Urbanization bias II. An assessment of the NASA GISS urbanization adjustment method

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

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

Abstract

NASA GISS are currently the only group calculating global temperature estimates that explicitly adjust their weather station data for urbanization biases. In this study, their urbanization adjustment procedure was considered.

A number of serious problems were found with their urbanization adjustments: 1.) The vast majority of their adjustments involved correcting for "urban cooling", whereas urbanization bias is predominantly a warming bias. 2.) The net effect of their adjustments on their global temperature estimates was unrealistically low, particularly for recent decades, when urbanization bias is expected to have increased. 3.) When a sample of highly urbanized stations was tested, the adjustments successfully removed warming bias for the 1895-1980 period, but left the 1980s-2000s period effectively unadjusted.

In an attempt to explain these unexpected problems, a critical assessment of their adjustment procedure was carried out. Several serious flaws in their procedure were identified, and recommendations to overcome these flaws were given.

Overall, NASA GISS' urbanization adjustments were found to be seriously flawed, unreliable and inadequate. Until their adjustment approach is substantially improved, their global temperature estimates should be treated with considerable caution.

Citation:
R. Connolly, and M. Connolly (2014). *Urbanization bias II. An assessment of the NASA GISS urbanization adjustment method*, Open Peer Rev. J., 31 (*Clim. Sci.*), ver. 0.1 (non peer reviewed draft).
URL: http://oprj.net/articles/climate-science/31

Version: 0.1 (non peer-reviewed)
First submitted: January 8, 2014.
This version submitted: January 31, 2014.

1 Introduction

This paper is the second of three companion papers in which we investigate the influence of urbanization bias on global temperature estimates constructed from weather station records. In Paper I, we reanalyse a number of studies which have concluded that this influence is small or negligible. We find a number of flaws with each of those studies, which make their conclusions invalid[1]. In Paper III, we assess the extent of urbanization bias in the main weather station dataset used for constructing the current global temperature estimates, i.e., the Global Historical Climatology Network. We find that urbanization bias is a systemic problem within that dataset, and that the extent of the problem has been seriously underestimated[2]. Only one of the groups currently estimating global temperature trends from weather station records explicitly attempts to adjust their data to account for urbanization bias - National Aeronautics and Space Administration's Goddard Institute for Space Studies, henceforth NASA GISS. In this paper, we assess in detail the urbanization adjustment method applied by NASA GISS.

It is well-known that urban areas tend to be warmer than equivalent rural areas (a phenomenon referred to as the "urban heat island"[3–6]). Since at least the 19th century, associated with dramatic world population growth[7], there has been a continuous increase in urban development. In recent decades, this urbanization appears to have been accelerating, particularly since the 1980s[8, 9]. As a result many of the weather stations, which may initially have been

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

located in relatively rural (or only moderately urbanized) locations, have been encroached by urban sprawl over the years.

If the urban heat island near a weather station increases, it introduces an artificial warming trend into the recorded temperatures, i.e., urbanization bias. This is a problem for global temperature estimates because, although urban areas still only cover 1% of the Earth's land surface area, about half of the weather stations used for constructing global temperature estimates are located in or near urban areas.

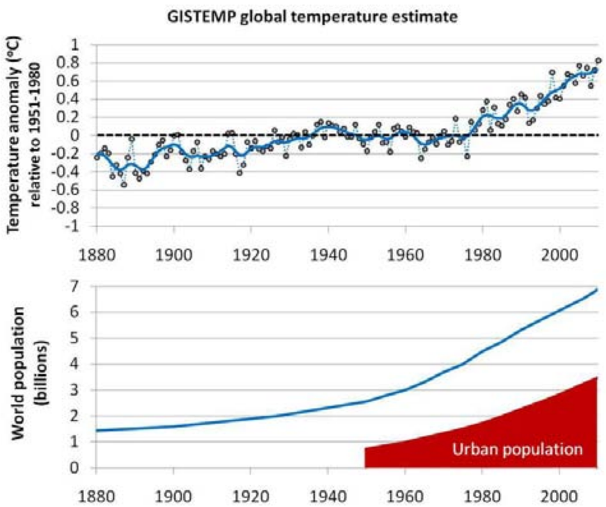


Figure 1: *Top: NASA GISS' global temperature estimates (land-only) expressed as deviations from the 1951-1980 mean - data downloaded from the NASA GISS website in November 2011. The solid blue lines corresponds to the 11-point binomial smoothed mean. Bottom: World population since 1880 (data from About.com website), and world population living in urban areas since 1950 (U.N. World Urbanization Population website)[8].*

As we discuss in Paper I, a number of studies have suggested that urban heat islands, while real and substantial, do not substantially affect the various global temperature estimates[1]. On the basis of these studies, most of the groups estimating global temperature trends from weather station records do not make any explicit attempt to remove the bias[10–14]. However, in Paper I, we show that the studies which had claimed the bias to be negligible were flawed[1].

NASA GISS is currently the only group which makes an explicit attempt to remove urbanization bias from their data before constructing their estimates[15–17]. The net effect of their urbanization adjustments on the trends of their global temperature estimates is quite small, and their estimate shows a similar amount of "global warming" to the estimates of the groups that ignore the urbanization problem, e.g., see Figure 3.1 of Ref. [18]. Initially, this might suggest that the effect of urbanization bias on global temperature estimates is only slight. However, from Figure 1, we can see a reasonable correlation still exists between NASA GISS' estimate and urban population growth. It is at least plausible that their urbanization adjustments were insufficient. In Papers I[1] and III[2], we show that urbanization bias is a substantial problem in current weather station-based global temperature estimates. So, it is surprising that NASA GISS calculate it to have such a small net effect.

With this in mind, it is worth carefully assessing the reliability of NASA GISS' urbanization adjustments. That is the purpose of this paper. In Section 2, we summarise the data used by NASA GISS, and the theory behind their adjustments. In Section 3 we describe a number of critical problems which are apparent from the results of their adjustments. We find that the adjustments applied by NASA GISS are inadequate and seem to have introduced about as many biases as they removed. We identify several serious flaws in their approach, which could explain these problems in Section 4. Finally, in Section 5, we offer some concluding remarks.

2 Theory behind the NASA GISS urbanization adjustments

Hansen et al., 1999 outlines the basic approach adopted by NASA GISS to remove urbanization bias from their weather station records[15]. They describe some later modifications to this approach in Hansen et al., 2001[16] and Hansen et al., 2010[17]. They also discuss other aspects of their global temperature estimates in Hansen et al., 2006[19].

The first step they take is to classify each station as either urban or rural. In their original 1999 version they did this by using estimates of populations in the vicinity of the stations[15]. However, currently, they use satellite-based estimates of the night light brightness associated with the co-ordinates of the stations[17]. Under both approaches, about half of the stations are identified as rural and half as urban.

NASA GISS explicitly assume that the only non-climatic biases they need to consider are those due to urbanization. They 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"*[15]. On this basis, they reason that the urbanization bias associated with a given urban station can be estimated (and then removed), by comparing the temperature trends of the urban station with the average trend of all the nearby rural stations.

To construct a rural average for an urban station, they require several neighbouring rural stations whose records at least partially overlap with that of the urban station. "Neighbouring" is initially defined as being within 500km of the urban station, but if that does not include enough rural stations, this is increased to 1000km. The contribution each neighbour makes to the rural average decreases as the distance from the urban station increases. If there are not at least three neighbouring, rural, stations with a common period[1] of at least 20 years with the urban station, then they are unable to adjust the urban station's record, and the station is not included in their global temperature estimates. Typically, between 5 and 7% of the urban stations are dropped in this way.

NASA GISS then estimates the urbanization bias at the urban station using a bi-linear approximation, comprising two segments, with each segment having a separate slope. The slopes of the two segments are determined by linear least squares fitting to the difference between the urban station and the rural average[2].

This adjustment approach may be better understood by considering the example adjustment shown in Figure 2. The unadjusted record (top panel) for the urban station at Sky Harbor International Airport in Phoenix, AZ (USA) shows a very strong warming trend since the start of its record. However, this strong warming is absent from the rural average constructed from its rural neighbours. NASA GISS define the difference between the urban record and the rural average as the *"urbanization bias"*, and therefore calculate their urbanization adjustment using the bi-linear fit of the difference (middle panel). This adjustment is then added to the unadjusted record yielding the adjusted record in the bottom panel.

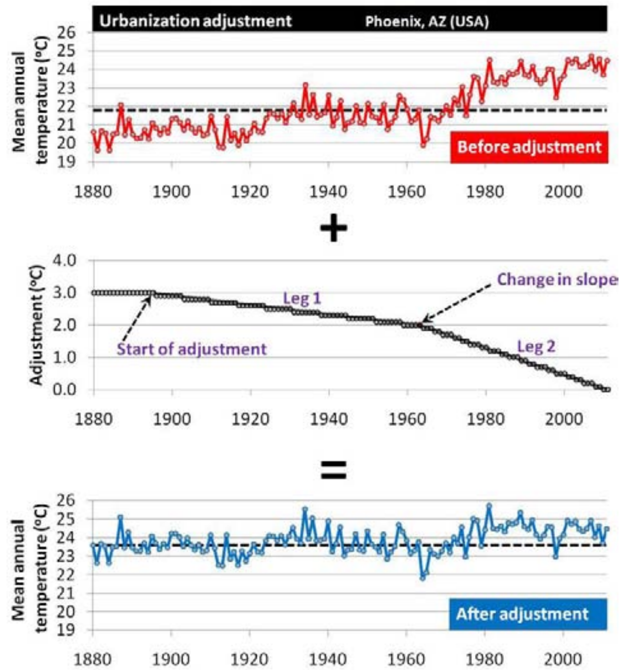


Figure 2: *Example of how NASA GISS adjust station records to account for urbanization bias. The values in the middle panel (black circles) are added to the red ("before adjustment") record in the top panel to yield the blue ("after adjustment") record in the bottom panel. The record shown is for the station at Sky Harbor International Airport in Phoenix, AZ (USA); 33.43° N, 112.02° W, GHCN ID=42572278000.*

An unusual feature of the NASA GISS adjustment algorithm is is that it applies the urbanization adjustments retrospectively inverted. In other words, instead of subtracting warming from the more recent part of the Phoenix record (in the above case), NASA GISS add warming to the earlier part. This is a counter-intuitive approach - if a weather station becomes warmer due to urbanization bias, NASA GISS treat the new warmer temperature as "normal" and increase the earlier temperatures to match.

NASA GISS' decision to take this approach appears somewhat arbitrary, and as we discuss in Section 3.5 has a number of problems. Nonetheless, when they use their station records for calculating their global temperature estimates, they first convert each record into an "anomaly record", by subtracting the 1951-1980 mean temperature for each station from all of the annual temperatures of that station's

[1] The "period" of a station's record is defined as the years between the first and last data points, even if there are large gaps in between.

[2] As the adjustments are rounded off to 0.1°C, these slopes are not exactly linear, but rather have a staggered staircase-like shape.

record. In other words, for their global analysis, the *absolute* temperatures of individual stations do not matter, just the temperatures relative to the 1951-1980 mean. So, in that sense, it should not make much difference whether the absolute temperatures of the adjusted records are all too high (or low), once they are all out by the same temperature.

The reason why NASA GISS use a bi-linear fit for their adjustments, rather than a simpler linear fit is to allow *"some time dependence in the rate of growth of the urban influence"*[15].

In the current version, the year separating the two linear segments (or "legs") is allowed to vary to give the best fit of the two legs to the rural average-urban difference. However, in the original 1999 version, this year was fixed as 1950[15].

2.1 Relevant data used and produced by NASA GISS

The global temperature estimate compiled by NASA GISS is the "GISS Surface Temperature Analysis", often referred to by its acronym:"GISTEMP". This estimate is updated monthly and is available from NASA GISS' website: http://data.giss.nasa.gov/gistemp/.

For their weather records, the main dataset used by NASA GISS is the Global Historical Climatology Network (GHCN)[20], which is provided by the NOAA National Climatic Data Center. However, NASA GISS also use an additional dataset to include more Antarctica stations. This dataset is provided by the Scientific Committee on Antarctic Research, or SCAR, and downloaded through the READER (REference Antarctic Data for Environmental Research) project at http://www.antarctica.ac.uk/met/READER/. Finally, they use an updated version of the record for the Hohenpeissenberg station in Germany. This version was taken from Hans Erren's website at http://members.casema.nl/errenwijlens/co2/europe.htm. Only stations with at least 20 years of data are considered in the NASA GISS analysis. For this reason, not all of the stations in the datasets mentioned are used.

The source code used by NASA GISS is available from their website at http://data.giss.nasa.gov/gistemp/sources/. After NASA GISS' release of their source code in 2007[21], the voluntary "Clear Climate Code Project" was co-founded by Nick Barnes. Barnes et al. ported the somewhat antiquated code (which was mostly written in Fortran90) into Python, using a more modern and accessible programming style. The results are available on the http://clearclimatecode.org/ website, and they appear to have produced results similar to the original GISTEMP code[22]. However, for our analysis, we used the original source code.

2.2 Time-line of NASA GISS' urbanization adjustments

NASA GISS have made a number of modifications to their approach in the years since Hansen et al., 1999 introduced their original approach[15]. Some of these modifications have been described in Hansen et al., 2001[16] and Hansen et al., 2010[17], and the rest are documented on the NASA GISS website. But, it may be helpful to briefly review these modifications.

- 1999

Hansen et al., 1999 introduced NASA GISS' original urbanization adjustment approach[15]. In this version, the year separating the two legs of the adjustments was fixed as 1950. Urban stations were identified on the basis of the population size associated with the station. Stations could be either "rural" (population $< 10,000$), "small town" ($10,000 \leq$ population $\leq 50,000$), or "urban" (population $> 50,000$). "Small town" stations were not adjusted for urbanization bias, but were not included in the rural averages either. All rural stations within 1000km of the urban station were included in the rural average.

- 2001

Hansen et al., 2001 updated their approach[16]. For the U.S. component of their dataset, they started using a homogeneity-adjusted version. Their method was changed to allow the year separating the two legs of the adjustment to vary. The maximum distance a rural station could be from the urban station was reduced from 1000km to 500km. However, if there were not at least three rural stations within 500km with an overlap of 2/3 of the record, the maximum distance for that station was increased back to 1000km. They also started adjusting "small town" stations as well as "urban" stations.

For stations in the United States (and nearby Canada and Mexico regions), they switched to using satellite night-light intensities to identify urban stations, instead of the population-based metric.

- July 2003

They started using a more complete record for the Hohenpeissenberg weather station than the one in their main dataset.

- September 2007

They published their code, and began providing public access to their monthly calculations[21].

- 2009

They switched to using a new version of the dataset they used for their U.S. component. This used a different set of homogeneity adjustments.

- 2010

The satellite night-lights metric used for identifying urban stations was expanded to apply to all stations, not just stations in the United States. We discuss the impacts this change had in Sections 4.2 and 4.3.

Hansen et al., 2010[17] summarised their updates since Hansen et al., 2001[16].

- December 2011

They switched to using a homogeneity-adjusted version of their main global dataset. They stopped using a separate dataset for their U.S. component, and adjusting the St. Helena and Lihue station records.

They also stopped publishing their intermediate monthly calculations, but switched to simply providing their finished products.

