] J. Oerlemans. "Extracting a climate signal from 169 glacier records". *Science* 308 (2005), pp. 675–677. DOI: 10.1126/science.1107046.

[140] J. M. Grove. "The initiation of the "Little Ice Age" in regions round the North Atlantic". *Clim. Change* 48 (2001), pp. 53–82. DOI: 10.1023/A:1005662822136.

[141] M. Hulme. "Claiming and adjudicating on Mt Kilimanjaro's shrinking glaciers: Guy Callendar, Al Gore and extended peer communities". *Sci. Culture* 19 (2010), pp. 303–326. DOI: 10.1080/09505430903214427.

[142] G. H. Roe and M. A. O'Neal. "The response of glaciers to intrinsic climate variability: observations and models of late-Holocene variations in the Pacific Northwest". *J. Glaciol.* 55 (2009), pp. 839–854. DOI: 10.3189/002214309790152438.

[143] S. J. Holgate. "On the decadal rates of sea level change during the twentieth century". *Geophys. Res. Lett.* 34 (2007), p. L01602. DOI: 10.1029/2006GL028492.

[144] J. A. Church and N. J. White. "A 20th century acceleration in global sea-level rise". *Geophys. Res. Lett.* 33 (2006), p. L01602. DOI: 10.1029/2005GL024826.

[145] B. F. Chao, Y. H. Wu, and Y. S. Li. "Impact of artificial reservoir water impoundment on global sea level". *Science* 320 (2008), pp. 212–214. DOI: 10.1126/science.1154580.

[146] Y. Wada et al. "Global depletion of groundwater resources". *Geophys. Res. Lett.* 37 (2010), p. L20402. DOI: 10.1029/2010GL044571.

[147] R. W. Stewart. "Sea-level rise or coastal subsidence?" *Atm. Ocean* 27 (1989), pp. 461–477. DOI: 10.1080/07055900.1989.9649347.

[148] J. P. M. Syvitski et al. "Sinking deltas due to human activities". *Nature Geosci.* 2 (2009), pp. 681–686. DOI: 10.1038/ngeo629.

[149] É. Ostanciaux et al. "Present-day trends of vertical ground motion along the coast lines". *Earth Sci. Rev.* 110 (2012), pp. 74–92. DOI: 10.1016/j.earscirev.2011.10.004.

[150] N.-A. Mörner. "Estimating future sea level changes from past records". *Glob. Planet. Change* 40 (2004), pp. 49–54. DOI: 10.1016/S0921-8181(03)00097-3.

[151] R. S. Nerem et al. "Comment on "Estimating future sea level changes from past records" by Nils-Axel Mörner". *Glob. Planet. Change* 55 (2007), pp. 358–360. DOI: 10.1016/j.gloplacha.2006.08.002.

[152] N.-A. Mörner. "Comment on comment by Nerem et al. (2007) on "Estimating future sea level changes from past records" by Nils-Axel Mörner (2004)". *Glob. Planet. Change* 62 (2008), pp. 219–220. DOI: 10.1016/j.gloplacha.2008.03.002.

[153] M. E. Mann et al. "Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia". *Proc. Nat. Acad. Sci.* 105 (2008), pp. 13252–13257. DOI: 10.1073/pnas.0805721105.

[154] F. C. Ljungqvist. "A new reconstruction of temperature variability in the extra-tropical northern hemisphere during the last two millennia". *Geograf. Ann. A* 92 (2010), pp. 339–351. DOI: 10.1111/j.1468-0459.2010.00399.

[155] R. Connolly and M. Connolly. "Global temperature changes of the last millennium". 16 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/16.

[156] R. Wilson et al. "A matter of divergence: Tracking recent warming at hemispheric scales using tree ring data". *J. Geophys. Res.* 112 (2007), p. D17103. DOI: 10.1029/2006JD008318.

[157] D. E. Parker. "Urban heat island effects on estimates of observed climate change". *Wiley Inter. Rev. Clim. Change* 1 (2010), pp. 123–133. DOI: 10.1002/wcc.21.

[158] D. Pearlmutter, P. Berliner, and E. Shaviv. "Urban climatology in arid regions: current research in the Negev desert". *Int. J. Clim.* 27 (2007), pp. 1875–1885. DOI: 10.1002/joc.1523.

