Quantifying the Effect of Urbanization on
U.S. Historical Climatology Network Temperature Records

Zeke Hausfather[1*], Matthew J. Menne[2], Claude N. Williams[2], Troy Masters[3],
Ronald Broberg[4], and David Jones[5]

[1]C3 Energy
[2]NOAA/National Climatic Data Center
[3]University of California at Los Angeles
[4]University of Colorado, Colorado Springs
[5]Climate Code Foundation

Corresponding Author:

Zeke Hausfather

hausfath@gmail.com

Abstract

An assessment quantifying the impact of urbanization on temperature trends from the U.S. Historical Climatology Network (USHCN) is described. Stations were first classified as urban and non-urban (rural) using four different proxy measures of urbanity. Trends from the two station types were then compared using a pairing method that controls for differences in instrument type and via spatial gridding to account for the uneven distribution of stations. The comparisons reveal systematic differences between the raw (unadjusted) urban and rural temperature trends throughout the USHCN period of record according to all four urban classifications. Based on these classifications, urbanization accounts for 14% to 21% of the rise in unadjusted minimum temperatures since 1895 and 6% to 9% since 1960. The USHCN-Version 2 homogenization process effectively removes this urban signal such that it becomes insignificant during the last 50-80 years. In contrast, prior to 1930, only about half of the urban signal is removed. Accordingly, the NASA Goddard Institute for Space Studies urban-correction procedure has essentially no impact on USHCN version 2 trends since 1930, but effectively removes the residual urban-rural temperature trend differences for years before 1930 according to all four urban proxy classifications. Finally, an evaluation of the homogenization of USHCN temperature series using subsets of rural-only and urban-only reference series from the larger U.S. Cooperative Observer (Coop) Network suggests that the composition of Coop stations surrounding USHCN stations is sufficiently "rural" to limit the aliasing of urban heat island signals onto USHCN-Version 2 temperature trends during homogenization.

1. Introduction

Urbanization has long been recognized as having the potential to impact near-surface temperature readings by altering the sensible and latent heat fluxes in affected areas [e.g., *Mitchell*, 1953; *Oke*, 1982; *Arnfield*, 2003]. The concentration of high thermal mass impermeable surfaces in urbanized regions commonly leads to higher surface temperatures compared to those in less developed or rural areas, especially at night [*Oke*, 1982; *Parker*, 2010]. To mitigate the potential for an urban bias in temperature records used for climate monitoring, stations that comprise the U.S. Historical Climatology Network (USHCN) were selected to be largely from rural or small town locations [*Quinlan et al.*, 1987; *Menne et al.*, 2009]. Still, station locations tend to be correlated with inhabited areas. Relative to the percentage of total land area that is built up, "urban" observation stations are likely overrepresented in general, even in networks like the USHCN.

Given the potential for urban biases, a number of studies have been undertaken to quantify the impact of the "urban heat island" (UHI) signal on land surface air temperature trends globally (e.g., *Peterson et al.* 1999; *Parker*, 2006; *Jones et al.*, 2008; *Hansen et al.*, 2010) and regionally within the USA (e.g*., Kukla et al.*, 1986; *Karl et al.*, 1988; *Gallo et al.*, 1999, *Gallo and Owen*, 2002; *Peterson*, 2003; *Peterson and Owen*, 2005). Unfortunately, quantifying the impact of urbanization on temperature trends faces multiple confounding factors. For example, an instrument originally installed in an urban environment may well have warmer absolute temperatures than one in a nearby rural area, *ceteris paribus*, but will not necessarily show a higher trend over time unless the composition of the city or the

microclimate around the sensor changes in such a way to cause the city observations to further diverge from temperatures at nearby rural locations [*Jones and Lister*, 2010], or the nature of urban land use leads to an amplifying of warm events whose frequency may change with time [*McCarthy et al.,* 2010]. It follows that urban heat island effects will lead to larger temperature trends compared to rural areas only if UHI-related factors cause incremental increases over rural temperatures during the period over which the trend is calculated [*Boehm*, 1998]. Moreover, cooling biases can be introduced into the temperature record when stations move from city centers to more rural areas on the urban periphery. This may have occurred, for example, during the period between about 1940 and 1960 when stations were moved from urban centers to newly constructed airports [*Hansen et al.*, 2001] and, in the case of the USHCN, airports, waste water treatment plants and other locations that lie outside the urban core [*National Climatic Data Center*, 2012]. Conversely, an instrument that is constructed in a relatively rural area that becomes more urban over time may exhibit a warming bias, and stations in small towns are not necessarily free of urban influences.

To further complicate matters, changes associated with urbanization may have impacts that affect both the meso-scale (10^2-10^4 meters) and the micro-scale (10^0-10^2 meters) signals. Small station moves (e.g., closer to nearby parking lots/buildings or to an area that favors cold air drainage) as well as local changes such as the growth of or removal of trees near the sensor may overwhelm any background UHI signal at the meso-scale [*Boehm*, 1998]. Notably, when stations are located in park-like settings within a city, the microclimate of the park can be

isolated from the urban heat island "bubble" of surrounding built up areas [*Spronken-Smith and Oke,* 1998; *Peterson*, 2003]. Further, changes in observation practice such as time of observation and instrument changes may lead to artifacts (inhomogeneities) in the data record that complicate the quantification of urban heat island signals [*Peterson*, 2003], especially if these changes are correlated with urban form.

Here, an analysis is described whose aim is to quantify the potential UHI contribution to U.S. temperature trends by more fully controlling for external factors that impact the trends but are otherwise unrelated to urbanization. A range of estimates for the UHI contribution to average U.S. temperature trends is provided by making use of four separate ways to differentiate urban and rural station environments to help assess uncertainty associated with identifying urban environments. The impact of data homogenization on the UHI signal is also evaluated. Homogenization is necessary to account for shifts in the station-based data caused by historical changes in the circumstances behind surface temperature measurement (e.g., changes in instrument type, station relocations) rather than by true changes in the climate. The artifacts caused by these kinds of changes have large, systematic impacts on U.S. temperature trends [*Menne et al.* 2009; *Williams et al.* 2012]. Consequently, homogenized datasets are essential for evaluating temperature changes from the observational record [*Venema et al.* 2012; *Lawrimore et al.*, 2011; *Hansen et al.* 2010; *Vose et al.*, in press]. Benchmarking the approach to homogenizing the U.S. monthly temperature data has essentially reaffirmed

previous assessments regarding the nature and impact of these artifacts on USHCN temperature trends [*Williams et al.* 2012] .