2.3 Our analysis

At several stages over the period August 2010 to November 2011, we downloaded the output files which NASA GISS generate every month, as intermediate calculations before constructing their global temperature estimates. We downloaded these output files from their public *ftp* website at ftp://data.giss.nasa.gov/pub/gistemp/GISS_Obs_analysis/. These output files provide details of the urbanization adjustments NASA GISS carry out. We analysed the format of these output files using the source code NASA GISS used for generating them (available from their website at http://data.giss.nasa.gov/gistemp/sources/). We then wrote a number of computer scripts using the Python programming language to systematically analyse the individual urbanization adjustments carried out for a given month. The scripts we used are included in the Supplementary Information, along with some sample input and output files.

NASA GISS changed their main dataset in December 2011. Unfortunately, as we mentioned in Section 2.2, they also stopped publishing their intermediate output files at this time. Indeed, at the time of writing, NASA GISS had removed their public *ftp* website. As a result, we were unable to use our detailed monthly analysis to study the effects of this change in dataset. However, in Paper III, we compare the dataset they used before December 2011 to the one they have been using since[2]. So, in Section 4.5.2, we are able to offer some discussion of the effects of this change in datasets.

3 Problems with NASA GISS' urbanization adjustments

In this section, we will summarise our main observations on the urbanization adjustments applied by NASA GISS before generating their global temperature estimates. We found several different problems with their adjustments, and we will describe each of them in turn. In Section 4, we will discuss flaws in the approaches NASA GISS uses for generating these adjustments, which may explain why these problems occur.

3.1 Most adjustments seem physically unrealistic

As described in Section 2, NASA GISS' adjustments comprise a bi-linear adjustment for each station identified as urban. The annual value of each station's adjustment for a given year is then subtracted from the raw monthly (and hence, annual) mean temperatures of that station for that year.

The slope of each of the two legs of the adjustments can be any value between -1 and 1. As the two slopes can be different, they can each be of either sign. A negative slope will reduce the amount of warming which occurred during a leg in the adjusted record, i.e., it will counteract an "urban warming" bias. A positive slope on the other hand will increase the amount of warming, i.e., it will counteract an "urban cooling" bias.

Therefore, we can categorise each of NASA GISS' adjustments into four types, based on the slopes of the two legs. We denote adjustments where both slopes are negative as Type 1, those where both slopes are positive (or zero) as Type 2. For adjustments where the two slopes are of opposite sign, we denote

Survey date	Type 1 Urban warming	Type 2 Urban cooling	Type 3 Warm/cool	Type 4 Cool/warm	Total	Rural	Skipped
Feb 2008†	447	300	1342	1440	3529	2488	250
Aug 2010	457	265	1191	1098	3011	3124	164
Jan 2011	459	260	1184	1108	3011	3127	176
Jul 2011	451	266	1182	1112	3011	3131	180
Oct 2011	461	261	1181	1113	3016	3132	177
Nov 2011	455	265	1177	1117	3014	3136	177
Slope 1	< 0	≥ 0	< 0	≥ 0			
Slope 2	< 0	≥ 0	≥ 0	< 0			

Table 1: *The four types of adjustments used in the NASA GISS analysis, and the frequency with which they were used at the time of each of our surveys. †Data for February 2008 survey was downloaded from the Climate Audit website. Positive slopes reduce the amount of warming, i.e., counteract "urban warming". Negative slopes increase the amount of warming, i.e., counteract "urban cooling". Types 3 and 4 are adjusted for both "urban warming and cooling".*

those where the first leg's slope is positive or zero, and the second leg's slope is negative as Type 3, and adjustments with the signs the other way around as Type 4. Examples of all four types are shown on the next few pages in Figures 3, 4, 5 and 6.

We carried out surveys of the NASA GISS urbanization adjustments at five times over the period from August 2010 until November 2011. In December 2011, NASA GISS stopped publishing the intermediate calculations we used for carrying out these surveys, and so our analysis stops then. Another researcher, McIntyre, carried out an analysis of the NASA GISS urbanization adjustments in March 2008 for his Climate Audit blog[23], and so we were also able to carry out a partial survey using his data from that analysis, which we downloaded from the Climate Audit website.

The frequencies of each of the types of adjustments made by NASA GISS during each of the surveys are listed in Table 1. Although there is some variability from survey to survey in the total numbers of each type, the frequencies of the different adjustments are fairly consistent. There is a relatively large change between the March 2008 survey and the others, but as we discussed in Section 2.1, there were a few significant changes in the NASA GISS approach between 2008 and August 2010. We will discuss the impact of one of these (the switch to night-light brightnesses as an urbanization metric) in Sections 4.2 and 4.3.

An unexpected result which can be seen from Table 1 is the relatively small number of adjustments which are of Type 1 (only about 12-15%). These are the adjustments which remove an urban warming bias. When urbanization bias is referred to in terms of temperature records, it is usually assumed to be of this type. However, the vast majority of NASA GISS' adjustments include "urban cooling" adjustments - either for the entire adjustment (Type 2) or else for half of the adjustment (Types 3 and 4). As we will discuss in Section 3.2, urbanization bias is predominantly a warming bias, so it is unrealistic that NASA GISS should identify such a high incidence of urban cooling biases.

Whether urbanization bias causes a warming or cooling bias at an individual station, it is difficult to see how urbanization at a station could cause a "warming bias" for several decades, but then spontaneously switch to causing a "cooling bias" (Type 3, e.g., Figure 5), or vice versa (Type 4, e.g., Figure 6). However, from Table 1, it can be seen that these two types of "tag-team" (or "bipolar"[23]) adjustments comprise most of NASA GISS' urbanization adjustments ($\sim$ 76-79%). This suggests that the "urbanization biases" NASA GISS has identified are not genuine urbanization biases.

3.2 Unusually high incidence of "urban cooling"

One of the most striking features of NASA GISS' urbanization adjustments is that about half of their adjustments are to counteract "urban cooling", i.e., their Type 2 adjustments, the second leg of their Type 3 adjustments and the first leg of their Type 4 adjustments. NASA GISS justify the inclusion of urban cooling adjustments with the following:

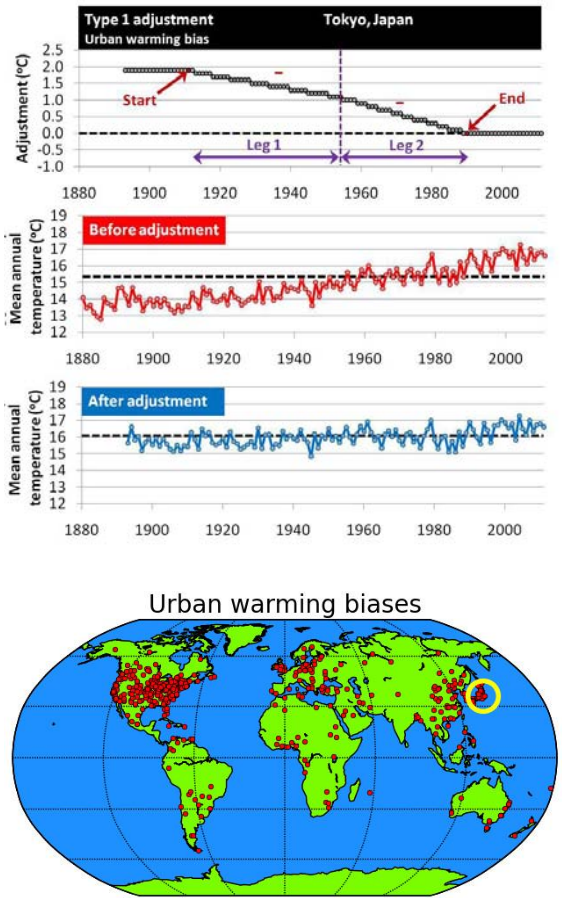


Figure 3: *Example of a Type 1 adjustment to remove "urban warming biases". The values in the top panel (black circles) are added to the red "before adjustment" record to yield the blue "after adjustment" record. The bottom panel shows the locations of the Type 1 adjustments from the November 2011 survey, with the example station (Tokyo, Japan) highlighted in yellow outline.*

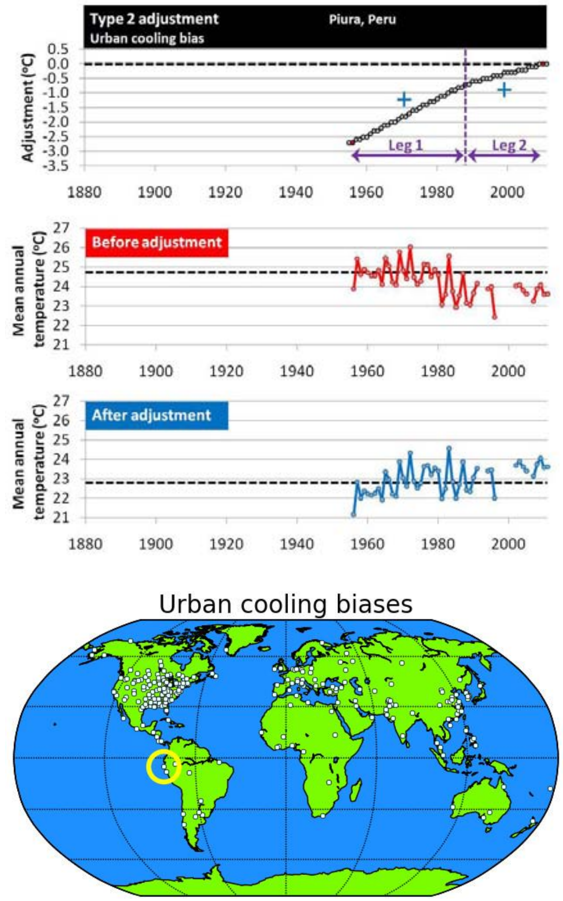


Figure 4: *As for Figure 3, except for Type 2 adjustments, which remove "urban cooling biases".*

> *"Anthropogenic effects can also cause a non-climatic cooling, for example, as a result of irrigation and planting of vegetation, but these effects are usually outweighed by urban warming."* - Hansen et al., 1999[15]

This seems a rather vague, and unsatisfactory explanation. More recently, Hansen et al., 2010 offered an alternative justification for their large number of urban cooling adjustments:

> *"A significant urban cooling can occur, for example, if a station is moved from central city to an airport and if the new station continues to be reported with the same station number and is not treated properly as a separate station in the global analysis."* - Hansen et al., 2010[17]

This is a worrying explanation for several reasons.

1. It misleadingly implies that NASA GISS actually make attempts to identify stations which have undergone station moves, and then treat such moved stations *"properly as a separate station in the global analysis"*, when they currently do not do this.

2. It implies that they consider a station move to be an "urbanization bias". This is inappropriate as the moving of a station does not have an influence on the development of neighbouring urban heat islands, so while it can introduce bias, it is not one of "urbanization". Indeed, it is a bias which is not limited to "urban" stations, but to

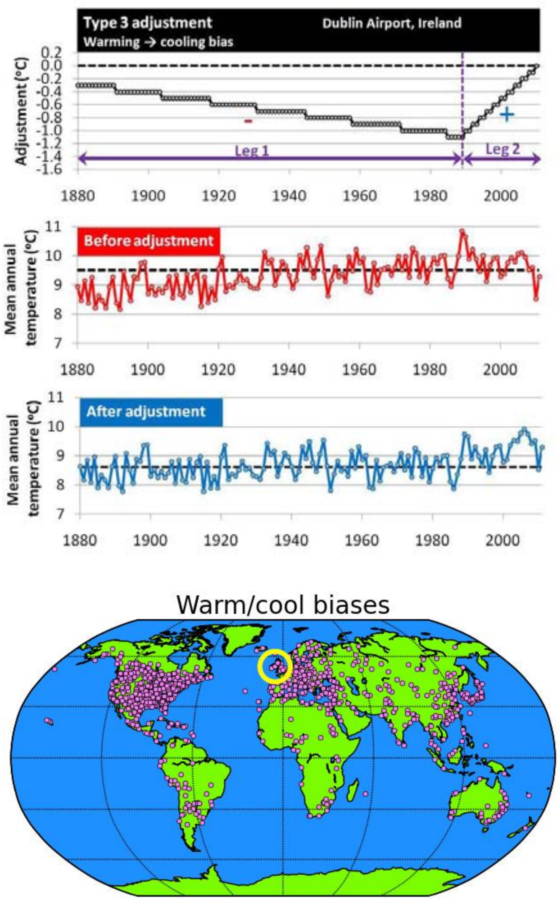


Figure 5: *As for Figure 3, except for Type 3 adjustments, which remove "urban warming biases" for the first part of the record, and "urban cooling biases" from the second part of the record.*

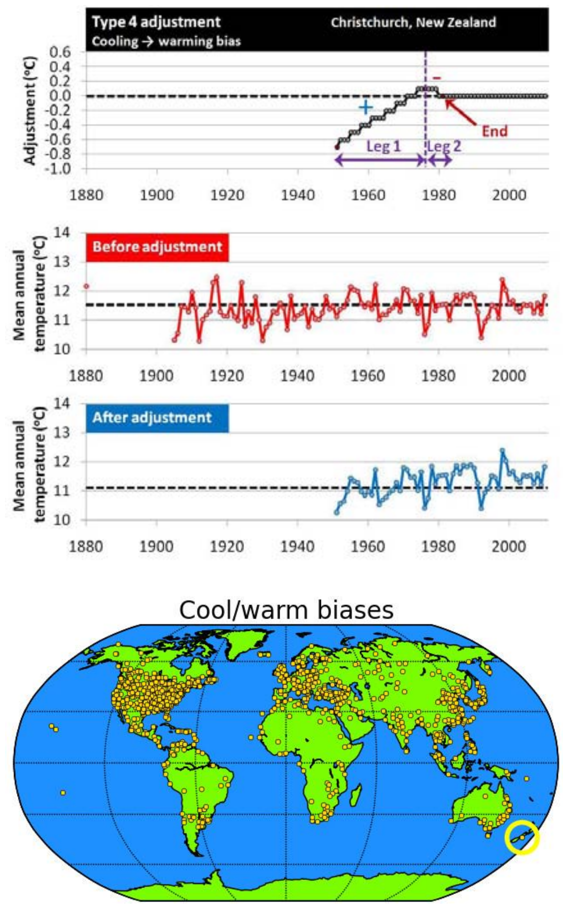


Figure 6: *As for Figure 3, except for Type 4 adjustments, which remove "urban cooling biases" for the first part of the record, and "urban warming biases" from the second part of the record.*

all stations which undergo station moves. Station moves are quite common for stations with multi-decadal records[24, 25].

3. It implies that they consider it acceptable to treat station move biases in the same manner as actual urbanization bias. Station moves are more likely to produce step biases rather than the trend biases which NASA GISS' bi-linear adjustments are designed for. Step and trend biases have different statistical properties, as we discuss in Section 4.5.1, and treating them as equivalent can increase the twin risks of failing to identify specific biases or misidentifying a bias where there is none, e.g., see DeGaetano, 2006[26].

From reviewing the extensive literature on urban climatology (e.g., see Arnfield, 2003[5], or our review in Paper I[1]), it seems highly unlikely that long-term "urban cooling" trends from the above scenarios or others are a dominant feature of the urban development which has occurred since the late 19th century.