[159] A. M. Rizwan, L. Y. C. Dennis, and C. Liu. "A review on the generation, determination and mitigation of Urban Heat Island". *J. Environ. Sci.* 20 (2008), pp. 120–128. DOI: 10.1016/S1001-0742(08)60019-4.

[160] S. De Schiller and J. M. Evans. "Training architects and planners to design with urban microclimates". *Atm. Environ.* 30 (1996), pp. 449–454. DOI: 10.1016/1352-2310(94)00139-1.

[161] W. D. Solecki et al. "Mitigation of the heat island effect in urban New Jersey". *Environ. Haz.* 6 (2005), pp. 39–49. DOI: 10.1016/j.hazards.2004.12.002.

[162] M. L. Imhoff et al. "A technique for using composite DMSP/OLS 'City Lights' satellite data to map urban area". *Rem. Sens. Environ.* 61 (1997), pp. 361–370. DOI: 10.1016/S0034-4257(97)00046-1.

[163] A. Schneider, M. A. Friedl, and D. Potere. "A new map of global urban extent from MODIS satellite data". *Environ. Res. Lett.* 4 (2009), p. 044003. DOI: 10.1088/1748-9326/4/4/044003.

[164] Z. Hausfather et al. "Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records". *J. Geophys. Res.* in press (2013). DOI: 10.1029/2012JD018509.

[165] P. Shabecoff. "Global warming has begun, expert tells Senate". New York Times. 24 June, 1988. Accessed: 2012-01-20. (Archived by WebCite at http://www.webcitation.org/64pxkL5d0). URL: http://www.nytimes.com/1988/06/24/us/global-warming-has-begun-expert-tells-senate.html?pagewanted=all&src=pm.

[166] P. D. Jones et al. "Northern Hemisphere surface air temperature variations: 1851-1984". *J. Clim. Appl. Meteor.* 25 (1986), pp. 161–179. DOI: 10.1175/1520-0450(1986)025<0161:NHSATV>2.0.CO;2.

[167] P. D. Jones, T. M. L. Wigley, and P. B. Wright. "Global temperature variations between 1861 and 1984". *Nature* 322 (1986), pp. 430–434. DOI: 10.1038/322430a0.

[168] P. D. Jones, S. C. B. Raper, and T. M. L. Wigley. "Southern Hemisphere surface air temperature variations: 1851-1984". *J.*

Clim. Appl. Meteor. 25 (1986), pp. 1213–1230. DOI: 10.1175/1520-0450(1986)025<1213:SHSATV>2.0.CO;2.

[169] P. D. Jones et al. "A grid point surface air temperature data set for the Northern Hemisphere". In: *Technical Report #022 (TR022)*. U. S. Department of Energy, Office of Energy Research, Office of Basic Energy Sciences, Carbon Dioxide Research Division. Washington, D.C. 20545. URL: http://www.cru.uea.ac.uk/st/.

[170] P. D. Jones et al. "A grid point surface air temperature data set for the Southern Hemisphere, 1851-1984". In: *Technical Report #027 (TR027)*. U. S. Department of Energy, Office of Energy Research, Office of Basic Energy Sciences, Carbon Dioxide Research Division. Washington, D.C. 20545. URL: http://www.cru.uea.ac.uk/st/.

[171] B. Sun et al. "A comparative study of ASOS and USCRN temperature measurements". *J. Atmos. Oceanic Tech.* 22 (2005), pp. 679–686. DOI: 10.1175/JTECH1752.1.

[172] C. E. Duchon. "Temperature trends at San Juan, Puerto Rico". *Bull. Amer. Meteor. Soc.* 67 (1986), pp. 1370–1377. DOI: 10.1175/1520-0477(1986)067<1370:TTASJP>2.0.CO;2.

[173] M. A. Lokoshchenko and A. A. Isaev. "Influence of Moscow City on the air temperature in central Russia". Poster presentation P.6.13 to "5th International Conference on Urban Climate". 1-5th September, 2003. Łódź, Poland. 2003. URL: http://nargeo.geo.uni.lodz.pl/~icuc5/text/P_6_13.pdf.