Homogenization of the USHCN monthly version 2 temperature data does not specifically target changes associated with urbanization. Rather, the procedure used involves identifying and accounting for shifts in the monthly temperature series that appear to be unique to a specific station--the assumption being that a spatially isolated and sustain shift in a station series is caused by factors unrelated to background climate variations [*Menne et al.* 2010]. Given that UHI-related changes may manifest as highly localized shifts or creeping changes, the focus in this analysis is to determine to what extent homogenization is removing apparent, local urban influences on the USHCN temperature record. Because homogenization may be removing local shifts caused by land use changes at non-urban stations, the same methodology used here could be applied to evaluating the impact of other types of land use changes.

The paper is organized as follows. Some additional background and motivation for the study are provided in section 2. The datasets and methods are discussed in section 3. Results are presented in section 4. Conclusions are provided in section 5.

2. Background and motivation

Motivation for assessing urban influences on temperature trends comes largely from interest in quantifying the contribution of urbanization in overall temperature trends relative to other factors. To that end, measures of ambient

population [*Kukla et al.*, 1986] and satellite-derived nightlights [*Gallo et al.*, 1999] have been used to differentiate urban and rural environments. Using these measures, monthly temperatures from U.S. weather stations designated as urban have been found to have decadal trends as much as 0.12°C/decade higher than those classified as rural [*Kukla et al.* 1986]. Because differences of this magnitude represent a non-negligible fraction of the likely background climate change signal, *Karl et al.* [1988] developed a specific adjustment to control for the apparent contribution of the urban heat island signal in USHCN temperature data. After adjusting for shifts in the data associated with time of observation and other changes documented in station histories, the *Karl et al.* [1988] evaluation suggested that an additional urban bias was present in the USHCN average temperature of about 0.06°C during the period from 1900 to 1984. Essentially all of the bias was associated minimum temperatures in urban areas, which were about 0.13°C higher on average than rural areas; maximum temperatures appeared to have little urban bias.

The *Karl et al.* [1988] UHI correction was used to produce the USHCN (version 1) fully adjusted USHCN monthly temperature data until the release of version 2 [*Menne et al.* 2009]. As in version 1, the version 2 release includes bias adjustments for time of observation and other station history changes, but version 2 also includes adjustments for changes (inhomogeneities) that are not documented in digital station histories (roughly 50% of all changes). Providing adjustments for both documented and undocumented station changes reduced the overall magnitude of minimum temperature trends from USHCN stations more than the

fully adjusted version 1 temperatures even though version 1 contained the additional *Karl et al.* [1988] UHI adjustment. The reason for this may be that the more comprehensive homogenization in version 2 removes the impact of incremental, but previously unidentified step changes associated with meso and micro-scale urbanization factors, or, that signal arising from local UHI trend changes are sometimes aliased (i.e., inadvertently accounted for) onto estimates of the more comprehensive version 2 step-type adjustments [*Menne et al.* 2009]. In any case, the development of a method for identifying and adjusting undocumented shifts appeared to account for more than of the signal attributed to urban effects on minimum temperatures by *Karl et al.* [1988]. Thus, no separate UHI-specific correction was provided in USHCN version 2.

Another reason that the *Karl et al.* [1988] corrections were not used in version 2 is that they are monotonic functions of city population; that is, these adjustments always reduced minimum temperature trends based on the total population of the city. In contrast, *Hansen et al.* [1999, 2001, 2010] have used a nightlights-based method that forces urban (and "peri-urban") station trends to conform to surrounding rural trends in the NASA Goddard Institute for Space Studies (GISS) surface temperature analysis. In the process, the *Hansen et al.* approach actually increases the trend for about 40% of urban stations. The fact that so many urban trends are larger after the urban adjustment likely reflects the degree to which the confounding factors discussed above can mitigate or otherwise obscure potential urban heat island signals.

For the U.S. data contribution to the NASA GISS analysis, *Hansen et al.* [2001, 2010] use the USHCN data that has been adjusted by NOAA/NCDC for time of observation and station history changes, but apply their own UHI adjustment. The GISS urban adjustment reduced the otherwise adjusted USHCN version 1 temperature trends by an additional 0.15°C/century, more than twice that of *Karl et al.* [1988] method [*Hansen et al.* 2001] even though the NASA GISS UHI corrections are not monotonic. Using the USHCN version 2 adjusted data, the impact of the GISS UHI correction is on the order of 0.07°C/century [*Hansen et al.* 2010].

The differential impacts of these approaches to assessing and correcting for the UHI are indicative of the need to better frame the uncertainty of urban influences on temperature trends in the U.S. As noted more recently by *Peterson* [2003] and *Peterson and Owen* [2005], this requires controlling for the many confounding issues like differences in instrumentation and other observation practices that may blur the urban signal. Whereas *Peterson* [2003] and *Peterson and Owen* [2005] focused primarily on a snapshot of mean urban-rural differences, here we build on their work by looking specifically at the time evolution of urban-rural differences. We use four rather than two proxy measures of urbanity and quantify the impact of data homogenization on the apparent UHI signal, focusing in particular on the potential magnitude of residual UHI contamination and whether there is evidence that homogenization transfers UHI bias from urban to non-urban station series.

3. Methods

The Conterminous United States (CONUS) has some of the most dense, publicly available digital surface temperature data in the world with over 7000 Cooperative Observer (Coop) Network Program stations reporting daily maximum and minimum temperature for at least 10 of the network's 120-plus year history. A subset of 1218 stations, generally those with long records, comprises the USHCN [*Menne et al.*, 2009]. This highly sampled surface temperature field allows for the comparison of subsets of station data in a manner that avoids inherent biases due to changes in spatial coverage. The Coop Program also now maintains accurate geolocational information on the present location of observing stations, with coordinates expressed in degrees, minutes and seconds (roughly 30 meter accuracy) available for most stations. This also allows for the accurate indexing of current Coop station locations against high-resolution georeferenced datasets that are useful to delineating urban and non-urban areas.

Because there is not an obvious meso-scale metric that determines the impact of urban form on temperature in all situations, we examined four different measures of urbanity that are available as georeferenced datasets: satellite-derived nightlights, urban boundary delineations, percent of impermeable surfaces, and historical population growth during the period where high-resolution data is available (1930 to 2000). These four measures, which represent different snapshots of urban boundaries, were used to classify a station as urban or non-urban by retrieving the pixel values coincident with the each station's coordinates. In cases where the proxy for urban form involved continuous measurements (all but urban boundaries), a cutoff point to divide stations between urban and rural was chosen

based on urban designations present in the literature (e.g. *Hansen et al.*, 2010 for

nightlights; *Elvidge et al.* 2007 for impermeable surface area). Each of these proxies

is described in section 3.a. below.