It is true that some classes of urbanization can (under certain conditions) lead to either a reduction in local heat islands, or in some cases to "urban cooling". For instance, Tereshchenko & Filonov, 2001[27] found that during the wet season (June-July), a "cool island" developed in Guadalajara, Mexico (a large tropical, high elevation city). One could argue then that urbanization in this case led to cooling. However, when averaged over the entire year, the annual trend for that region was of urban warming. In dry, hot desert areas, urban features can sometimes lead to cooler daytime temperatures, but these only occur when certain conditions are met, and are also often

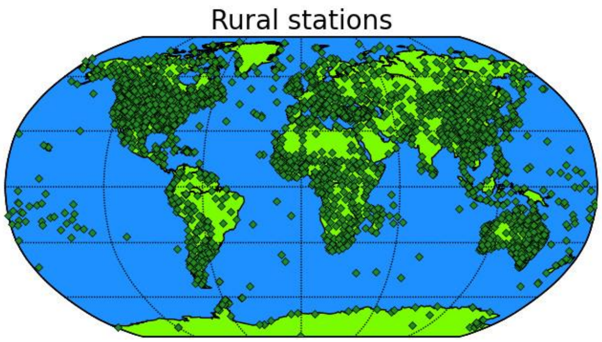


Figure 7: *Stations used by NASA GISS in November 2011 which were identified as rural.*

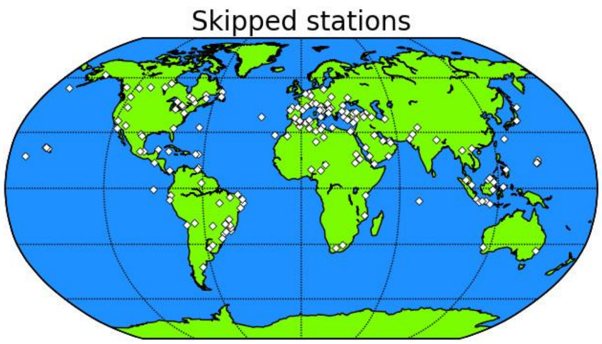


Figure 8: *Urban stations which NASA GISS dropped from its November 2011 analysis, because they did not have enough rural neighbours to construct a rural average.*

associated with warmer nighttime temperatures[28, 29].

Parks and green areas in cities are often cooler than the surrounding areas[30, 31] and are dubbed "park cool islands"[30]. But, this is generally thought to be a partial mitigation of the urban heat island phenomenon, rather than being an example of increasing urbanization in itself leading to cooling.

Some urban areas also appear to have started off cooler than neighbouring rural areas[32], perhaps due to the location of the urban area, but it is not the urban-rural difference itself which matters for global temperature estimates, but the *trends* of station records over time[33]. These in general seem to increase with increasing urbanization[32].

In recent years, there has been a lot of interest in modifying urban planning and development to deliberately counteract urban heat islands[27–31, 34–38] - for example, by the careful planning of urban vegetation[32, 37, 39–42], or by the use of high-albedo surfaces in urban areas to reflect sunlight away[32, 37, 42], e.g., light coloured roofs. But, this typically is an expensive, difficult, politically complex, and intentional process. The motivation for such urban plans is typically to counteract a problem in the area of expanding urban heat islands.

This all suggests that long-term "urban cooling" is unlikely to have been a frequent spontaneous feature of urban development for the stations being used by NASA GISS. In other words, if NASA GISS' urbanization adjustments are genuinely removing urbanization bias, then only a small fraction (at most) of their adjustments should be for "urban cooling". The fact that roughly half of their adjustments are for urban cooling, suggests their adjustment approach is unreliable.

3.3 Unrealistic net adjustments

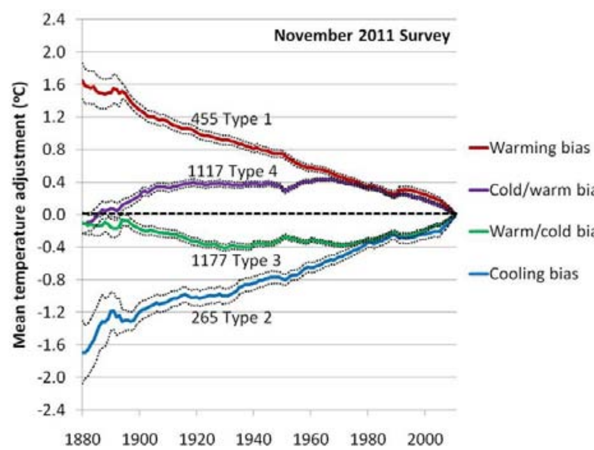


Figure 9: *Gridded mean for the adjustments of each of the four types applied by NASA GISS in the November 2011 survey. The dotted lines on either side of the four lines correspond to confidence intervals of twice the standard error of the annual mean adjustments.*

Figure 9 shows the mean adjustments applied by NASA GISS in November 2011, for each of the four different adjustment types. To construct these curves, we assigned all the stations to $5° \times 5°$ latitude by longitude grid boxes, for each of the four subsets of stations in Figures 3-6. We then calculated the mean adjustment applied in each grid box. Then, we weighted each box by the cosine of the latitude of the middle of each box, since higher latitude grid boxes have a smaller surface area. Finally, we calculated the mean adjustments of the weighted grid boxes to obtain the mean global adjustments.

We can see from Figure 9 that when the adjustments are sub-setted by adjustment type, the magnitude of the urbanization adjustments is quite substantial. For instance, the linear trend for the mean Type 1 adjustment is $-1.14°C$/century, while the linear trend for the mean Type 2 adjustment is $+1.04°C$/century.

The trend of NASA GISS' global temperature estimates (e.g., Figure 1) shows several non-linear aspects, and so it is inaccurate to describe the long-term trend using a linear fit. Nonetheless, if we *approximate* the trend as linear, the long-term trend of Figure 1 is about $+0.63°C$/century. In other words, the average magnitudes of NASA GISS' individual urbanization adjustments are comparable to (and often greater than) their estimates of "global warming".

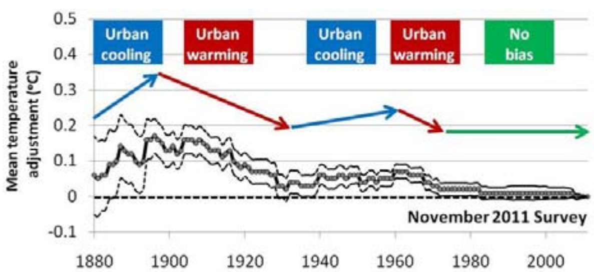


Figure 10: *Gridded mean of all of the adjustments applied by NASA GISS in the November 2011 survey. The dotted lines correspond to confidence intervals of twice the standard error of the annual mean adjustments.*

Figure 10 shows the gridded mean adjustments when calculated for all stations, i.e., the net effect on the gridded mean trends of the urban stations. These net adjustments are much smaller than the mean adjustments of the subsets. Again, the trend is not linear, it can be approximated by a linear trend of only about $-0.10°C$/century.

Since NASA GISS identifies about half of their stations as urban (see Table 1), the net effect on their global temperature estimates (e.g., the plot in Figure 1) is only about half of the net effect on the urban subset. That is, the net global effect of their urbanization bias adjustments is only about $-0.05°C$/century.

This explains why the NASA GISS global temperature estimates are so similar to the global temperature estimates of the groups that do not attempt urbanization adjustments (e.g., see Figure 3.1 of Ref [18]). As Hansen et al., 1999 noted when they developed the NASA GISS urbanization adjustments, the *net* effect of their adjustments on global trends is small[15]. However, from Figure 9, we can see that this is *not* because the individual adjustments are small, but rather because the mean adjustments of the different types are mostly cancelled by an approximately "mirror image" set of adjustments of the opposite sign, i.e., in the November 2011 survey, the 455 Type 1 adjustments are roughly balanced by the 265 Type 2 adjustments and the 1177 Type 3 adjustments are roughly balanced by the 1117 Type 4 adjustments.

Since the mean magnitudes of the adjustments are so large, it is important to confirm that they are reasonably accurate. We can see evidence that the adjustments are unreliable by considering the net adjustments of Figure 10 in detail. For instance, the adjustments for two periods (1880s-1890s and 1930s-1960s) are for urban cooling. As we discussed in Section 3.2, urbanization bias is mostly a problem of urban warming. So, even if some individual stations genuinely experienced some urban cooling, the fact that, when averaged globally, the NASA GISS adjustments have periods of *net* urban cooling seems physically unrealistic.

Another problem is that the slopes of the adjustments seem to have been getting closer to zero over time. We saw from Figure 1 that there has been a dramatic increase in population since 1880, and particularly for recent decades this growth has been greatest in urban areas[7–9]. So, regardless of the *sign* of urbanization bias, we would expect the *magnitude* of urbanization bias to have increased, not decreased, as the world has become more urbanized. In particular, the fact that NASA GISS' net urbanization adjustments are almost zero for the post-1970s period is a major problem.

All of these factors suggest that NASA GISS' urbanization adjustments are unreliable. In the next section, we will describe an additional test which confirms their unreliability.

3.4 Incomplete adjustments of highly urbanized stations

Identifying which station records are affected by urbanization bias is not a simple problem. As we discuss in Paper I[1], some stations that are located *near* an urban area might actually be far enough away to be unaffected, while some stations that are located in an area which is relatively rural (e.g., in a small town) may be affected if the station is located near enough to where the urbanization occurred. If most of the urbanization development occurred before the station was set up, the urbanization bias might not have changed much during the station record. But, in some regions, even a small amount of urban development can lead to a substantial warming bias, e.g., weather observers in climatically harsh areas, such as Arctic permafrost regions, may have substantially improved insulation and shelter in the areas near where they work, over the years.

Nonetheless, we would expect that urbanization bias should, in general, be relatively large at stations located in highly urbanized metropolises. So, if NASA GISS' urbanization adjustment method is at all reliable, their adjustments should be relatively large for these stations. With this in mind, we carried out a test on the results of the November 2011 survey, by calculating the mean trends and adjustments for a subset of highly urbanized stations.

Using metadata accompanying the station records[43], we identified the most highly urbanized stations in terms of both night-light brightness and associated population. We only selected those stations with an associated population of at least 2 million. Of those stations, we only kept those described as "bright" by Peterson et al., 1999[44] and with at least three times the brightness of Imhoff et al., 1997[45]'s urban threshold. Some of these stations are dropped from NASA GISS's final estimates as there are too few rural stations in the vicinity to meet their requirements, e.g., the São Paulo station in Brazil. For this reason, we removed these skipped stations from our subset.

116 stations met these criteria. Their locations (as well as the skipped stations) are shown in Figure 11. However, particularly for the U.S., where there is a relatively high station density, some of these stations were located in the same urban metropolises. So, many of these stations are too close to each other to be distinguishable on the map in Figure 11. In total, there were stations from a total of 47 different urban metropolises from around the world included in the subset. The list of stations in the subset is provided in the Supplementary Information.

The mean temperature trends of the subset both before and after NASA GISS' urbanization adjustments are shown in Figure 12. To calculate these trends, we first converted each of the station records into a "temperature anomaly" record, relative to 1951-1980. In other words, we subtracted the mean temperature of each station over the 1951-1980 period from all of the temperatures in that station's record. We then applied the gridding approach described in Section 3.3 to our subset. This gave us a gridded global temperature anomaly for each year. This procedure was carried out twice - once for the unadjusted data, and once for the adjusted data.

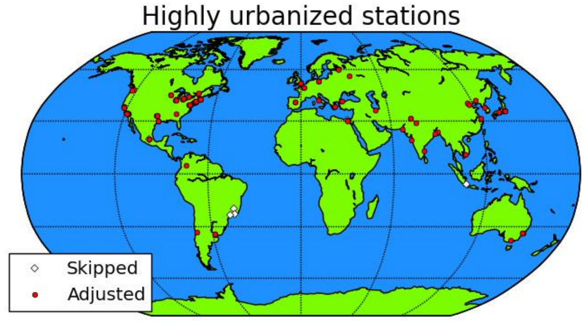


Figure 11: *Locations of those stations identified as highly urbanized in terms of night lights with an associated population of >2 million. The "Skipped" stations were dropped from NASA GISS' final analysis, and hence we do not consider them in our analysis.*

The unadjusted subset shows a strong, almost continuous, warming trend. If we approximate this warming as linear, this gives a trend of about $+1.33°C$/century over the period 1880-2011. This is more than twice the linear trend of NASA GISS' global temperature estimate (Figure 1), which we mentioned in Section 3.3 was about $+0.63°C$/century.

In other words, the highly urbanized subset shows considerably more warming than the average for the full dataset. This suggests that a substantial component of this warming is urbanization bias. So, if the NASA GISS urbanization adjustments are reliable, they should have substantially reduced the trend for the adjusted subset. From the bottom panel of Figure 12, we can see that, *up to about 1980*, the adjustments have indeed substantially reduced the warming trend, e.g., the 1895-1980 linear trend is reduced from $+1.02°C$/century for the unadjusted subset to $+0.21°C$/century for the adjusted subset. However, after about 1990, there is almost no reduction, and the 1990-2011 linear trend for both subsets is almost the same ($+2.16°C$/century for unadjusted and $+2.04°C$/century for adjusted).

This is more immediately obvious from Figure 13, where the gridded mean adjustment for the subset is plotted. Although the mean adjustments do not begin until about 1895, there is an almost linear mean

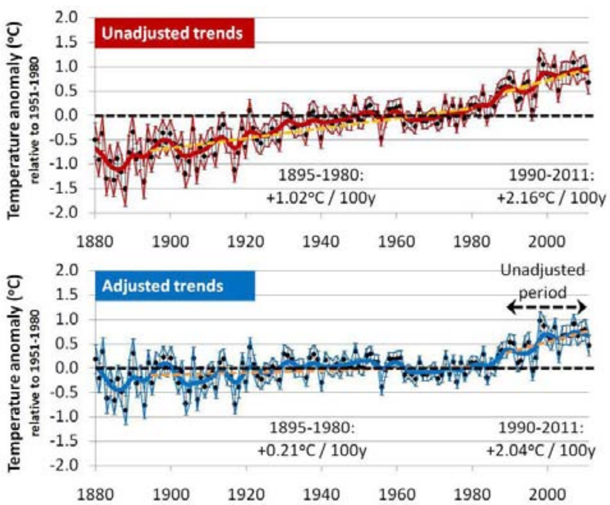


Figure 12: *Gridded mean temperature trends of the highly urbanized stations in Figure 11, before (top) and after (bottom) NASA GISS's urbanization adjustments. Thick lines correspond to 11-point binomial smoothed trends, and the error bars correspond to twice the standard error of the annual mean.*

adjustment from 1895 until the 1980s (1895-1980 linear trend of -0.79°C/century). But, this adjustment begins to dramatically decrease in the 1980s, and by the 1990s, there is only a slight adjustment (1990-2011 linear trend of -0.13°C/century).

The reduction in NASA GISS' adjustments since the 1980s is in direct contrast to the actual urbanization of the associated metropolises. As can be seen from the bottom panel of Figure 13, the total population of the 47 urban metropolises associated with the stations has more than trebled since 1950 (129 million in 1950 to 434 million in 2010). While population is not an exact measure of urbanization[46, 47], it is a reasonable indicator. So, the fact that the adjustments for the subset begin *decreasing*, rather than increasing, in the 1980s suggests a serious flaw in the NASA GISS urbanization adjustments.

We note that by removing a lot of urbanization bias from the pre-1980s records, but not much from the post-1980s records, this artificially makes global temperatures for recent decades appear more unusual than if they had been unadjusted. This is important, because the public seems particularly interested in establishing which years are globally *"the hottest on record"*, e.g., see Refs. [48–50].