[174] P. D. Jones and A. Moberg. "Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001". *J. Clim.* 16 (2003), pp. 206–223. DOI: 10.1175/1520-0442(2003)016<0206:HALSSA>2.0.CO;2.

[175] A. Robock. "Russian surface temperature data set". *J. Appl. Meteor.* 21 (1982), pp. 1781–1785. DOI: 10.1175/1520-0450(1982)021<1781:TRSTDS>2.0.CO;2.

[176] K. Ya. Vinnikov, P. Ya. Groisman, and K. M. Lugina. "Empirical data on contemporary global climate changes (temperature and precipitation)". *J. Clim.* 3 (1990), pp. 662–677. DOI: 10.1175/1520-0442(1990)003<0662:EDOCGC>2.0.CO;2.

[177] Q. Li et al. "Urban heat island effect on annual mean temperature during the last 50 years in China". *Theor. Appl. Clim.* 79 (2004), pp. 165–174. DOI: 10.1007/s00704-004-0065-4.

[178] Q. Li et al. "Assessment of the uncertainties in temperature change in China during the last century". *Chi. Sci. Bull.* 55 (2010), pp. 1974–1982. DOI: 10.1007/s11434-010-3209-1.

[179] W. Hughes and R. Jr. Balling. "Eastern Australia temperature variations 1930-1992". *Submitted to J. Clim., but rejected* (1994). URL: http://www.warwickhughes.com/papers/eastoz.htm.

[180] L. Kamél. "A homogeneous temperature record for southern Siberia". *Submitted to Geophys. Res. Lett., but rejected* (2004). URL: http://climateaudit.org/2009/12/21/climategatekeeping-siberia/.

[181] T. R. Karl et al. "A new perspective on recent global warming: Asymmetric trends of daily maximum and minimum temperature". *Bull. Amer. Meteor. Soc.* 74 (1993), pp. 1007–1023. DOI: 10.1175/1520-0477(1993)074<1007:ANPORG>2.0.CO;2.

[182] K. Gallo, D. R. Easterling, and T. C. Peterson. "The influence of land use/land cover on climatological values of the diurnal temperature range". *J. Clim.* 9 (1996), pp. 2941–2944. DOI: 10.1175/1520-0442(1996)009<2941:TIOLUC>2.0.CO;2.

[183] C.-J. Shiu, S. C. Liu, and J.-P. Chen. "Diurnally asymmetric trends of temperature, humidity, and precipitation in Taiwan". *J. Clim.* 22 (2009), pp. 5635–5649. DOI: 10.1175/2009JCLI2514.1.

[184] K. Trusilova et al. "Urbanization impacts on the climate in Europe: Numerical experiments by the PSU-NCAR Mesoscale Model (MM5)". *J. Appl. Meteor. Clim.* 47 (2008), pp. 1442–1455. DOI: 10.1175/2007JAMC1624.1.

[185] L. Zhang et al. "Effect of data homogenization on estimate of temperature trend: a case of Huairou station in Beijing Municipality". *Theor. Appl. Climatol.* 115 (2014), pp. 365–373. DOI: 10.1007/s00704-013-0894-0.

[186] A. Dai, K. E. Trenberth, and T. R. Karl. "Effects of clouds, soil moisture, precipitation, and water vapor on diurnal temperature range". *J. Clim.* 12 (1999), pp. 2451–2473. DOI: 10.1175/1520-0442(1999)012<2451:EOCSMP>2.0.CO;2.

[187] R. G. Quayle et al. "Effects of recent thermometer changes in the Cooperative Station Network". *Bull. Amer. Meteor. Soc.* 72 (1991), pp. 1718–1723. DOI: 10.1175/1520-0477(1991)072<1718:EORTCI>2.0.CO;2.

[188] D. E. Parker. "Maximum and minimum temperatures: a backward and a forward look". *Atm. Res.* 37 (1995), pp. 3–9. DOI: 10.1016/0169-8095(94)00062-I.

[189] S. Fall et al. "Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends". *J. Geophys. Res.* 116 (2011), p. D14120. DOI: 10.1029/2010JD015146.