3.a. Datasets used to classify station types

Satellite Nightlights

Satellite-derived brightness values associated with the COOP Network

stations (including the USHCN) were taken from the Global Radiance Calibrated

Nighttime Lights dataset produced by the Earth Observation Group using

instruments flown on Defense Meteorological Satellite Program (DMSP) satellites.

We used the data from the F16 satellite recorded between 2005-11-28 and 2006-

12-24. The values we associate with each station are linearly interpolated from the

4 neighbor pixels in the image file and are converted to radiance by multiplying by

1.51586×10^{-10} giving a result in Watts sr^{-1} cm^{-2} [*Baugh et al.*, 2010]. To determine

a radiance value threshold for designating urban stations that is consistent with the

32 microWatts/m^2/sr/micrometer used in *Hansen et al.* [2010] (who used data from

Imhoff et al., 1997), we divided radiance values by the optical bandwidth of the F16

satellite (0.7 micrometers), resulting in a cutoff of 14.78 (i.e, $32 \div 0.7 \times 1.51586$) as

the equivalent value for the 2005-2006 satellite nightlight series. This is rounded to

the nearest integer (15) for the purpose of assigning a cutoff to separate urban from

non-urban pixels.

Urban Boundaries (GRUMP)

For the urban boundaries urbanity proxy, we use binary designations from the Global Rural-Urban Mapping Project (GRUMP), produced by the Center for International Earth Science Information Network (CIESIN) of the Earth Institute at Columbia University. GRUMP designations are based on the identification of urban areas using national census data (including the National Imagery and Mapping Agency database of populated places). GRUMP purports to identify cities and towns with populations exceeding 1,000 residents. Urban boundaries surrounding identified cities and towns are estimated based on DMSP Operational Linescan System (OLS) data from 1994-1995 as well as data from the Digital Chart of the World's Populated Places (DCW) [*Balk et al.*, 2004].

Impermeable Surfaces

The Impervious Surface Area (ISA) for pixels coincident with Coop stations is taken from the Global Distribution and Density of Constructed Impervious Surfaces dataset produced by the Earth Observation Group. The 1km resolution data used for this study was derived from 30-meter ISA data generated by the US Geological survey as described in *Elvidge et al.* [2007]. The data product has a nominal date of 2000-2001 and represents the percentage of the surface area that is comprised of manmade structures such as roads, buildings and parking lots. Station latitude and longitude were used to reference the dataset and extract the percentage of impervious surface in the surrounding 1 km. To determine the urban/non-urban classification a cut off of ten percent was employed. As noted by *Schuler* [1994] and *Arnold and Gibbons* [1996], the impacts to hydrology typically begin above this

figure. ISA values below ten percent were classified as rural. This approach is consistent with though somewhat more conservative than the recent work of *Potere et al.* [2009], who used a figure of twenty percent for detecting urban extent.

Population Growth

For the population growth proxy, we utilized Gridded 1 km Population Estimates for the Conterminous U.S., 1930-2000. This dataset was also used by *Peterson and Owen* [2005] and *Peterson* [2003] to classify USHCN stations into urban and non-urban categories. The gridded population was created using two U.S. Census Bureau data sets: The 2000 U.S. Census Bureau 1 km² population density grid for CONUS (*National Geophysical Data Center/NESDIS/NOAA*, 2002) and tabular U.S. Census county data [*U.S. Census Bureau*, 2002]. Urban sites were defined as those characterized by a 1930 to 2000 population growth of greater than or equal to 10 people per square kilometer, which yields similar sized numbers of urban and non-urban stations as shown in Table 1. While there is no available justification in the literature for this or any specific 1930-2000 population growth cutoff as a proxy for urbanization, this value was chosen to be reasonably conservative and to produce an urban/rural division generally in line with the other urbanity proxies. As Table 1 indicates, the GRUMP, Nightlights, and Population Growth urbanity proxies result in a relatively even distribution of stations in the rural and urban categories while the ISA proxies identifies the majority of stations as rural. Information on retrieving these datasets are provided as supplementary information.

Table 1: Number of USHCN stations classified by urbanity for each urbanity proxy.

Note that four stations could not be classified using the ISA urbanity proxy due to

dataset limitations.

Proxy Name	Urban Boundaries	Nightlights	ISA	Pop Growth
Rural Stations	608	594	857	685
Urban Stations	610	624	357	533

3.b. Calculation of rural and urban temperature trend differences

Urban-rural temperature differences were calculated by sub-setting the

USHCN station data according to the urban/non-urban station classifications

described above (for simplicity non-urban stations are referred to as rural). To

examine the possible UHI signal present in the USHCN temperature record, we use

two different but complimentary methods to compare urban and rural station

temperatures: station pairing and spatial gridding.

Station Pairing Method

The station pairing method creates pairs of nearby urban and non-urban

(rural) stations as classified by the four urban proxy measures. Pairs were created

by forming all possible permutations of urban and rural stations, excluding those

that were more than 161 kilometers (100 miles) apart; that had differing or

unknown sensor equipment types (e.g. Maximum Minimum Temperature Sensors [MMTS] versus Liquid in Glass Thermometers in Cotton Region Shelters [CRS]); or cases in which both stations currently have MMTS sensors but installation dates differ by more than 5 years. This pairing method yields a set of proximate urban/rural station pairs for each classification method that should be relatively unaffected by bias introduced through sensor-type transitions [*Quayle et al.*, 1991; *Menne et al.*, 2009]. Time series of monthly maximum and minimum temperature anomalies relative to a 1961-1990 baseline were calculated for all urban and rural series. Difference series for each urban and rural station pairings were then created for the full period of the USHCN version 2 records (1895 to present).

More specifically, the approach in the station pairing method was to take all permutations of urban and rural stations and produce a set containing unique pairs but non-unique occurrences of individual urban and rural station series (see Table 2). For example, a specific urban station would create distinct pairs with all surrounding rural stations within 100 kilometers with the same instrumentation type. To avoid overweighting regions with large numbers of adjacent urban and rural stations (and thus disproportionately more possible station pair combinations) we weight the urban-rural differences by the inverse of the number of stations pairs associated with each unique urban station. The mean urban-rural differences for unique urban stations are averaged for each month to obtain a best estimate of the underlying urban-rural temperature differences.

Table 2: Number of total urban/rural station pairs and unique urban stations by urbanity proxy.