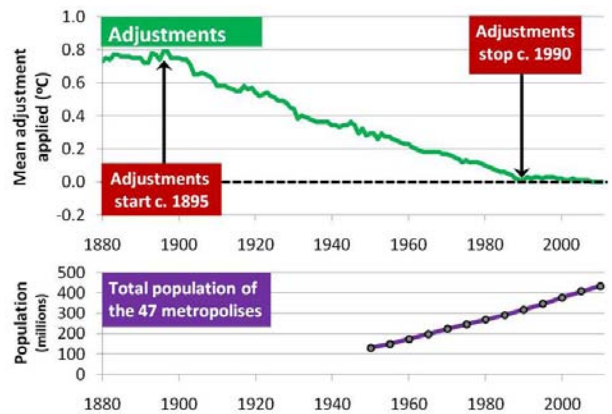


Figure 13: *Top: Gridded mean adjustments applied by NASA GISS to the temperature trends of the highly urbanized stations in Figure 11. Bottom: Total population of the 47 urban metropolises associated with those stations. Population figures taken from the United Nations Population Division website[8].*

3.5 Poor documentation of a non-intuitive approach

Although Hansen et al., 2010 boasted of transparency in describing their analysis and providing their source code and data[17], and they have published several relatively long articles describing their global temperature analysis[15–17, 19], there are quite a few non-intuitive and/or unexpected features and implications of their analysis for which they provided little or no discussion or justification. Some key features were not even described in their articles, but only revealed after a careful inspection of the code, e.g., the "extension rule" which we discuss in Section 4.1. Although NASA GISS should be commended for publishing their source code, it was only after considerable public pressure that they finally did so in 2007[21].

In addition, it is disappointing that, in December 2011, they ceased publishing their intermediate calculations. It was these intermediate calculations which enabled us to carry out most of our analysis for this article, so this meant we were unable to assess the impacts of their December 2011 change in datasets in as much detail, although we do discuss the impacts in general terms in Section 4.5.2.

Their approach of applying temperature adjustments in reverse chronological order, i.e., adding the calculated current urban bias to the start of the record, rather than subtracting it from the end of the record, appears unwise for several reasons:

- It is non-intuitive: why should we claim that, for instance, the Tokyo record was artificially "too cold" in 1900 because it *now* has a substantial urban heat island, rather than recognising that it is currently artificially too warm (Figure 3)?
- By forcing their adjustments to be zero for the most recent year, they have to identify the other end of their adjustments (i.e., the exact onset of the urbanization bias), with a much higher precision and accuracy than they would otherwise. As urbanization bias is generally a progressive phenomenon, it is easier to accurately identify its presence near the end of a record after it has grown substantial, than to accurately pinpoint the year in which urbanization bias began affecting the record.
- As urbanization is an ongoing phenomenon, it means that the entire set of adjustments must be recalculated and changed each month, as the latest data arrives. This can lead to considerable confusion comparing results from one month to the next, as each urban station's record continually has its "history rewritten".
- As NASA GISS have been rather terse in justifying their basis for taking this approach[15–17], it is liable to lead to suspicion amongst those sceptical of the reliability of NASA GISS' global temperature estimates[51]. Indeed, Hansen et al., 2010 recently complained of being the victims of unfair suspicion from their critics[17]. Perhaps this is part of the reason.

Although Hansen et al. briefly discussed their theories as to why urbanization bias could in some cases be a "cooling bias"[15–17], they do not explicitly discuss any examples of their Type 2, 3 and 4 adjustments. Indeed, the only two adjustment examples they explicitly discussed were of Type 1 (Tokyo, Japan and Phoenix, Arizona, USA)[15]. As a result, many users of NASA GISS' global temperature estimates might have mistakenly assumed that the "urbanization bias adjustments" were predominantly ones for urban warming bias, i.e., the usual assumption (Section 3.2). As we saw in Section 3.1, the vast majority of their adjustments involve correcting for urban cooling, i.e., the opposite of what would be expected. So, it is surprising that Hansen et al. did not explicitly discuss examples of these counter-intuitive adjustments.

In the previous sections, we described a number of unusual results of the NASA GISS adjustments which appear to contradict the generally accepted views on urbanization bias (e.g., see Ref. [5] or our review in Paper I[1]). When results contradict previous expectations, this should inspire researchers to look carefully at their results and methods, and the basis for the previous expectations. However, we could find little discussion of the divergence between the NASA GISS adjustment results and conventional views on urbanization bias. This suggests to us that NASA GISS have either only carried out a very limited analysis of their own results, or else have not adequately considered how their results compared to the literature expectations.

It is possible that part of this is due to confirmation bias, since it appears that many of the authors who were involved in the development and testing of the NASA GISS adjustments, had already concluded beforehand that urbanization bias was a small, possibly negligible, problem for global temperature estimates[52, 53]. Indeed, from Section 10 of Hansen et al., 2010[17], it seems NASA GISS consider their global temperature estimates to be politically sensitive, and as a result are concerned that, if critical analysis of their estimates revealed any flaws, they could be *"interpreted and misrepresented as machinations"*[17]. This suggests that they might have been reluctant to rigorously test their analysis, in case their tests revealed problems.

Whatever the reasons, it seems that the NASA GISS urbanization adjustment algorithm had not been subjected to sufficiently rigorous testing.

4 Flaws in NASA GISS' urbanization adjustment approach

In Section 3, we identified a number of serious problems with the urbanization adjustments that NASA GISS apply to their weather station data before generating their global temperature estimates. In this section, we discuss several flaws we have identified in the approaches they take to calculate these adjustments. We will try to offer suggestions as to how these flaws could be overcome.

4.1 Extension rule leaves urbanization bias in records

An unusual feature of the NASA GISS urbanization adjustment approach is their "extension range" rule. As we discuss in Paper III, there is a serious shortage of rural stations with very long records[2]. In particular, there is a sharp drop in the number of available rural records in their data set after 1990[44]. As NASA GISS require at least three rural neighbours to construct their rural averages for each urban station, this means that the rural averages often are unable to cover the entire period of the urban record.

Without a rural average for a given period, the NASA GISS adjustment algorithm is unable to determine what the urbanization bias for that period should be. This means that their urbanization adjustment cannot begin until there are at least three rural neighbours with overlapping records and has to stop if the number of neighbours drops below three.

As the periods of the urban record before and after the period of overlap with the rural average are not adjusted for urbanization bias, a reasonable approach would be to discard the unadjusted periods of the urban records. However, NASA GISS appear to have decided that this would involve shortening the urban records too much. Instead, they use an extension range rule: if the overlap between the urban record and its rural average is shorter than the urban record, they can include some of the longer part of the urban record effectively unadjusted. If the urban record ends after the rural average, then the adjustment for all of the remaining years is set to a constant of zero. If the urban record starts before the rural average, then the adjustments of all years up to the start are set to whatever the adjustment for the first year of the rural average is.

Examples of the extension rule in action can be seen in some of the example adjustments we discussed earlier. For the Phoenix, AZ (USA) station in Figure 2, there is an extension range at the start of its adjustment period, i.e., the adjustments do not start until 1895. For the Christchurch (New Zealand) station in Figure 6, there is an extension range at the end of its adjustment period, i.e., the adjustments end in 1980. The Tokyo (Japan) station in Figure 3 has extension ranges at both the start and the end of its adjustment period, i.e., the adjustments are only carried out for the 1914-1990 period.

The extension rule does not appear to be directly discussed in any of the published literature on their method[15–17], but can be confirmed by inspecting their Fortran code ("PApars.f"), or by analysing the results of their adjustments. They also mention that:

> *"An adjusted urban record is defined only if there are at least three rural neighbors for at least two thirds of the period being adjusted."*
> - Hansen et al., 1999[15]

Since they require at least three rural neighbours in order to construct a rural average, this statement does implicitly acknowledge that up to one third of an urban record can be included unadjusted via the extension rule. In any case, as we will see, the extension rule has quite a significant impact on their global temperature estimates. So, we find it surprising that they do not appear to have explicitly discussed the rule or its implications.

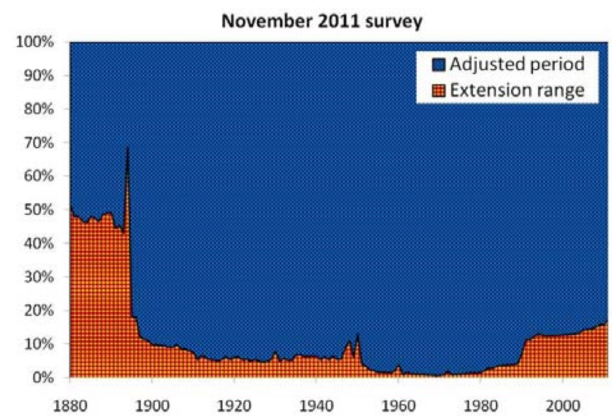


Figure 14: *Frequency of "extended periods", compared to the adjusted periods for the urban records. Taken from the November 2011 survey.*

Figure 14 shows, for each year, the percentage of urban records which are unadjusted as a result of being in the extension range. We can see that the percentage is very high (nearly 50%) for much of the late 19th century, but then reduces to about 5-10%. From 1951-1980, the percentage is lower still. However, in the 1980s, the percentage increases again, and has consistently been greater than 10% since 1991.

Figure 15 illustrates the relative distributions of the unadjusted and adjusted stations for six different years (1880, 1895, 1950, 1980, 1990, 2000). We can see that when there are large numbers of unadjusted stations, they are often clustered together. This makes sense - if there are not enough rural stations to construct a continuous rural average for one urban station, then the other urban stations in the

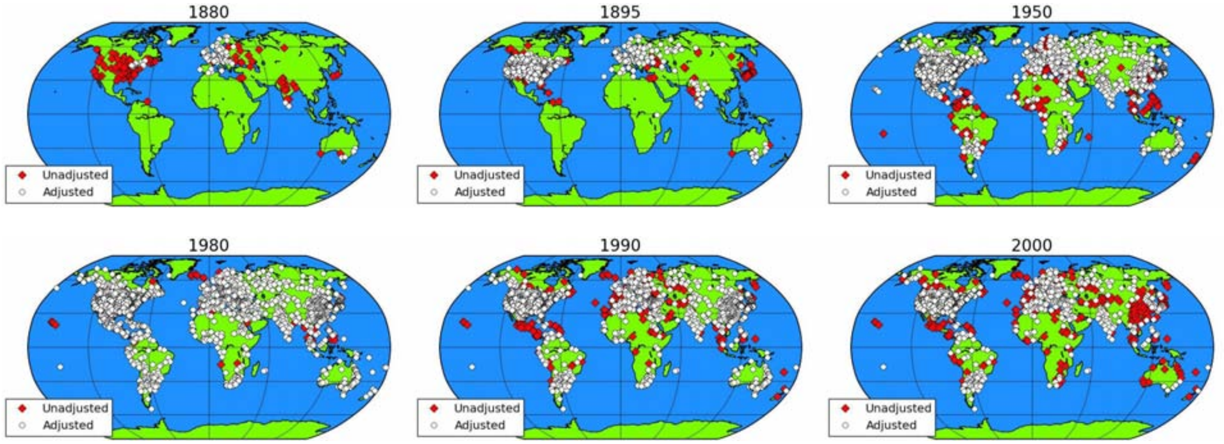


Figure 15: *Maps showing the locations of stations identified by GISTEMP as subject to urbanization bias, with data for 1880, 1895, 1950, 1980, 1990 and 2000. Red diamonds correspond to urban records that GISTEMP keeps unadjusted for that year, as there are too few rural neighbours. White circles correspond to urban records which have been adjusted for that year. The white circles were added after the red diamonds, so some red diamonds may be obscured by white circles.*

vicinity are likely to also have that problem. However, this leads to a double problem for those regions. When NASA GISS are calculating the gridded temperatures for those regions, they will be including a large number of urban stations unadjusted in those years. As they are unadjusted, they are likely to contain urbanization bias. In addition, because there were not enough rural stations to construct a rural average for those years, there will not be many rural stations contributing to the grid.

We know that urbanization has increased dramatically over the 20th century, e.g., see Figure 1 or Refs. [7–9]. So, the magnitude of urbanization bias was probably smaller during the late 19th century. For this reason, it is possible that the urbanization biases introduced into the global temperature estimates by the extension rule are relatively small for the late 19th century period, even though about half of the urban records are unadjusted then. However, since the 1980s, there has been a large increase in urbanization. If this has also led to an increase in urbanization bias (which seems probable), then the fact that more than 10% of the post-1990 records for urban stations are unadjusted is a serious concern. This can be easily seen by considering in detail the example of the Tokyo (Japan) station shown earlier in Figure 3.

Tokyo, the capital of Japan, is well-known to currently have a very large urban heat island[13, 54–59], which stretches out more than 30km[55, 58, 59]. However, the present existence of an urban heat island at the site of a weather station does not in itself indicate that the *trends* of its weather records suffer from urbanization bias. For instance, if the urban heat island has remained static for the entire record, then the temperatures for all years would be biased by a similar amount, and so there would be no overall trend from the bias[33].

The problem, then, is not how large the current urban heat island at the station in Metropolitan Tokyo is, but rather how much has it grown since the record began. Fujibe et al. found that there has indeed been considerable growth of the bias since the start of the record[13, 56, 57], and since the urban bias stretches quite far[58, 59] and Japan is a highly urbanized country, it is plausible that the rural stations which they used to estimate the bias are themselves partially affected by urbanization bias[60], a concern which Fujibe et al.[13, 56, 57] hint at.

As can be seen from Figure 3, NASA GISS does identify a quite substantial growth in Tokyo's urban heat island of 1.9°C over the course of its record. Their adjustments begin in 1914, with the slope increasing in the 1950s for the second leg of the adjustment, indicating an acceleration in the urban heat island growth. However, the adjustments end abruptly in 1990, despite the Tokyo record continuing up to present. As a result the adjusted Tokyo record shows a fairly flat trend from the late 19th century until

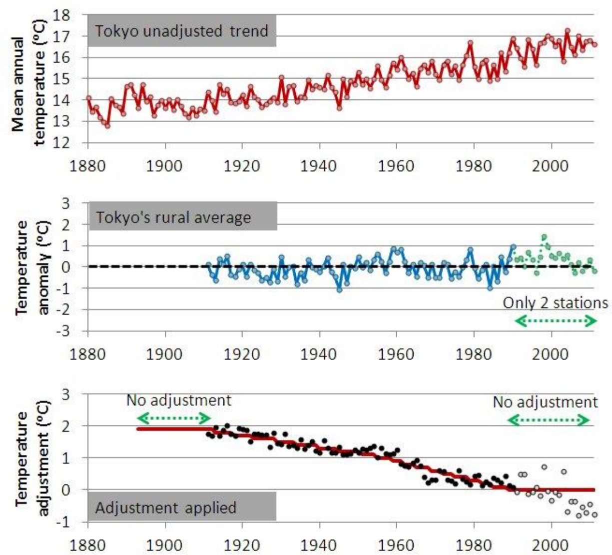


Figure 16: *Top: Unadjusted mean annual temperature trends for the Tokyo station. Middle: The rural average for the Tokyo station. Bottom: The red line represents the adjustment applied to the Tokyo record in November 2011. Circles represent the annual differences between Tokyo and the rural average, rescaled to match the red line. Hollow circles were constructed from just two stations (Hachijojima and Shionomisaki), and so those years are unadjusted following NASA GISS' "extension rule". 1.15°C was added to the rural average values to allow direct comparison.*

1990, at which point a "warming" trend begins. This warming trend continues to present.