[190] R. S. Vose, D. R. Easterling, and B. Gleason. "Maximum and minimum temperature trends for the globe: An update through 2004". *Geophys. Res. Lett.* 32 (2005), p. L23822. DOI: 10.1029/2005GL024379.

[191] K. Makowski, M. Wild, and A. Ohmura. "Diurnal temperature range over Europe between 1950 and 2005". *Atmos. Chem. Phys.* 8 (2008), pp. 6483–6498. DOI: 10.5194/acp-8-6483-2008.

[192] L. Brito-Castilo, S. C. Díaz Castro, and R. S. Ulloa Herrera. "Observed tendencies in maximum and minimum temperatures in Zacatecas, Mexico and possible causes". *Int. J. Clim.* 29 (2009), pp. 211–221. DOI: 10.1002/joc.1733.

[193] V. Gray. "Regional temperature change". *Submitted to Geophys. Res. Lett., but rejected* (1999). URL: http://www.john-daly.com/guests/regional.htm.

[194] M. J. Menne, C. N. Jr. Williams, and R. S. Vose. "The U.S. Historical Climatology Network monthly temperature data, version 2". *Bull. Amer. Meteor. Soc.* 90 (2009), pp. 993–1007. DOI: 10.1175/2008BAMS2613.1.

[195] R. A. Spronken-Smith and T. R. Oke. "The thermal regime of urban parks in two cities with different summer climates". *Int. J. Rem. Sens.* 19 (1998), pp. 2085–2104. DOI: 10.1080/014311698214884.

[196] A. Robock. "Volcanic eruptions and climate". *Rev. Geophys.* 38 (2000). See also correction published in 2007 (doi:10.1029/2007RG000232), pp. 191–219. DOI: 10.1029/1998RG000054.

[197] S. McIntyre. "Trends in Peterson 2003". Climate Audit blog. 4 August, 2007. URL: http://climateaudit.org/2007/08/04/1859/.

[198] K. P. Gallo. "Evaluation of temperature differences for paired stations of the U. S. Climate Reference Network". *J. Clim.* 18 (2005), pp. 1629–1636. DOI: 10.1175/JCLI3358.1.

[199] T. C. Peterson and T. W. Owen. "Urban heat island assessment: Metadata are important". *J. Clim.* 18 (2005), pp. 2637–2646. DOI: 10.17775/JCLI3431.1.

[200] G. T. Johnson et al. "Simulation of surface urban heat islands under 'ideal' conditions at night: Part 1. Theory and tests against field data". *Boundary-Layer Meteor.* 56 (1991), pp. 275–294. DOI: 10.1007/BF00120424.

[201] C. J. G. Morris, I. Simmonds, and N. Plummer. "Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city". *J. Appl. Meteor.* 40 (2001), pp. 169–182. DOI: 10.1175/1520-0450(2001)040<0169:QOTIOW>2.0.CO;2.

[202] I. Eliasson and B. Holmer. "Urban heat island circulation in Göteborg, Sweden". *Theor. Appl. Clim.* 42 (1990), pp. 187–196. DOI: 10.1007/BF00866874.

[203] R. C. Jr. Balling and R. S. Cerveny. "Long-term associations between winds speeds and the urban heat island of Phoenix, Arizona". *J. Clim. Appl. Meteor.* 26 (1987), pp. 712–716. DOI: 10.1175/1520-0450(1987)026<0712:LTABWS>2.0.CO;2.

[204] R. D. Bornstein. "Observations of the urban heat island effect in New York City". *J. Appl. Meteor.* 7 (1968), pp. 575–582. DOI: 10.1175/1520-0450(1968)007<0575:OOTUHI>2.0.CO;2.

[205] T. R. Oke. "The distinction between canopy and boundary-layer urban heat islands". *Atmosphere* 14 (1976), pp. 268–277. DOI: 10.1080/00046973.1976.9648422.

[206] G. Farmer et al. *Documenting and explaining recent global-mean temperature changes. Final Report to the Natural Environment Research Council. Contract No. GR3/6565*. Climatic Research Unit, U.K., June 1989, 141pp. URL: http://www.cru.uea.ac.uk/cru/pubs/pdf/Farmer-1989-NERC.pdf.