Proxy Name	Urban Boundaries	Nightlights	ISA	Pop Growth
Total Station Pairs	1684	1809	1446	1392
Unique Urban Stations	437	470	271	390

The trend and confidence intervals for two periods, 1895-2010 and 1960-2010, are calculated from the station pair data using a weighted regression with clustered standard errors, with unique urban stations used for both the weighting and clustering. Standard errors are clustered by unique urban station because station pairs contain non-unique occurrences of individual urban and rural stations (e.g. a single urban station might be paired with four different rural stations), and treating each station pair as independent would result in erroneously narrow confidence intervals. As mentioned previously, each urban-rural pair is given a weight in the regression proportionate to the inverse of the number of station pairs that share the same urban station.

The station pairing method allows us to control for both spatial coverage and sensor type, avoiding potential complications introduced by differing locations of urban and rural stations as well as urban-correlated bias in the transition to MMTS sensors in the 1980s. The results will not necessarily be as representative of the

entire CONUS temperature field as those produced by spatial gridding, however, as station pairing does not explicitly weight based on spatial coverage.

Spatial Gridding Method

The spatial gridding method is used to create separate gridded fields for the conterminous U.S. using the subsets of urban and rural station series (and separately for maximum and minimum temperatures) as classified by each urban proxy measure. Station temperatures are converted to anomalies relative to a 1961-1990 baseline period, and station series that fall within 2.5° latitude x 3.5° longitude grid cells are averaged together and each grid cell average is cosine weighted to produce a CONUS average time series. The CONUS average urban and rural station series are then differenced. Trends and confidence intervals for the urban-rural differences during the 1895-2010 and 1960-2010 periods are calculated by regressing against the date using an AR(1) model to account for autocorrelation.

The gridding method described above is commonly used by NOAA/NCDC to produce spatially averaged time series for climate monitoring. In addition to this method, results using the gridding method described in *Menne et al.* [2009; 2010] are provided as supplementary information.

3.c. USHCN version 2 monthly temperature data

Urban-rural differences for mean monthly maximum and minimum temperatures were calculated using four different versions of the USHCN version 2 monthly temperature data. The four versions were used to help quantify the

magnitude of the UHI in the underlying raw (unhomogenized) data, to isolate the impact of data homogenization on the UHI signal, and to evaluate impact of the GISS UHI correction when applied as an addition correction over and above homogenization. The dataset versions include

1   time of observation-only adjusted data (called TOB);

2   adjusted version 2 (TOB + pairwise homogenization adjustments; v2);

3   adjusted version 2 data produced by running the pairwise homogenization algorithm using (a) neighboring series classified only as rural (v2-rural neigh); and, (b) neighboring series classified only as urban (v2-urban neigh).

4   adjusted version 2 data with the GISS UHI correction (TOB + pairwise homogenization + GISS UHI adjustments; v2+step2)

Each of these variants is described below.

Time of Observation Bias-Adjusted data (TOB)

The TOB station series are the raw monthly temperature data adjusted only for the time-of-observation bias [*Schaal and Dale,* 1977; *Karl et al.,* 1986]. The time of observation bias is an artifact of the starting/ending hour for the 24-hour interval over which the maximum and minimum temperature occurred. This bias is unrelated to any physical artifacts associated with urbanization and only leads to biased trends when the time of observation changes through time. However, such changes are likely more prevalent at rural stations, which are commonly run by volunteer observers who have been systematically transitioning from afternoon to

morning observation times [*Menne et al.*, 2009]. In order to remove the time of

observation bias as a confounding factor in assessing UHI impacts, we use data

adjusted according the method described by *Karl et al.* [1986] and *Vose et al.* [2003].

Results using completely unadjusted (raw) data are provided as supplementary

information using the *Menne et al.* [2009; 2010] gridding method.

Data adjusted by the Pairwise Homogenization Algorithm (USHCN version 2)

 Running the TOB-adjusted data through the Pairwise Homogenization

Algorithm (PHA; *Menne and Williams*, 2009) produces the USHCN version 2 fully

adjusted data [*Menne et al.*, 2009]. The PHA works by identifying and removing

abrupt shifts in monthly temperature series that appear to be unique to a particular

station. The shifts can be caused by small station moves, a change in

instrumentation, or, possibly, from the local impacts of any kind of land use change.

The shifts are identified via automated pairwise comparisons of monthly

temperature series in which the relative homogeneity of a given station's series is

evaluated by looking for breaks in differences series formed between the target

station and a number of highly correlated neighboring series. The adjustments are

based on the median shift magnitude calculated from pairwise temperature

differences between the target and neighbors before and after the shift. For any

particular target adjustment, the neighbor pool is drawn from those that appear to

be homogeneous according to the PHA for a minimum period (24 months) before

and after the target shift. The PHA does not specifically target urban station changes.

Rather, the algorithm targets all shifts that appear to be unique to a particular

station. Removing these local signals at all stations (rural and urban alike) produces temperature trend fields that more accurately reflect the general background climate signal than the raw data.

For version 2, USHCN monthly temperatures were compared to sets of highly correlated neighboring series within the larger Coop Network. Details regarding the mechanics of the PHA and evaluations of the algorithm's efficiency can be found in *Menne and Williams* [2009] and *Williams et al.* [2012a]. Version 2.0 of the adjusted monthly data was released in 2008 based on PHA version "52d". Urban-rural differences in version 2.0 adjusted data are discussed below. An evaluation of the UHI signal in a new version of the dataset—termed version 2.5—is provided as supplementary information using the Menne et al. [2009, 2010] gridding method. Version 2.5 fully homogenized data are produced by algorithm version "52i", which contains some bug fixes relative to version 52d [Williams et al., 2012b].

To evaluate the potential for UHI bias to be transferred from urban Coop stations that may be used as neighbors in the homogenization of USHCN station records, we also ran the USHCN station series through the PHA using only Coop stations that were classified as rural in one case and using only stations classified as urban in the other according to the same set of four urban proxies.

Version 2 homogenized data with the additional NASA/GISS "GISTEMP" UHI correction

Finally, we apply the GISTEMP urban heat island adjustment (described *Hansen et al.*, 2010) to the version 2.0 series to see how it addresses any remaining

urban-related signal from the homogenized monthly temperature records. The GISTEMP UHI correction adjusts the trend of stations classified as urban or peri-urban to match the trend of a distance-weighted composite record made from nearby rural stations. An urban station is adjusted only if there are at least three nearby rural stations with values that overlap at least two-thirds of the urban station's period of record. Periods and urban stations that fail the rural station requirement are excluded from the GISS analysis. Rural stations are ideally selected to be within 500 km of the urban station, but in some cases could be as far as 1000 km away to meet the selection requirement. Note that in performing this adjustment only rural stations from USHCN have been used. This contrasts with the usual GISTEMP analysis which will use any suitable rural stations in GHCN, possibly including stations not in USHCN (such as in Canada and Mexico). Given the spatial density of stations in USHCN we expect any differences in adjustment to be minimal.