The unusual adjustment pattern for the Tokyo station is a result of the extension rule. NASA GISS has eight stations within 500km of the Tokyo station, which they identify as rural (Hachijojima, Shionomisaki, Katsuura, Aikawa, Oshima, Irozaki, Nikko and Miyakejima). But, records for all but two of those stations (Hachijojima and Shionomisaki) finish in 1990. As a result, they cannot calculate a rural average for the post-1990 period. Therefore, because of the extension rule, they keep the Tokyo record *unadjusted* for that period. Similarly, while Tokyo has two rural neighbours with records beginning in 1906, a third neighbour does not begin until 1911, and so Tokyo has an extension range at the beginning of its record too.

It can be seen from Figure 16 that the increasing divergence between the Tokyo record and its rural average continued after 1990. So, if NASA GISS had continued their adjustment on the basis of the two remaining rural stations, the urbanization adjustments would have continued to be substantial. If they had simply dropped the post-1990 portion of the Tokyo record, they would have also avoided including urbanization bias. However, by using the extension rule, they have kept the post-1990 urbanization bias associated with the Tokyo record.

We saw in Figures 12 and 13 that similar incomplete adjustments appear in our gridded subset of 116 highly urbanized stations. This suggests that the problems caused by the extension rule, which we illustrated for the Tokyo station, are systemic.

4.2 Identification of urban stations is often based on inaccurate locations

Since Hansen et al., 2010[17], NASA GISS have been relying on the night-light brightness associated with the station co-ordinates to decide if a station is "urban" or "rural". However, sometimes the co-ordinates NASA GISS have for a station are incorrect. This can have serious consequences in regions with low station densities. We can illustrate this by considering the case of the Riyadh station in Saudi Arabia.

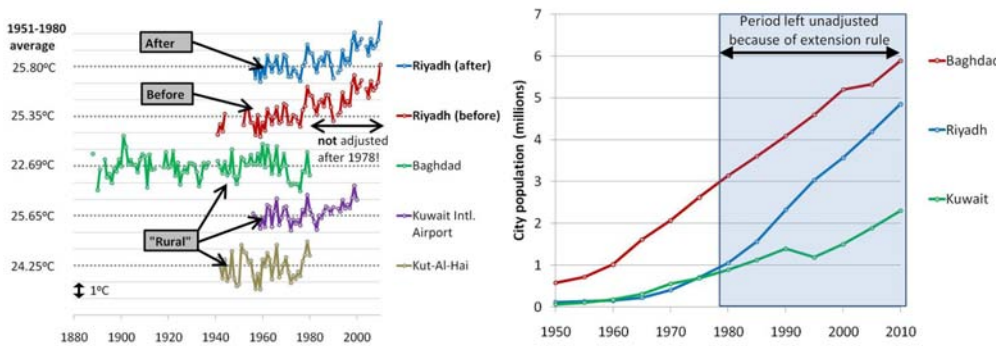
NASA GISS identifies the Riyadh station as urban, but as can be seen from Figure 17a, there are only three neighbouring stations identified as rural (Baghdad, Kut-Al-Hai and Kuwait International Airport) within the required 1000km, and with an overlap of more than 20 years with Riyadh's record. This is enough to construct a rural average for Riyadh, but only just.

As the capital of Iraq, Baghdad is one of the largest cities in the Middle East, and so it is quite surprising that the Baghdad station is identified as "rural". However, a close inspection using Google Earth (and the NASA Earth City Lights overlay), reveals that the co-ordinates NASA GISS use for the station ($33.23°N$, $44.23°E$) are on the outskirts of the Baghdad metropolis, and so its associated night-brightness is relatively low. So, this explains their identification.

From Figure 17b, it is clear that there is a serious error for their co-ordinates for Kuwait International Airport, however. The co-ordinates they use are for a location in the Persian Sea, i.e., more than 30km away from the actual airport. Obviously, the night-brightness at such a location (in the middle of the sea) is very low. This is why NASA GISS incorrectly

(a) **(b)**

Figure 17: *Google Earth aerial photographs of: (a) Riyadh and its three "rural" neighbours (using NASA Earth City Lights overlay) and (b) close-up of Kuwait International Airport, showing the actual station co-ordinates, and the co-ordinates used by NASA GISS*



(a) **(b)**

Figure 18: *(a) Temperature records of Riyadh and its three "rural" neighbours. "Before" and "After" refers to Riyadh's record before and after urbanization adjustment. (b) City population trends for Riyadh, Kuwait and Baghdad. Taken from the United Nations Population Division Home Page (`http://www.un.org/esa/population/unpop.htm`).*

identified the station as "rural".

The consequence of this can be seen from the temperature records in Figure 18a. It can be seen that it is only during the period after Kuwait International Airport's record begins (1956) and before Baghdad and Kut-Al-Hai's records end (1980) that Riyadh has three overlapping rural neighbours. For this reason, the post-1980 Riyadh record is unadjusted, due to the extension rule discussed in Section 4.1. But, even if there were enough stations to construct a rural average for the post-1980 period, because Kuwait International Airport has been misidentified as "rural", its urbanization bias would be incorporated into the "rural average" and so the Riyadh adjustment would have been incomplete.

The mistaken identification of Kuwait International Airport as "rural" has left urbanization bias in NASA GISS' estimates for the region in two ways:

1. NASA GISS does not attempt to adjust the Kuwait record because it is "rural".
2. The urbanization adjustment for Riyadh is inadequate as an urban station is mistakenly used for constructing the rural average.

If NASA GISS had correctly identified Kuwait International Airport as an urban station, then the urbanization bias of both stations could have been removed from the gridded temperatures for the region[3].

[3] In this case, the biases would have been removed by the two

For this reason, NASA GISS' current night brightness approach to identifying urban stations is very reliant on their station co-ordinates being accurate. O'Neill has noted on his blog quite a few other NASA GISS stations with inaccurate co-ordinates[61]. There are also other problems with their current method for identifying stations as urban, as we discuss in the next section.

4.3 Inappropriate use of U.S.-calibrated urbanization metric for rest of world

When NASA GISS first developed their urbanization adjustment method, they used estimates of the populations associated with each station as a metric for identifying which stations were urban, rural, or intermediate ("small town")[15]. These were the estimates provided by Peterson & Vose, 1997[43] when they developed the Global Historical Climatology Network dataset used by NASA GISS. However, there were several problems with this metric:

- The somewhat ad-hoc nature with which the population estimates were compiled, meant that it was really more a qualitative rather than a quantitative identification.
- Many of the population figures were probably out-dated by the late 1990s.
- It is plausible that a station in the centre of a small town may have observed more urbanization bias than a station on the outskirts of a large city. Therefore the population of the nearest town associated with a station is not necessarily the best indicator of the urbanization experienced by the station.
- Population growth is only approximately related to urban growth[46, 47].

As a result, NASA GISS decided to consider alternative metrics for identifying stations as rural or urban.

Imhoff et al., 1997 took a composite image of "night-time city lights" for the continental United States, made from 231 orbital swaths gathered by the Defense Meteorological Satellite Program's Operational Linescan System (DMSP/OLS) over a six month period between 1994 and 1995[45]. They then, by trial-and-error, established a threshold value of night brightness which corresponded to urban areas in several U.S. metropolitan areas. By using this threshold value, they were then able to construct a map of urbanization for the U.S. This map showed reasonable agreement with U.S. Census-derived population densities, suggesting that it could be used as a reliable proxy for urbanization in the U.S.[45].

stations being dropped from their estimates, since there would then only have been two rural stations left for constructing the rural averages.

NASA GISS decided in 2001 to switch to using Imhoff et al.'s dataset to identify urban stations in the U.S.[16], instead of the population-based method. However, although this dataset provided night-light brightness values for most of the planet[62], the urban threshold they calculated was calibrated using the types of urbanization which occurred in the U.S. It was therefore not appropriate for using outside the U.S.:

> *"...although this technique worked well in the United States (i.e., in a developed country), it is untested in lesser-developed countries where the type of infra-structure development and its associated nighttime lighting may be different"* - Imhoff et al., 1997[45].

For this reason, NASA GISS only applied the night-lights criteria to stations in or near the U.S.[16], and continued to use the population metric for the rest of the world.

In 2010, NASA GISS changed their mind and decided to use Imhoff's U.S.-calibrated night-light brightness values to classify *all* of their stations[17]. Their justification for doing this is as follows:

> *"The relation between population and night light radiance in the United States is not valid in the rest of the world as energy use per capita is higher in the United States than in most countries. However, energy use is probably a better metric than population for estimating urban influence, so we employ* [the same threshold] *as the dividing point between rural and urban areas in our global night light test of urban effects."* - Hansen et al., 2010[17].

The night-light brightness values do indeed seem to be better correlated to energy use (and GDP) than to population density[62]. But, it is unclear why Hansen et al. assume *a priori* that energy use is a better indicator of urbanization bias than population density. Naïvely, one might suppose that all of these proxies (population density, energy use, night-light brightness) are indicators of urbanization, and so the distinction is irrelevant. However, as NASA GISS uses a threshold value for its urbanization proxy, it is critical that a reasonable threshold value is used.

Since the U.S. has an anomalously high per-capita electricity usage[17, 45, 62], the urban threshold Imhoff et al. had chosen for the U.S. might not be sensitive enough in other countries[45]. For example, the U.S. consumed 3,906 billion kilowatt hours of electricity in 2008, versus 601 billion kilowatt hours by India (from US Energy Information Administration). In comparison, USA had a population of 310 million to India's 1,225 million in 2010 (from UN Population Division). So, USA only has a quarter of the population, but still uses 8 times as much electricity as India, i.e., in 2008, U.S. electricity usage was more than 25 times that of India, *per capita*.

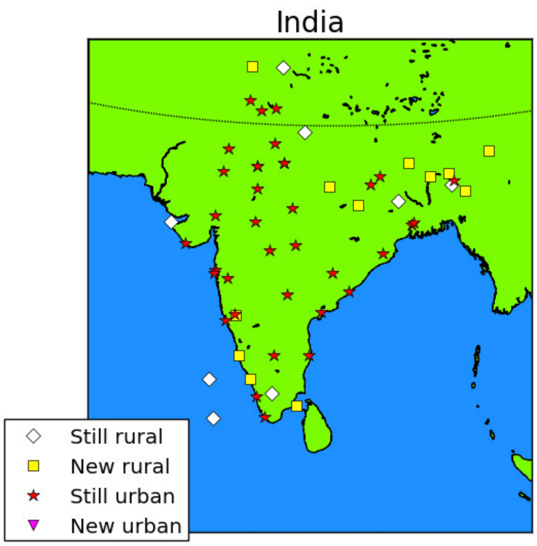


Figure 19: *Map showing the locations of all the NASA GISS stations for India (i.e., those with a weather station country code of "207"). White diamonds correspond to stations which remained "rural" and red stars to stations which remained "urban" with Hansen et al., 2010[17]'s transition to using U.S. calibrated night light brightness for identifying urban stations. Yellow squares correspond to those stations which were "urban" with the original identification, but are now identified as "rural". None of the rural India stations were changed to "urban" under the new scheme.*

NASA GISS' new night-lights threshold has more than doubled the number of "rural" stations for the Indian subcontinent (from 8 to 20) by including stations which were classified as urban under the old population-based threshold, but did not meet the new U.S. calibrated night-lights threshold. The new threshold failed to reclassify any extra stations as urban (see Figure 19). Instead, the number of stations NASA GISS attempt to adjust for urbanization bias has *decreased* from 46 to 34 for India.

By analysing the locations associated with the "'new rural" stations using Google Earth, and the 2001 Census of India, it appears that a number of these stations are in, or near, highly urbanized areas. For example, Figure 20 shows four of the twelve stations reclassified as rural - Dhubri, Gauhati, Pamban and Srinagar, and their corresponding station records. If we assume that the station co-ordinates used by NASA GISS are accurate (although, see Section 4.2), then the Pamban station is located less than 2km from the town of Rameswaram, with a 2001 population of $\sim 38,000$. Dhubri station appears to be located near the centre of another town, Dhubri (2001 population $\approx 64,000$). Gauhati appears to be located at an international airport on the western outskirts of the city of Guwahati (2001 population $\approx 819,000$), while the Srinagar station is in the middle of the city (2001 population $\approx 988,000$). So, it is quite plausible that some of these stations may have been affected by urbanization.

We can see that, for India, the new threshold is less strict, and more likely to mistake stations with urbanization bias as "rural". As we discussed in Section 4.2, when this happens it causes two serious problems for NASA GISS' urbanization adjustments. First, it means that the stations mistakenly identified as rural will be included unadjusted. Second, their trends, which may have urbanization bias, can be incorporated into the "rural averages" which are used to estimate the urbanization bias of its neighbours. If the rural averages inadvertently include any urbanization bias, then the NASA GISS approach will underestimate the magnitude of the urbanization bias in those urban stations it does adjust.

For these reasons, with the NASA GISS approach, it is probably better to have a stricter threshold even if it might falsely identify some rural stations as "urban", rather than a laxer threshold which will leave stations with urbanization bias unidentified. In this sense, their 2001 decision to use Imhoff et al.'s U.S. calibrated night light threshold for their U.S. stations was probably a good idea. But, it does not seem a good idea for the India stations.

NASA GISS did consider the possibility that extending their U.S. calibrated threshold to the rest of

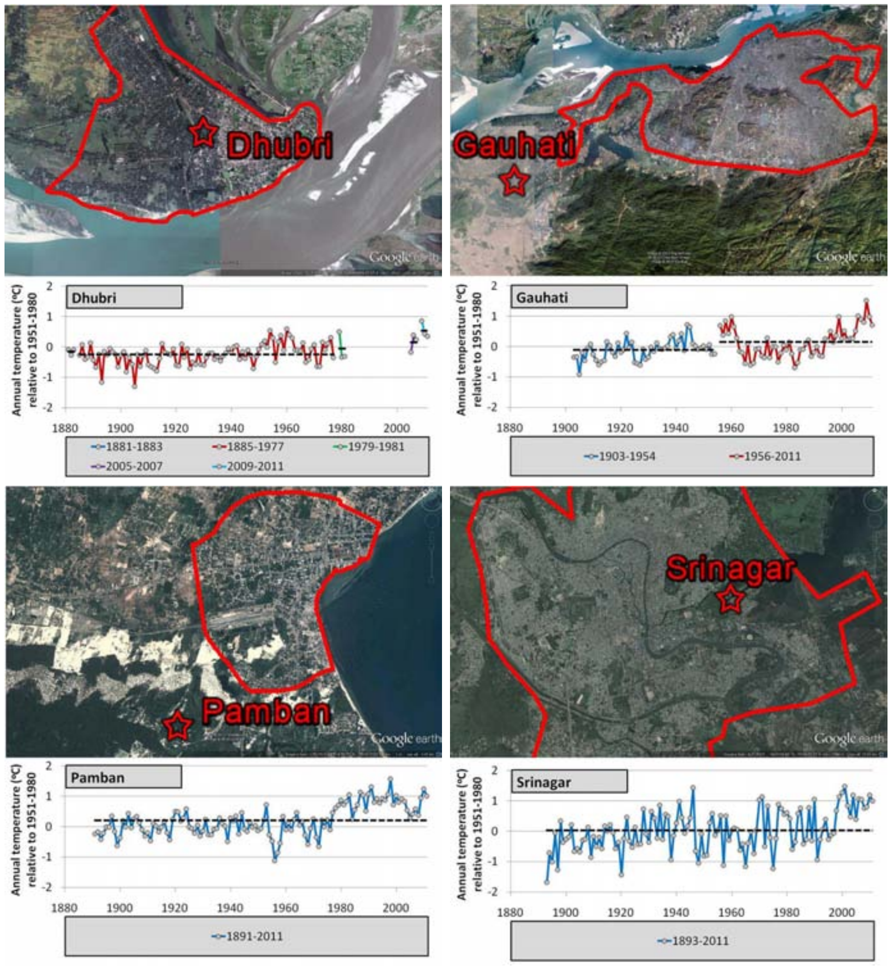


Figure 20: *Google Earth aerial photographs of four of the twelve "new rural" stations in India, and their annual temperature anomalies. Red lines approximate the outline of major towns/cities in the vicinity of the stations.*

the world, might be problematic for some places:

> *"This night light criterion is stricter than the population criterion in the United States... However, as we will see, the opposite is true in places such as Africa"* - Hansen et al., 2010[17].

To investigate how serious a problem this was, we used the country codes associated with each station to calculate (for each of the seven continents) the percentages of stations identified as "urban" and "rural" under both the original (population-based) and new (U.S. calibrated night brightness) thresholds. We show the changes in percentages in Figure 21. The only continent which showed an increase in the strictness of the urban threshold was North America. Aside from Antarctica, which is identified as 100% rural under both criteria, all of the other continents showed a decrease in the number of stations identified as urban. Hansen et al., 2010's claim that the night light criteria is less strict *"in places such as Africa"*[17] seriously underestimates the problem.

We agree that using associated populations as a metric for urbanization is not ideal, as they are only approximately related[46, 47]. However, adopting a U.S.-calibrated night brightness as a replacement metric seems unwise. Perhaps, a combination of different metrics could be used instead.

4.4 Limited availability of long, complete, rural records

A major difficulty in attempting to calculate the magnitude of urbanization bias in urban station records

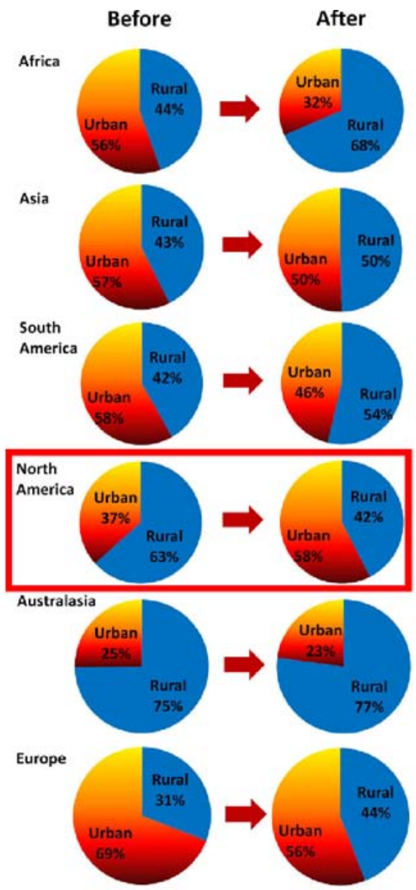


Figure 21: *Changes in stations identified as "urban" by continent after NASA GISS switched to using their U.S.-calibrated night brightness metric to identify urban stations. North America was the only continent with an increase in the percentage of stations identified as urban. All of the stations for Antarctica are considered rural under both metrics, and so are not shown.*

is the severe shortage of neighbouring rural stations with long and relatively complete records for comparisons. One problem is that heavily urbanized areas tend to be surrounded by moderately urbanized areas, and often the outskirts of an urban sprawl are still quite urbanized. This means that in the regions which are most likely to be severely affected by urbanization bias, the nearest rural neighbours may be a long distance away. For instance, we mentioned in Section 4.1 that the urban heat island associated with Tokyo (Japan) stretches out more than 30km[55, 58, 59].

Another problem is that the least urbanized areas are, by definition, sparsely populated. It would have been difficult for early observers to convince staff to maintain continuous daily records at these remote locations for several decades. In recent decades, the development of automatic weather stations has reduced this problem, but obviously this cannot provide us with records for the mid-20th century, or earlier. In the past, some meteorological organisations paid weather observers extra money to maintain weather records at remote, rural locations, e.g., daily observations were recorded manually almost continuously at the Mount Säntis weather station in Switzerland, from the time it was set up in 1882[63] until the installation of an automated weather station in the late 1970s[64]. However, it is difficult to find many stations for which long, continuous, records have been maintained, and which have not been affected by any urban development or modernisation over the course of that record. Indeed, in recent years, the location of the Mount Säntis station has become a popular mountain resort (Säntis der Berg).

In Paper III, we describe the shortage of long, complete, rural records in some detail[2]. However, for this study, two examples should suffice to illustrate the problems this shortage poses for the NASA GISS urbanization adjustments.

In Section 4.3, we noted that NASA GISS only have records for a few rural stations in India. 12 of the India stations NASA GISS currently identify as rural were identified as "urban" under their pre-2010 population-based urbanization metric. We saw in Figure 20 that several of these stations are likely to be affected by urbanization bias.

Figure 22 shows the temperature records for all eight of the India stations which are identified as rural by both the population-based and night brightness-based metrics. In other words, these are the stations NASA GISS has for India which are least likely to be affected by urbanization bias. There are several points to note about these stations and their records:

- From Figure 19 we can see that all of the eight stations are either in the mountains near the northern borders, or else coastal/island stations, while many of the urban stations are in central India. In other words, the rural stations are in climatically different regions from many of the urban stations they are being compared to.
- Most of the records have data gaps lasting several years.
- None of the records show much similarity to the "global temperature trends" of Figure 1.

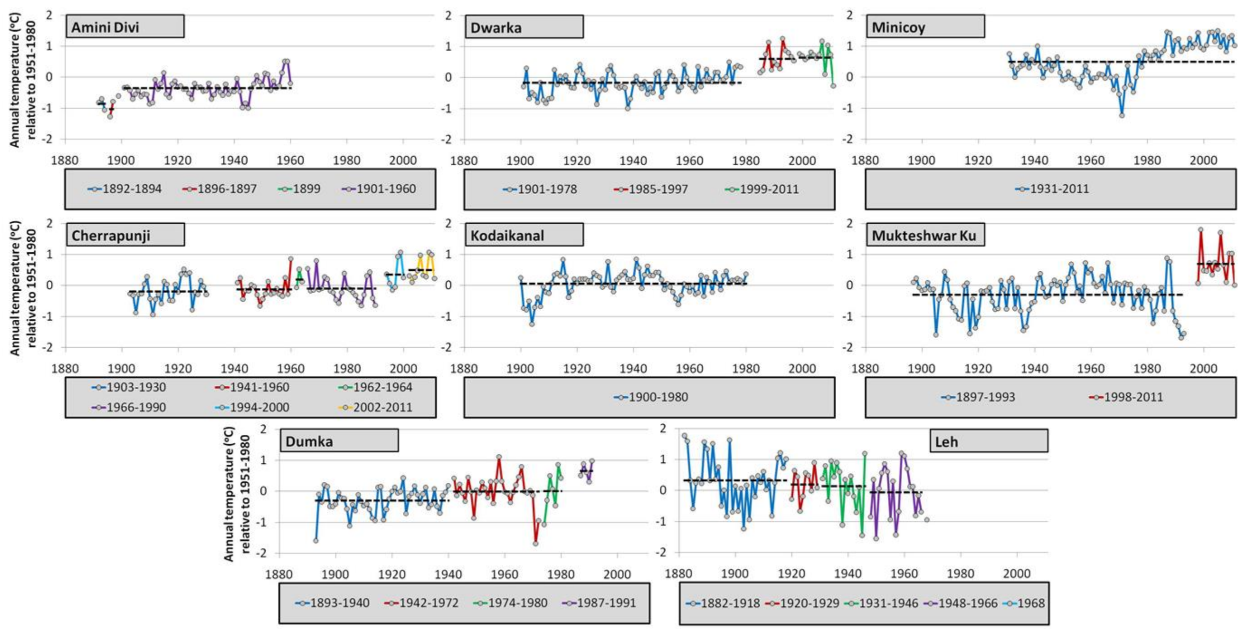


Figure 22: *Mean annual temperature trends of all 8 of the original rural stations on the Indian subcontinent used by NASA GISS. Black dashed lines represent the means of each period without missing data.*

- During the complete segments of the records (i.e., the segments in between data gaps), the temperature trends tend to oscillate above and below a mean value, i.e., long-term warming or cooling trends are generally absent. Again, this disagrees with the idea of long-term "global warming" implied by NASA GISS' global temperature estimates.
- When substantial "warming" or "cooling" does occur in the records, it often coincides with a missing data period, followed by a step change in mean temperatures. This is characteristic of non-climatic step-change biases, such as a station move or a change in instrumentation.
- There is a remarkable lack of coherence between stations in these warming/cooling trends. This agrees with the suggestion that many of the apparent trends in the records involve non-climatic biases.

For all of these reasons, rural averages constructed from these stations are unlikely to accurately describe the genuine climatic trends which their urban neighbours would have described if they did not have any urbanization bias. Hence, NASA GISS' estimates of the urbanization biases of the urban stations in India will be unreliable.

As it happens, the rural station records for India are relatively long compared with other parts of the world. It is instructive to consider the effects of NASA GISS' urbanization adjustments in regions where the rural records are shorter, e.g., in Peru.

We saw in Figure 4, that NASA GISS' urbanization adjustment for the Peruvian station, Piura, assumed that the record was biased by strong urban cooling. The magnitude of the adjustment was so large that it changed the long-term trend for the station from a strong cooling trend to a strong warming trend. As we discussed in Section 3.2, urbanization bias typically leads to artificial warming. So, it is worth investigating why NASA GISS calculate the bias to be the opposite sign.

From Figure 23, we can see that the adjustment is large in order to make the Piura trend match the warming trend of the rural average. *If* the rural average accurately represents the underlying climatic trends that Piura would have experienced if it was not urbanized then this would be a reasonable adjustment to make, since the difference in trends would presumably have been due to non-climatic problems with the Piura record. However, if the rural average is an *inaccurate* representation of the climatic trends then the adjustment would be completely inappropriate. So, it is important to look at how this rural

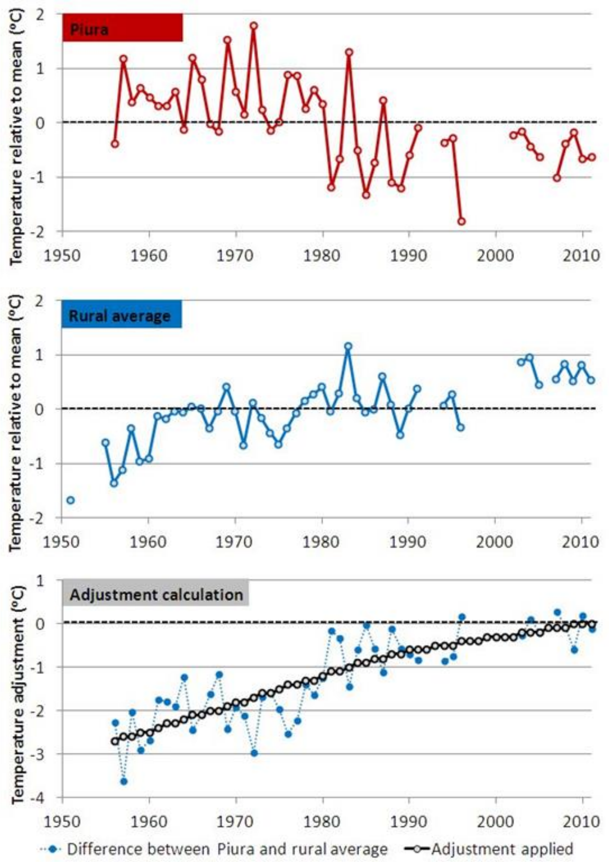


Figure 23: *Temperature records of Piura and its rural average for November 2011. For the bottom panel, 1.3°C was subtracted from the rural average values to allow direct comparison with the applied adjustment.*

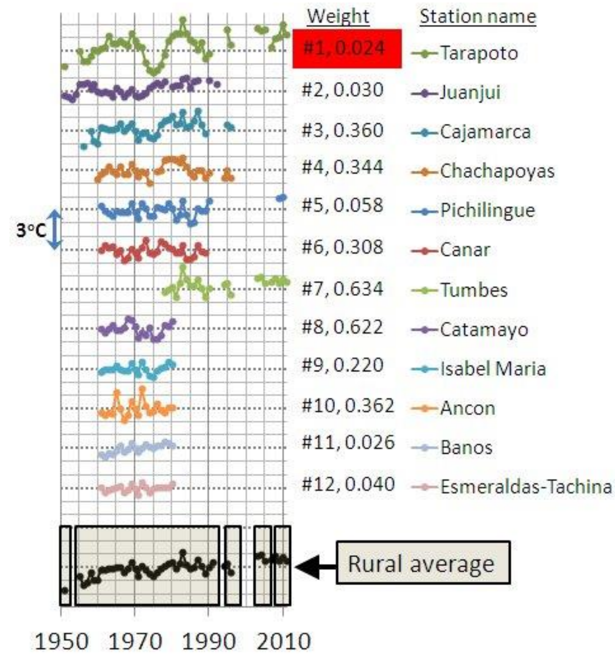


Figure 24: *Temperature anomaly records for Piura's twelve rural neighbours. The black line at the bottom which is labelled "Rural average" is the rural average NASA GISS calculated for Piura in November 2011, i.e., the middle panel of Figure 23.*

average was constructed.

Figure 24 shows the temperature trends of all twelve of Piura's rural neighbours. We find that, during the short periods when the stations overlap a lot of the records show similar fluctuations. For instance, they all suggest it was a relatively warm year in 1983. However, unfortunately, five of the stations only have 20 years of data (1961-1980), one of the stations (Canar) finishes in 1989, and three of the stations finish in the mid-1990s, as well as having a number of data gaps. Of the remaining three stations, all three of them have large gaps in their records, Pinchilingue only has one post-1990 value, and Tumbes' record only starts in the late 1970s.

The trend for the remaining station, Tarapoto, is quite unlike the other 11 stations. Since it is also the furthest of the stations from Piura, it only makes a relatively small contribution to the rural average for those years when there is data from some of the closer stations[4]. However, because the Tarapoto record is the most complete of the rural stations, there are several years (particularly in the 1950s and 2000s) when it is either the only station included in the rural average, or else one of just two or three stations.

Presumably, the most reliable portions of the rural average in this case are those years when the rural average was constructed from a large number of stations, most of which showed similar trends and fluctuations. Arguably, this is the period 1961-1990 (and possibly during the mid-1990s), during which there does not appear to be any major trend (either warming or cooling).

Essentially, the "warming" trend in Piura's rural average appears to be mostly due to one station which showed quite different trends from the others (Tarapoto), and one station whose record only began in the late 1970s (Tumbes). Neither of these records

[4]The relative weights of the rural stations to the "rural average" are inversely proportional to their distance from the urban station, and are listed on the left hand side of Figure 24.

show much similarity with the other ten rural stations. But, because the other records are so short, the long-term trends of the rural average are dominated by Tumbes and Tarapoto. This illustrates that the reliability of NASA GISS' adjustments can be seriously reduced if there is a shortage of rural neighbours with long records.