The scheme for identifying stations as urban has changed in the history of the GISTEMP analysis (see *Hansen et al.*, 1999; *Hansen et al.*, 2001; *Hansen et al.*, 2010); here we use nighttime radiances from the DMSP calibrated radiance product described earlier. The analysis was carried out using the ccc-gistemp software supplied by the Climate Code Foundation [*Barnes and Jones*, 2011]. The resulting version 2.0 series with the GISTEMP UHI correction should be essentially the same as the USHCN data used in NASA's GISTemp product, albeit with a slightly more up-to-date dataset used for determining nighttime brightness and separate application of the Step 2 (UHI correction) process to average monthly minimum and maximum data rather than applying it to the mean monthly data only.

This analysis described above produces estimates of urban-rural differences for each month from 1895-2010 for mean monthly minimum and maximum temperatures for the TOB, v2, v2+Step 2, and v2-rural neigh/v2-urban neigh variants for each of the four urbanity proxy via both station pairing and spatial gridding methods, resulting in 64 different distinct urban-rural differences for each month.

4. Results

4.1 Unhomogenized (TOB-Adjusted) Data

Figure 1, which summarizes the urban minus rural (urban-rural) trend differences for all data set versions, indicates that the USHCN unhomogenized (TOB-only adjusted) data contains significant urban warming signals ($p < 0.05$ for linear trend fit) over the period from 1895 to present in both the minimum and maximum temperatures according to nearly all urban classification and comparison methods (the exception being GRUMP and Nightlights maximum temperatures evaluated via spatial gridding).
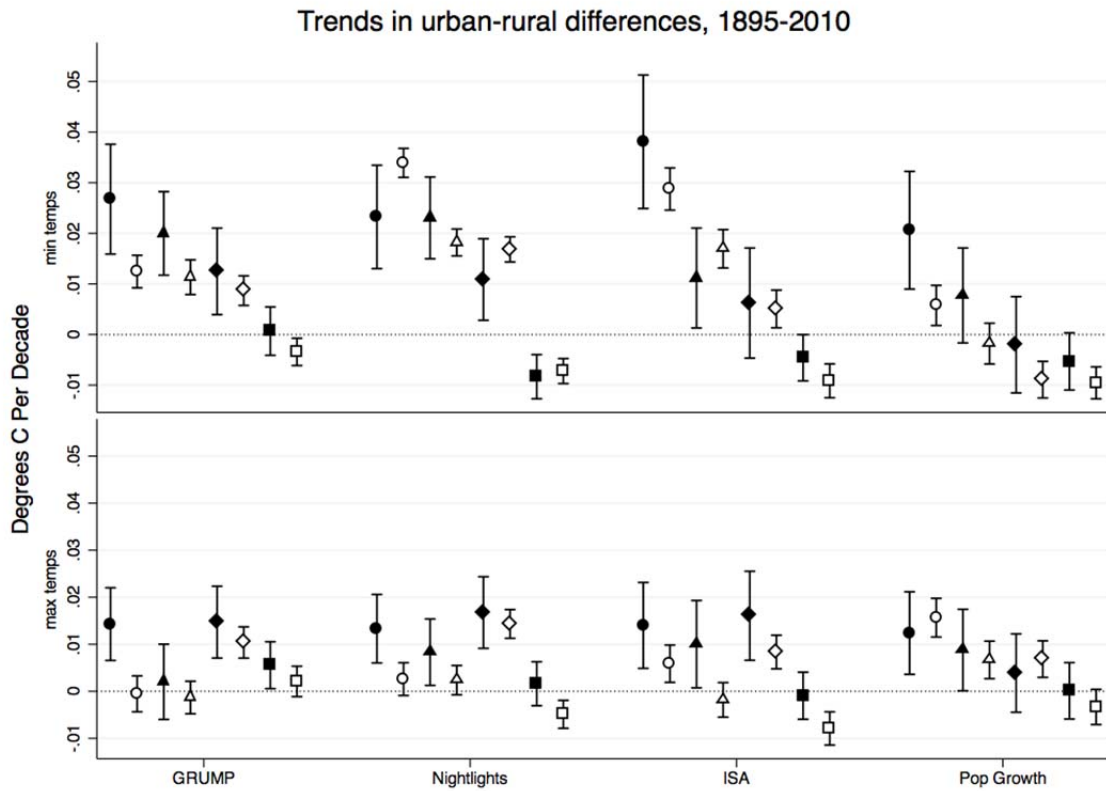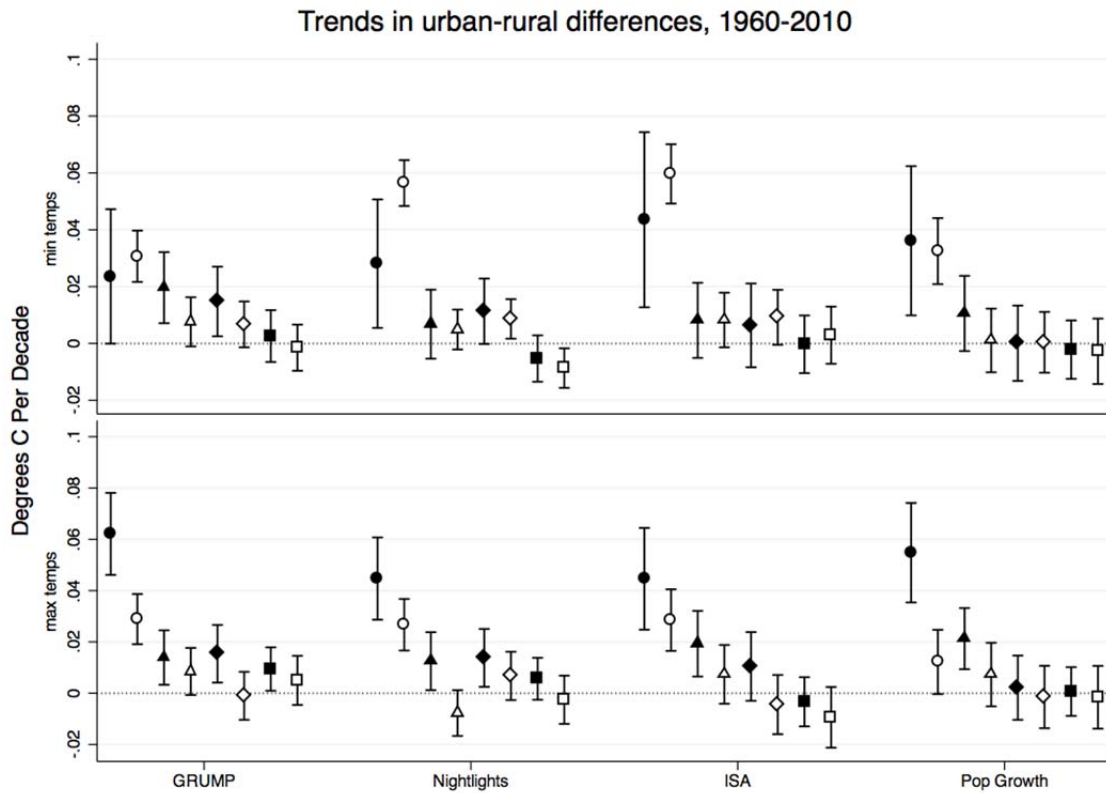
Figure 1: 1895-2010 trends and 95% confidence intervals in urban-rural differences by proxy type. Circles represent TOB adjusted data, Triangles represent version 2.0 data adjusted using rural neighbors only (v2-rural neigh), Diamonds represent version 2.0 homogenized data (v2), and Squares represent version 2.0 homogenized data with additional corrections using GISS's Step 2 method (Step 2). Solid shapes show results from the station pair method, and hollow shapes show results from the spatial gridding method.