4.5 Failure to account for other non-climatic biases

When NASA GISS initially introduced their adjustments in 1999, they used a dataset which had undergone no adjustments for non-climatic biases. This dataset was the unadjusted version 2 of the Global Historical Climatology Network (GHCN)[43]. Aside from two specific stations (Lihue, HI (USA) and St. Helena Island), they did not attempt to correct for any non-climatic biases, other than urbanization. They explicitly assumed that, other than urbanization, any biases would tend *"to average out in large area averages and in calculations of temperature change over long periods"*[15]. In 2001, they decided to switch to using a "homogeneity-adjusted" dataset for the U.S. component of their analysis[16]. In December 2011, they decided to switch to using a "homogeneity-adjusted" dataset for the rest of the world. This was the homogeneity-adjusted version 3 of the Global Historical Climatology Network[20]. As we mentioned in Section 2.1, they also stopped publishing their intermediate calculations then, and since we used these calculations for our surveys, our last survey occurs before this change-over, i.e., November 2011.

Each of these approaches makes different assumptions, and has its own problems. So, in this section, we will consider the problems of the different approaches separately. In Section 4.5.1, we will consider the flaws in the approach NASA GISS took until December 2011, i.e., assuming that non-climatic biases other than urbanization biases can be ignored. This is the approach described in their peer-reviewed documentation, i.e., Refs. [15–17, 19], and led to their global temperature estimate which was used in the most recent IPCC report[18].

Since December 2011, NASA GISS have been taking a different assumption, i.e., that the homogeneity adjustments applied to version 3 of the Global Historical Climatology Network dataset have successfully removed these other biases, without introducing replacement biases. As this is a relatively recent change, it has not been discussed much in the peer-reviewed literature yet. But, the global temperature estimates constructed from this approach have already received considerable media attention, e.g., Refs. [48–50]. We discuss the homogeneity adjustments applied to this replacement dataset in detail in Paper III[2], but in Section 4.5.2, we will also briefly consider the impacts of the December 2011 change in datasets on NASA GISS' global temperature estimates.

4.5.1 The effect of other biases on NASA GISS' urbanization adjustments

We saw in Figure 5 that NASA GISS' urbanization adjustment for the Dublin Airport (Ireland) station was a Type 3 ("urban warming then urban cooling") adjustment, during the November 2011 survey. As we discussed in Sections 3.1 and 3.2, urbanization bias typically leads to artificial warming, so "urban cooling" should not be a frequent occurrence, let alone urbanization bias which starts off causing urban warming, but then switches to causing urban cooling. Nonetheless, 39.1% of NASA GISS' adjustments in the November 2011 survey were of Type 3.

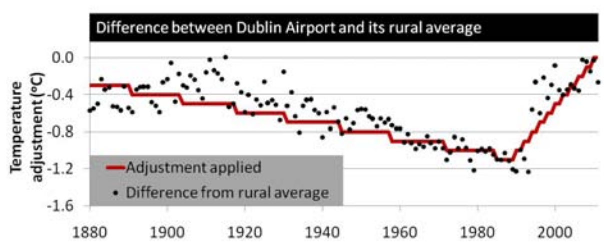


Figure 25: *Comparison between the urbanization adjustment applied by NASA GISS to the Dublin Airport, Ireland station (red line) and the difference between Dublin Airport and its rural average (black dots), for the November 2011 survey. 0.6°C was subtracted from the rural average values to allow direct comparison.*

We suggest that many of these unusual adjustments are due to the presence of other non-climatic biases in the urban records and/or the rural records, as well as urbanization bias. We illustrate how by using the example of the Dublin Airport station. Figure 25 compares the difference between Dublin Airport and its rural average to NASA GISS' bi-linear adjustment. We agree that, if the difference is to be modelled with a bi-linear adjustment, then NASA GISS' adjustment is probably the best approximation. The problem is that the bi-linear approximation is inappropriate in this case.

In Section 2, we summarised the basis for NASA GISS using this bi-linear approximation: Urbanization bias is a trend bias, and this trend may change over time (i.e., it is not strictly linear). For this reason, they use a bi-linear fit, to allow *"some time dependence in the rate of growth of the urban influence"*[15]. This is in itself a reasonable approximation. However, their method breaks down if there are any other non-climatic biases in the station records of either the urban station or its rural average.

There are many potential biases which may occur in any or all of the stations, whether "rural" or "urban". For instance, changes in station location[24], observation practice[24, 65], station microclimate[66], instrumentation used[67] or local land use[68] can all lead to non-climatic biases in station records. As a rough approximation, we can divide these biases into two types[26]:

1. "Step" biases, which involve some event (e.g., if the station is moved or nearby trees are cut down) which affects all subsequent temperature readings by a similar amount[5].

2. "Trend" biases, which involve a continuous, gradual change from year to year (e.g., an expanding urban heat island or the growth of nearby trees).

In Figure 26, we directly compare the Dublin Airport record to that of its rural neighbour, Valentia Observatory (Ireland). From the 1940s to the mid-1990s, we see a gradual reduction in the difference between the warmer Valentia Observatory and the colder Dublin Airport (Valentia Observatory is in the southwest of Ireland, which is climatically warmer). This suggests the possibility of urban warming at the Dublin Airport station. This agrees with the difference between Dublin Airport and its rural average (Figure 25), although this is not surprising, since Valentia Observatory is one of the longest rural stations in the area (see our discussion in Paper III[2]), and so is a major contributor to Dublin Airport's rural average. However, around 1994, this reduction is abruptly reversed.

[5]In reality, step biases do not necessarily affect all readings by the same amount, e.g., neighbouring trees may shelter the thermometer station from certain winds or increase its shading, but if there is annual variability in the types of winds and their directions, or the amount of cloud cover, the effect of cutting down those trees on mean monthly (or annual) temperatures may vary from year to year

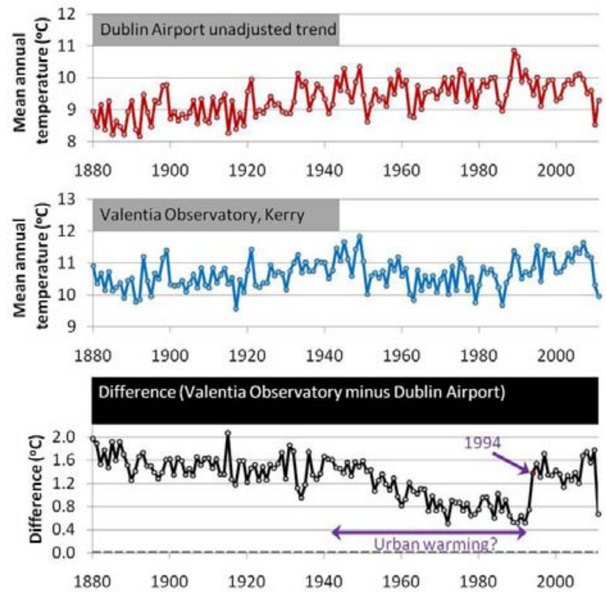


Figure 26: *Comparison between the unadjusted Dublin Airport station and the neighbouring rural station of Valentia Observatory (County Kerry, Ireland).*

The rapid change in the difference series suggests a non-climatic step change at either Valentia Observatory or Dublin Airport. We compare the Dublin Airport record to the record for another Dublin station, Phoenix Park, in Figure 27. The Phoenix Park station data is not in the dataset used by NASA GISS, but we were able to download it from the ECA&D project[69]. Again, there is an abrupt step change in 1994. This indicates that the bias is associated with the Dublin Airport record. According to comments in Table 3 of Sweeney, 2000[70], the location of the wind measurements for Dublin Airport was changed in 1994. It is probable that the location of the temperature measurements also changed then, and this would explain the step change.

We can now understand why NASA GISS calculated the urbanization bias at Dublin Airport as a Type 3 adjustment. The Dublin Airport record contains *both* a strong urban warming trend bias of about 0.7-1.0$°C$, and an abrupt "cooling" step bias (in 1994), also of about 0.7-1.0$°C$. Because their adjustment method only allows for bi-linear adjustments, this second (non-urbanization) bias confounded their algorithm and led to the false conclusion that the urbanization bias changed from warming to cooling.

If the Dublin Airport station was subject to two major biases - an urban warming trend bias *and* a

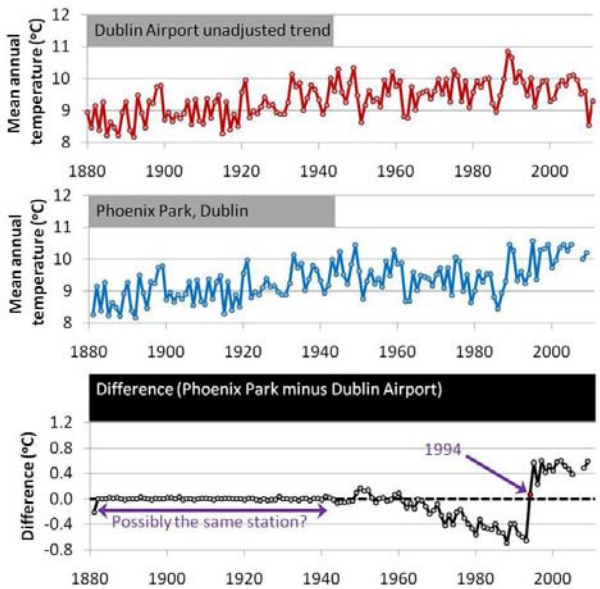


Figure 27: *Comparison between the unadjusted Dublin Airport record and temperatures for another Dublin station, Phoenix Park. The Phoenix Park record is not included in the dataset used by NASA GISS, but was constructed by applying the same December-January annual averaging used by NASA GISS to daily temperature data from the European Climate Assessment & Dataset project. Phoenix Park is a large public park located near the centre of Dublin City.*

station move step bias - then you might argue that it is a good thing that NASA GISS' adjustment unintentionally included the second bias as part of the "urbanization bias" - after all, both biases are non-climatic, and should be removed. The problem is that the urbanization bias model they used is unable to handle the superimposing of biases of different types. There are two main reasons for this.

First, step biases and trend biases have different properties, and using a trend adjustment to remove a step bias is problematic. In Paper III[2], we discuss de Gaetano, 2006's observation that treating a trend bias as a step bias leads to an underestimation of the bias, due to aliasing[26]. This is because the step bias is approximated by the mid-point of the trend bias, and so correcting for a step bias only removes *some* of the bias. The corollary of this is that treating a step bias as a trend bias leads to an overestimation of the step bias. In this case, NASA GISS over-corrected for a cooling step change in 1994 with a warming trend from 1989 to 2011.

Second, because the NASA GISS adjustments only are designed for one net bias for each leg of the adjustment, in order to account for the step change cooling bias in 1994, the second leg of the Dublin Airport adjustment cannot also correct for a warming trend bias. So, there is no correction for the *actual* urbanization bias during this second leg, i.e., 1989-2011.

It can be seen therefore that any biases other than urbanization bias which occur in an urban station's record[71] can easily confound the NASA GISS approach. However, the same can also occur if there are biases in the neighbouring rural stations. On the one hand, such biases would have less effect on an individual urban adjustment, since the rural average is constructed from the trends of at least three rural stations. However, on the other hand, the effect could be spread into many adjustments, since the biased rural station's record could be included in the rural averages of several nearby urban stations. In highly urbanized areas, there may be many urban stations which are being corrected, and only a few rural stations which are used for constructing the rural averages, e.g., the case of India which we discussed in Section 4.4. This means that those few rural stations need to be reasonably bias-free or else the NASA GISS approach could incorrectly contaminate a large number of urban stations with non-climatic trends.

As we discussed in Sections 3.1 and 3.2, urbanization bias is predominantly a warming bias, so most of NASA GISS' urbanization adjustments should be of Type 1. We suspect that a major reason why there were so many adjustments of the other types in all of our surveys (see Table 1) is that their adjustment technique was confounded by other non-climatic biases in the station records, as happened in the Dublin Airport example. If this is correct, then the cancelling-out of their "urban cooling" and "urban warming" adjustments, which we discussed in Section 3.3, was invalid, and their adjustments were not just inadequate, but may have actually introduced artificial biases into their estimates.

4.5.2 Problems with the new dataset used by NASA GISS

Figure 28 compares NASA GISS' global temperature estimates from November 2011 (i.e., one using the unadjusted version 2 of the Global Historical Climatology Network dataset) to that from December 2012 (i.e., one using the homogeneity-adjusted version 3 of the Global Historical Climatology Network dataset). The change in datasets has introduced a

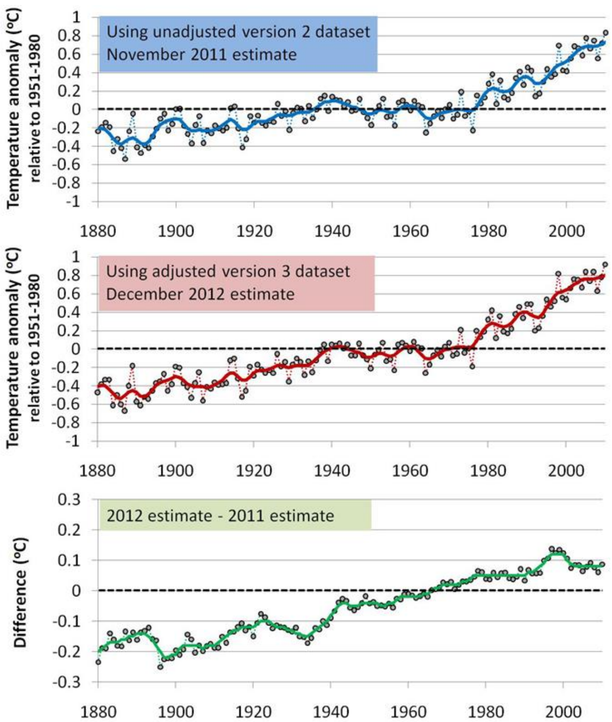


Figure 28: *The top two panels show the NASA GISS estimates of global temperature trends (land only) calculated in November 2011, using version 2 of the* unadjusted *Global Historical Climatology Network dataset (top); and in December 2012, using version 3 of the* adjusted *Global Historical Climatology Network dataset (middle). The bottom panel shows the difference between the two estimates (December 2012 estimate - November 2011 estimate). Solid lines represent 11-point binomial smoothed versions.*

substantial warming trend (bottom panel). If we approximate this trend as linear, this gives a trend of $+0.25°C$/century) over the 1880-2011 period. As we mentioned in Section 3.3, the global temperature estimates are not exactly linear, so linear trends are only crude approximations of the actual trends. Nonetheless, if we take this approximation, then the November 2011 estimate gives a trend of $+0.63°C$/century, while the December 2012 estimate gives a trend of $+0.88°C$/century (over the same 1880-2011 period).

In other words, a simple change in datasets used by NASA GISS has increased the "global warming" trend by about 40%. This is a substantial change in trends to occur from just changing datasets. So, it is worth investigating which of the two datasets is the more reliable, if either.

In 2010, the NOAA National Climatic Data Center introduced version 3 of the Global Historical Climatology Network[20], which NASA GISS uses for their main dataset. As part of this version, the National Climatic Data Center had updated their previous homogeneity adjustment approach to use the Menne & Williams, 2009 algorithm[25]. As mentioned above, until December 2011, NASA GISS had preferred to use the unadjusted version of the earlier dataset. But, they seem to have decided that this new homogeneity-adjusted dataset is more reliable. Unfortunately, NASA GISS also decided to stop publishing the intermediate calculations, which we used for our surveys, so we are unable to directly analyse the effects this change had on their individual adjustments. However, we can assess the reliability of the homogeneity-adjusted version of the dataset.