As expected, the urban signal is larger in minimum temperatures than in maximum temperatures. Urban-rural difference trends in minimum temperature range between 0.05 and 0.5 C per century in minimum temperatures for the 1895-2010 period for the unhomogenized data depending on the classification and comparison method (e.g. station pairing or spatial gridding).

Figure 2: 1960-2010 trends and 95% confidence intervals in urban-rural differences by proxy type. Circles represent TOB adjusted data, Triangles represent version 2.0 data adjusted using rural neighbors only (v2-rural neigh), Diamonds represent version 2.0 homogenized data (v2), and Squares represent version 2.0 homogenized data with additional corrections using GISS's Step 2 method (Step 2). Solid shapes show results from the station pair method, and hollow shapes show results from the spatial gridding method.

As shown in Figure 2, there is also evidence of a significant urban signal in the unhomogenized data during the past 50 years, with urban-rural difference trends of between 0.2 and 0.6 C per century across all urbanity proxies for the period 1960-2010. This large urban warming signal does not appear to be a result of any correlation between instrument changes and urban form because it occurs with a similar magnitude in both the station pairing method (which controls for instrument type) and the spatial gridding method (which does not).
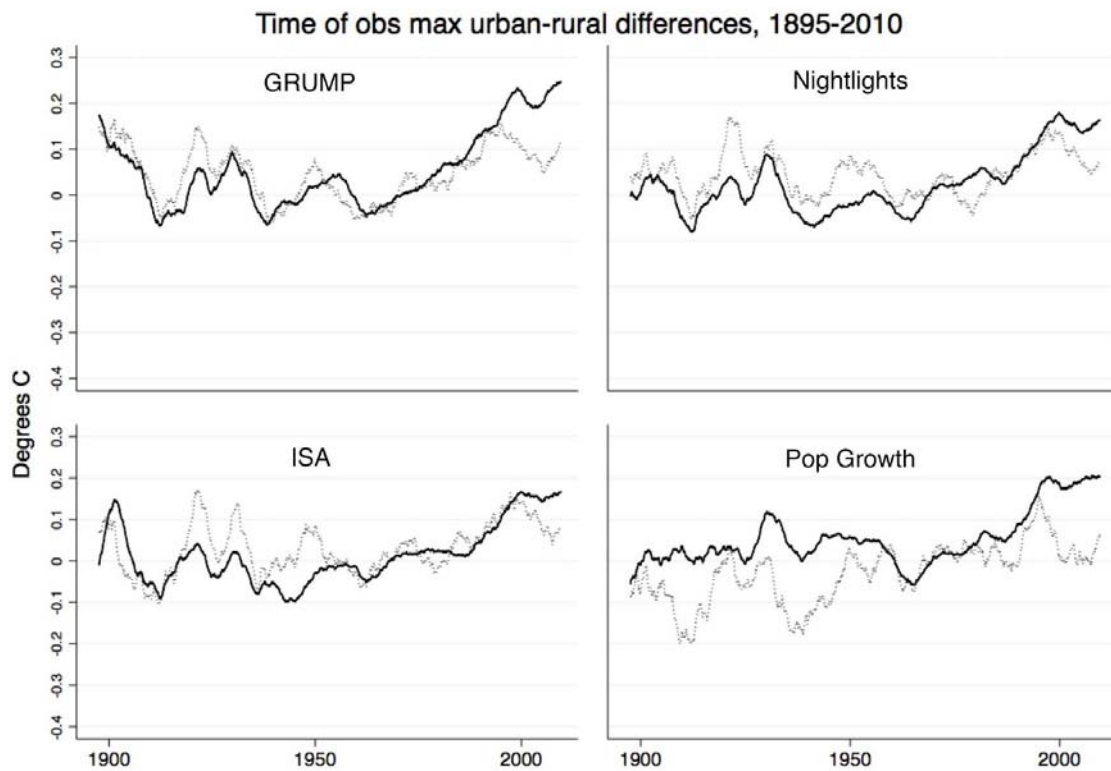
Time of obs min urban-rural differences, 1895-2010
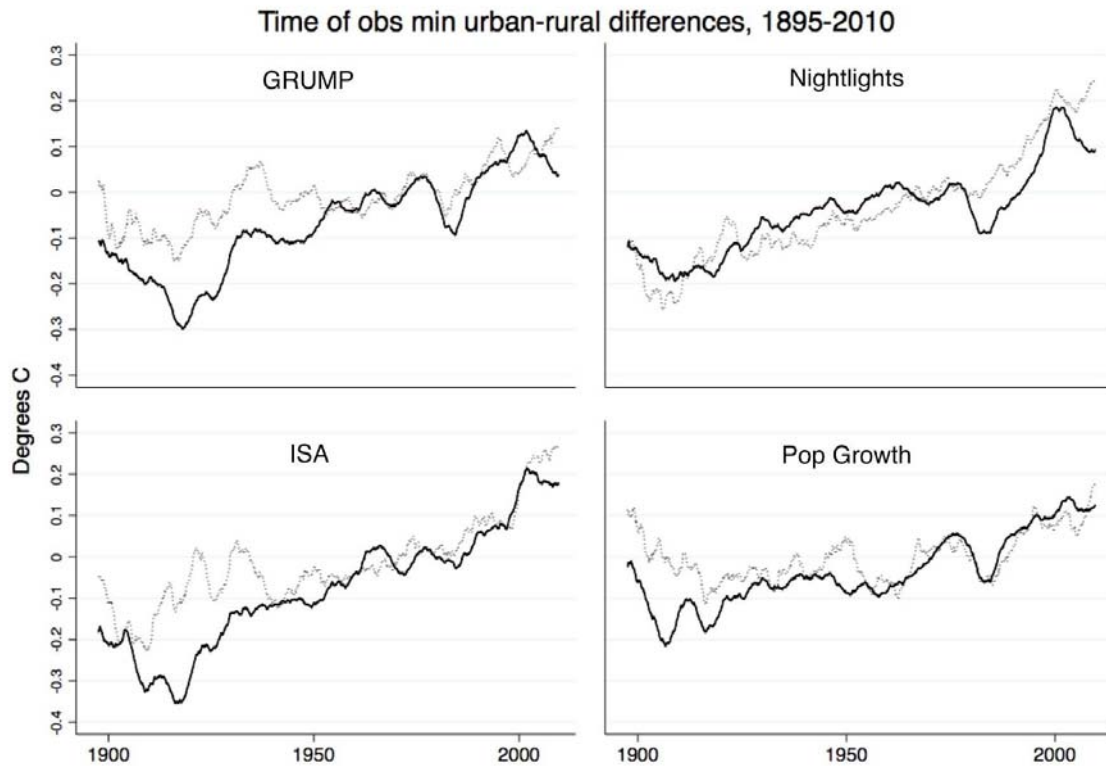
Figure 3: Running 5-year mean of urban and rural differences for time of observation-adjusted min USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.