It is worth noting that the homogeneity-adjustments of version 3 did successfully identify and correct for the 1994 step change at Dublin Airport. So, in some cases, the homogeneity adjustments improve the reliability of the dataset. However, as we discuss in Paper III, the adjustments also transferred urbanization bias from Valentia Observatory's urban neighbours, such as Dublin Airport, into the Valentia Observatory record[2]. We find that this "urban blending" between rural and urban stations is a systemic problem when the Menne & Williams, 2009 homogeneity adjustments[25] used are applied to a highly urbanized network such as the Global Historical Climatology Network dataset.

Since the rural stations in the homogeneity-adjusted dataset partially contain urbanization bias from urban blending, NASA GISS' critical assumption that their rural averages contains no urbanization bias breaks down for their new dataset. As mentioned in Sections 4.2 and 4.3, if the rural averages contain urbanization bias, then NASA GISS' method will underestimate the magnitude of the urbanization bias in urbanized stations.

We also note that the Menne & Williams, 2009 homogeneity adjustments can also lead to the blending (as opposed to removal) of other non-climatic biases, if the biases occur with a high frequency. Fall et al., 2011 have found that about 70% of the weather stations in the U.S. component of the dataset are currently sited in poorly-exposed locations[66]. In a separate paper, we show that this poor exposure can introduce a warming bias into the station records, and that the Menne & Williams, 2009 homogeneity adjustments is inadequate for removing this bias[72].

It is likely that such biases are also a problem for the rest of the dataset.

NASA GISS' decision to switch in December 2011 to a homogeneity-adjusted dataset has probably reduced the extent of the problems described in Section 4.5.1. However, because of the problems with the homogeneity-adjusted dataset, they have replaced these problems with new ones. As a result, the new global temperature estimates are still unreliable.

5 Conclusions

In this article, the adjustments applied by NASA GISS to remove urbanization bias from their global temperature estimates were assessed. We found a number of serious problems with their adjustments:

- The vast majority of their adjustments involved correcting for "urban cooling", but urbanization bias is predominantly a warming bias.
- The net effect of their adjustments was unrealistically low, and tended towards zero for recent decades, despite this being a period during which urbanization increased globally.
- For a subset of some of the most highly urbanized stations, their adjustments succeeded in removing much of the urban warming for the period 1895-1980. But, for the more recent period, almost no adjustment was applied, even though urbanization continued to increase.

A number of serious flaws were found in the current approach:

- The use of their "extension rule" to extend the length of the urban records they could use in their global temperature estimates is inappropriate, because they include these extended periods of the urban records unadjusted.
- Their method for identifying stations as urban assumes that the co-ordinates they have for the stations are accurate, but quite a few of their station co-ordinates are inaccurate.
- Their method for identifying urban stations is not sensitive enough.
- As we discuss in Paper III, the currently available weather station datasets have a severe shortage of records for rural stations which are both long and complete[2]. Their method is unable to adequately handle such a shortage.
- Their method assumes that the only non-climatic biases which need to be considered are urbanization biases. As a result, up until December 2011, their adjustments were confounded by the presence of other non-climatic biases, leading to spurious and inappropriate adjustments.
- In December 2011, they switched to using a dataset which had already been homogeneity-adjusted, and so, presumably, this problem has been reduced. However, as we discuss in Paper III, the homogeneity-adjustments used for that dataset are inadequate and often lead to the "blending" of non-climatic biases between stations, rather than their removal[2]. So, switching to this dataset has introduced new problems.

In principle, we agree with NASA GISS's idea of attempting to adjust their data to remove urbanization bias, before estimating global temperature trends. Unfortunately, the approach they developed does not seem to work - at least with the current datasets. Aside from their unjustified "extension rule", and the problems with their current urbanization metric, the main problem seems to be that their adjustment algorithm is only designed for correcting one type of bias, and is not designed for multiple biases.

We should recognise that both step biases *and* trend biases exist in many of the temperature records currently being used for constructing global temperature estimates. Homogenization methods which attempt to correct step biases without correcting for trend biases are inadequate, as are those that attempt to correct trend biases without correcting for step biases. The homogeneity adjustments applied to the dataset NASA GISS have been using since December 2011 take the former approach, while NASA GISS' urbanization adjustments take the latter approach. Neither approach is adequate.

We note that when Menne & Williams, 2009 were developing the homogenization algorithm currently used for homogenizing the main dataset used by NASA GISS, they initially considered an algorithm which could identify combinations of both trend biases and step biases. However, they abandoned that approach for one in which they only remove step biases, as they believed it would be too hard to remove trend biases[25]. It might be worth revisiting this decision. Probably, future homogenization attempts should try to correct for both types of bias, and do so simultaneously.

As a final note, the recent claims that *"all 9 of the hottest years on record have occurred since 1998"*[48–

50] which were based on NASA GISS' global temperature estimates probably need to be reconsidered. Until more adequate attempts have been made to remove (or at least substantially reduce) the non-climatic biases from NASA GISS' global temperature estimates, they should be treated with considerable caution.

Acknowledgements

We would like to thank NASA GISS for having made their data, code and calculations public and accessible via their GISTEMP website from 2007 to 2011. Google's Google Earth application was very useful for investigating the surrounding environments of individual weather stations.

We acknowledge the data providers in the ECA&D project[69] for providing the Phoenix Park temperature data. Data and metadata for the ECA&D project is available at http://www.ecad.eu.

J.P. McGowan and Don Zieman provided some useful feedback on an early draft of this article.

No funding was received for this research.

References

[1] R. Connolly and M. Connolly. "Urbanization bias I. Is it a negligible problem for global temperature estimates?" 28 (Clim. Sci.). Ver. 0.1 (non peer reviewed draft). 2014. URL: http://oprj.net/articles/climate-science/28.

[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] 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.

[4] 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.

[5] 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.

[6] 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.

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

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

[9] R. Kunzig. "Population 7 billion". *Nat. Geo.* (January 2011). URL: http://ngm.nationalgeographic.com/2011/01/seven-billion/kunzig-text.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] 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/.

[19] J. Hansen et al. "Global temperature change". *Proc. Nat. Acad. Sci.* 103 (2006), pp. 14288–14293. DOI: 10.1073/pnas.0606291103.

[20] J. H. Lawrimore et al. "An overview of the Global Historical Climatology Network monthly mean temperature dataset, version 3". *J. Geophys. Res.* 116 (2011), p. D19121. DOI: 10.1029/2011JD016187.

[21] J. E. Hansen. "'Peak oil' paper revised and temperature analysis code". Online essay published on Dr. Hansen's website on 7 September 2008. URL: http://www.columbia.edu/~jeh1/mailings/2007/20070907_peakrevandgistemp.pdf.

[22] N. Barnes and D. Jones. "Clear Climate Code: Rewriting legacy science software for clarity". *Software, IEEE* 28 (2011), pp. 36–42. DOI: 10.1109/MS.2011.113.

[23] S. McIntyre. "Positive and negative urban adjustments". Climate Audit blog post. 1 March 2008. 2008. URL: http://climateaudit.org/2008/03/01/positive-and-negative-urban-adjustments/.

[24] T. R. Karl and C. N. Jr. Williams. "An approach to adjusting climatological time series for discontinuous inhomogeneities". *J. Clim. Appl. Meteor.* 26 (1987), pp. 1744–1763. DOI: 10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.

[25] M. J. Menne and C. N. Jr. Williams. "Homogenization of temperature series via pairwise comparisons". *J. Clim.* 22 (2009), pp. 1700–1717. DOI: 10.1175/2008JCLI2263.1.

[26] A. T. DeGaetano. "Attributes of several methods for detecting discontinuities in mean temperature series". *J. Clim.* 19 (2006), pp. 838–853. DOI: 10.1175/JCLI3662.1.

[27] I. E. Tereschenko and A. E. Filonov. "Air temperature fluctuations in Guadalajara, Mexico, from 1926 to 1994 in relation to urban growth". *Int. J. Clim.* 21 (2001), pp. 483–494. DOI: 10.1002/joc.602.

[28] 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.

[29] E. Johansson. "Influence of urban geometry on outdoor thermal comfort in a hot dry climate: A study of Fez, Morocco". *Build.*

Environ. 41 (2006), pp. 1326–1338. DOI: 10.1016/j.buildenv.2005.05.022.

[30] W. T. L. Chow et al. “Observing and modeling the nocturnal park cool island of an arid city: horizontal and vertical impacts”. *Theor. Appl. Clim.* 103 (2011), pp. 197–211. DOI: 10.1007/s00704-010-0293-8.

[31] N. H. Wong and C. Yu. “Study of green areas and urban heat island in a tropical city”. *Habitat Int.* 29 (2005), pp. 547–558. DOI: 10.1016/j.habitatint.2004.04.008.

[32] H. Akbari, M. Pomerantz, and H. Taha. “Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas”. *Sol. Energy* 70 (2001), pp. 295–310. DOI: 10.

[33] 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.

[34] 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.

[35] 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.

[36] G. S. Golany. “Urban design morphology and thermal performance”. *Atm. Environ.* 30 (1996), pp. 455–465. DOI: 10.1016/1352-2310(95)00266-9.

[37] 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.

[38] L. Doulos, M. Santamouris, and I. Livada. “Passive cooling of outdoor urban spaces. The role of materials”. *Sol. Energy* 77 (2004), pp. 213–249. DOI: 10.1016/j.solener.2004.04.005.

[39] E. G. McPherson. “Accounting for benefits and costs of urban greenspace”. *Landscape Urb. Plan.* 22 (1992), pp. 41–51. DOI: 10.1016/0169-2046(92)90006-L.

[40] R. Avissar. “Potential effects of vegetation on the urban thermal environment”. *Atm. Environ.* 30 (1996), pp. 431–448. DOI: 10.1016/1352-2310(95)00013-5.

[41] L. Shashua-Bar and M. E. Hoffman. “Vegetation as a climatic component in the design of an urban street: An empirical model for predicting the cooling effect of urban green areas with trees”. *Energy Build.* 31 (2000), pp. 221–235. DOI: 10.1016/S0378-7788(99)00018-3.

[42] D. J. Sailor. “Simulated urban climate response to modifications in surface albedo and vegetative cover”. *J. Appl. Meteor.* 34 (1995), pp. 1694–1704. DOI: 10.1175/1520-0450-34.7.1694.

[43] T. C. Peterson and R. S. Vose. “An overview of the Global Historical Climatology Network temperature database”. *Bull. Amer. Meteor. Soc.* 78 (1997), pp. 2837–2849. DOI: 10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2.

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

[45] 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.

[46] 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.

[47] 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.

[48] D. Satterfield. “NASA: 2012 was 9th warmest year on record. The 9 warmest years have all occurred since 1998”. AGU Blogosphere post on 16 January 2013. Accessed: 2013-02-07. (Archived by WebCite® at http://www.webcitation.org/6EG8qjgnT). URL: http://blogs.agu.org/wildwildscience/2013/01/16/nasa-2012-was-9th-warmest-year-on-record-the-9-warmest-years-have-all-occurred-since-1998/.

[49] T. Ghose. “NASA: 2012 was 9th hottest year on record worldwide”. LiveScience article posted 15 January 2013. Accessed: 2013-02-07. (Archived by WebCite® at http://www.webcitation.org/6EG96wrgX). 2013. URL: http://www.livescience.com/26277-nasa-2012-ninth-hottest-year.html.

[50] S. Goldenberg. “2012 among the 10 warmest years on record, figures show”. The Guardian, 16 January 2013. Accessed: 2013-02-07. (Archived by WebCite® at http://www.webcitation.org/6EG8fHVn0). URL: http://www.guardian.co.uk/environment/2013/jan/16/2012-10-warmest-years-on-record.

[51] R. Hoven. “NASA’s rubber rule”. American Thinker blog post. Accessed: 2013-02-10. (Archived by WebCite® at http://www.webcitation.org/6EKZgRTV5). 2012. URL: http://www.americanthinker.com/blog/2012/09/nasas_rubber_ruler.html.

[52] J. Hansen et al. “Climate impact of increasing atmospheric carbon dioxide”. *Science* 213 (1981), pp. 957–966. DOI: 10.1126/science.213.4511.957.

[53] 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.

[54] T. Ichinose, K. Shimodozono, and K. Hanaki. “Impact of anthropogenic heat on urban climate in Tokyo”. *Atm. Environ.* 33 (1999), pp. 3897–3909. DOI: 10.1016/S1352-2310(99)00132-6.

[55] H. Tran et al. “Assessment with satellite data of the urban heat island effects in Asian mega cities”. *Int. J. Appl. Earth Obs. Geoinfo.* 8 (2006), pp. 34–48. DOI: 10.1016/j.jag.2005.05.003.

[56] 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.

[57] 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.

[58] S. Yamashita. “Detailed structure of heat island phenomena from moving observations from electric tram-cars in Metropolitan Tokyo”. *Atm. Environ.* 30 (1996), pp. 429–435. DOI: 10.1016/1352-2310(95)00010-0.

[59] T. Yokobori and S. Ohta. “Effect of land cover on air temperatures involved in the development of an intra-urban heat island”. *Clim. Res.* 39 (2009), pp. 61–73. DOI: 10.3354/cr00800.

[60] 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.

[61] P. O’Neill. “Are you acquainted with any of these GHCN stations?” Peter O’Neill’s Blog. 20 August, 2011. URL: http://oneillp.wordpress.com/2011/08/20/are-you-acquainted-with-any-of-these-ghcn-stations/.

[62] C. D. Elvidge et al. “Night-time lights of the world: 1994-1995”. *ISPRS J. Photogramm. Rem. Sens.* 56 (2001), pp. 81–99. DOI: 10.1016/S0924-2716(01)00040-5.

[63] Wettersäulen in Europa. “Säntis Wetterstation (in German)”. Photographic history of the Säntis weather station. Accessed: 2013-02-08. (Archived by WebCite® at http://www.webcitation.org/6EHP6vrIE). URL: http://www.wettersaeulen-in-europa.de/direct.htm?/saentis/saentis.htm.

[64] M. Begert, T. Schlegel, and W. Kirchhofer. “Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000”. *Int. J. Climatol.* 25 (2005), pp. 65–80. DOI: 10.1002/joc.1118.

[65] T. R. Karl et al. “A model to estimate time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States”. *J. Clim. Appl. Meteor.* 25 (1986), pp. 145–160. DOI: 10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2.

[66] 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.

[67] K. G. Hubbard and X. Lin. "Reexamination of instrument change effects in the U.S. Historical Climatology Network". *Geophys. Res. Lett.* 33 (2006), p. L15710. DOI: 10.1029/2006GL027069.

[68] R. C. Hale et al. "Land use/land cover change effects on temperature trends at U.S. Climate Normals stations". *Geophys. Res. Lett.* 33 (2006), p. L11703. DOI: 10.1029/2006GL026358.

[69] A. M. G. Klein Tank et al. "Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment". *Int. J. Climatol.* 22 (2002), pp. 1441–1453. DOI: 10.1002/joc.773.

[70] J. Sweeney. "A three-century storm climatology for Dublin 1715-2000". *Irish Geograph.* 33 (2000), pp. 1–14. URL: http://geographicalsocietyireland.ie/gsi_main/.

[71] R. Heino. "Homogeneity of the long-term urban data records". *Atm. Environ.* 33 (1999), pp. 3879–3883. DOI: 10.1016/S1352-2310(99)00130-2.

[72] 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.