For minimum temperatures, the urban warming signal over both century and half-century timeframes is larger in the more restrictive urban classification—ISA—that contains relatively few urban stations, and are smaller in the classifications—GRUMP, Nightlights, and Population Growth—that contain a more even split between urban and rural designations. The station pairing method often shows significantly larger urban warming than the spatial gridding method; however, the pairing method does not account for the potential biases related to the spatial distribution of the station pairs. As shown in Figure 3, the divergences between

station pairing and spatial gridding methods are particularly pronounced prior to 1950, which may be indicative of a larger geographic bias to the station pairs during that period. On the other hand, both methods produce similar results for periods after 1950.

As supplementary Figure SI.1 shows, the rural-urban differences are even larger in the raw minimum temperatures than in the TOB-adjusted data especially for the period since 1950 when time-of-observation changes were prevalent. However, as mentioned above, this difference is not likely driven by any physical phenomena related to UHI. Rather it likely reflects a higher frequency of time of observation changes at non-urban stations.

Figure 4: Running 5-year mean of urban and rural differences for time of observation-

adjusted max USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.

Maximum temperature urban-minus-rural trends in the unhomogenized (TOB) data are also significantly larger than zero over the period 1895 to 2010 for most urban classifications, but are smaller than the trends in minimum temperature urban-rural differences. They also show considerably less variation across urbanity proxy, with urban warming trends of around 0.08 to 0.22 C per century for the station pairing method and -0.04 to 0.2 C per century for the spatial gridding method. However, maximum temperature urban-rural difference trends are larger over the period 1960 to 2010, particularly in the GRUMP and Population Growth proxies where they exceed minimum urban minus rural trends. In this case, there is also a greater divergence between analysis methods, with the station pairing method showing much larger urban warming than the spatial gridding method, which again, likely reflects a spatial bias caused by the non-uniform distribution of station pairs.

By comparing the trends of rural stations to those of all USHCN stations, we can use the spatial gridding method to get an estimate of the extent to which overall CONUS minimum temperature trends over the past century may have been driven by the urban warming signal (see Table SI.1). By this estimate, the unhomogenized minimum temperature data from rural USHCN stations yields trends that are between 14 and 21 percent smaller on average over the period 1895-2010 period

than the trends from the full USHCN network. This difference decreases to between about 6 and 9 percent during the last 50 years.

4.2 Homogenized Version 2 Data (v2)

The pairwise homogenization algorithm (PHA) significantly reduces the difference between urban and rural minimum temperature trends according to all analysis methods and station classifications. This is particularly true over the 1960-2010 period, where a the vast majority of the urbanity proxies and methods indicate no significant urban warming in the minimum data. Maximum temperatures are a bit more mixed, though most proxies and methods show no significant urban warming in the maximum data over the period. As shown in Figure 5, there is still a small but significant minimum urban warming prior to 1960 in all urbanity proxies except for Population Growth. The station pairing method suggests some residual urban signal before 1960, but this residual signal is small in the spatial gridding method for all proxies after 1930.
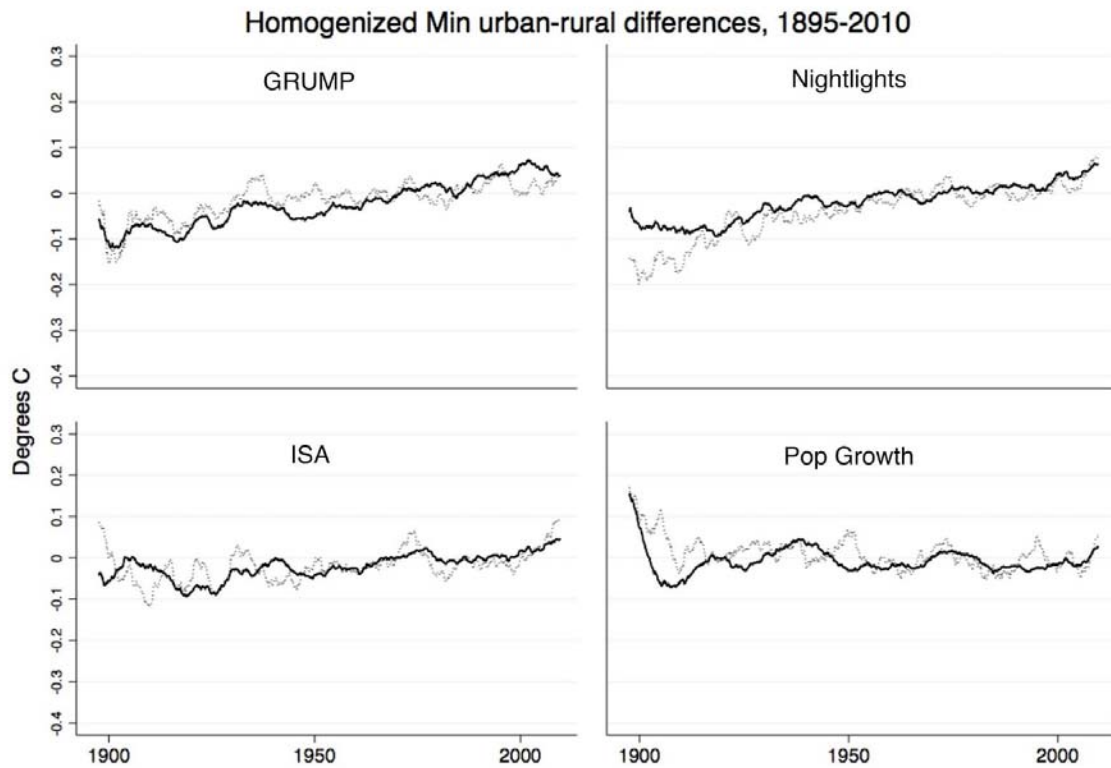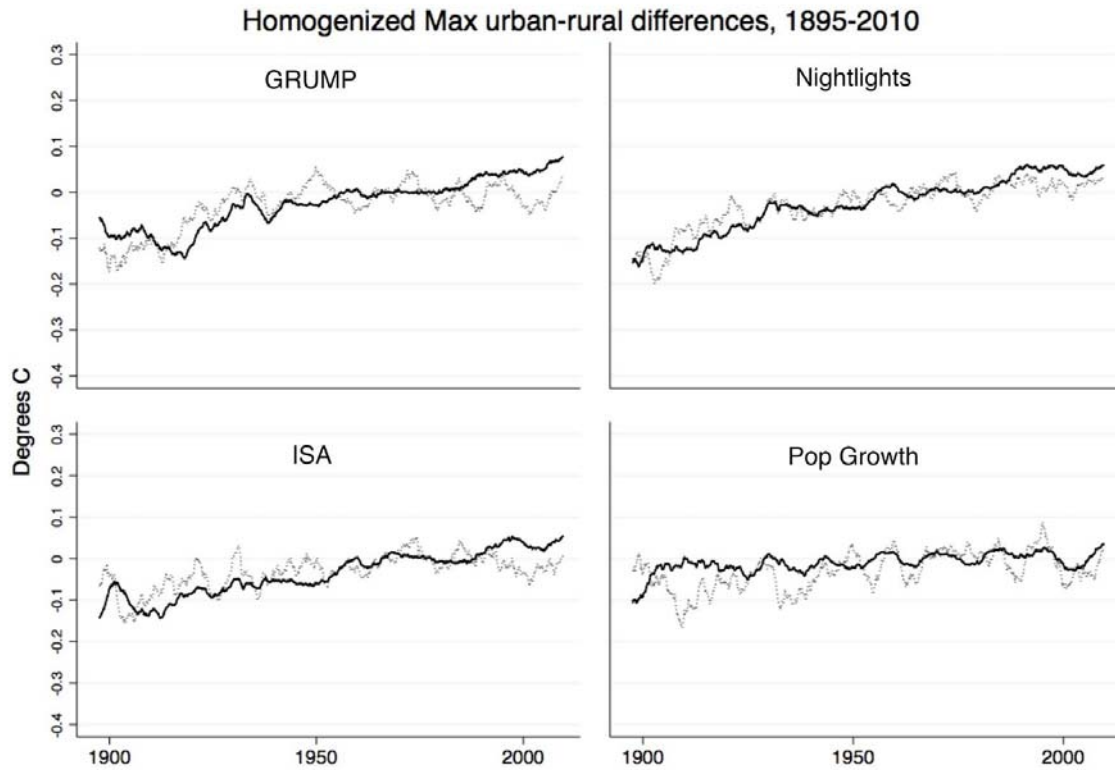
Figure 5: Running 5-year mean of urban and rural differences for v2 homogenized minimum temperature USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.

The effect of homogenization is most pronounced in the more restrictive urbanity proxies like ISA that contain relatively few urban stations and show larger urban warming trends prior to homogenization. The divergences between urban and rural temperatures that remain prior to 1930 even after homogenization are likely in part due to the combination or poorer metadata for that time period and fewer coop station records that can be used as neighbors.

Figure 6: Running 5-year mean of urban and rural differences for v2 homogenized max USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.

Notably, urban-rural differences in maximum temperatures over the century timeframe are in most cases not significantly reduced by homogenization, as shown in Figure 6.

Comparing homogenized rural HCN stations to all HCN stations, we find that rural stations show between 3 and 13 percent less average temperature (tave) warming over the 1895-2010 period, and a slight but not significantly different from zero reduction in warming over the 1960-2010 period (see Table SI.1). Thus, residual urban signals not removed by data homogenization appear to be significant

only for the period prior to 1960 and effectively only prior to about 1930. In summary, pairwise homogenization effectively removes the urban signal present in minimum temperature data from the last 50 to 80 years, and reduces it by around 50% or more for the period prior to 1930 (as can be seen when comparing Figure 3 and Figure 5).

4.3 Homogenized version 2 data with added GISTEMP correction (v2+step 2)

As reported in *Hansen et al.* [2010], applying the GISTEMP Step 2 UHI correction to the USHCN version 2 data has the impact of reducing the mean CONUS temperature trend from 0.73°C to 0.65°C over the period 1900-2009. As shown in Fig. SI.1, this reduction appears to result almost entirely from trend adjustments in the data for years prior to 1930. After 1930, the version 2.0 (52d) and version 2.5 (52i) data are not significantly impacted by the Step 2 adjustment. Moreover, this trend reduction is required only because of an urban signal in the early minimum temperature data, which get reduced by about 0.0113°C/decade by the Step 2 adjustment. The impact on maximum temperature is only 0.00288°C/decade. The average of these impacts is equivalent to the impact reported by *Hansen et al.* [2010]. As shown in Figures 7 and 8, the GISS Step 2 adjustment is effectively removing the residual urban signal in both minimum and maximum temperatures across all proxies without any significant over adjustment, even for the most restrictive definitions of urbanity.
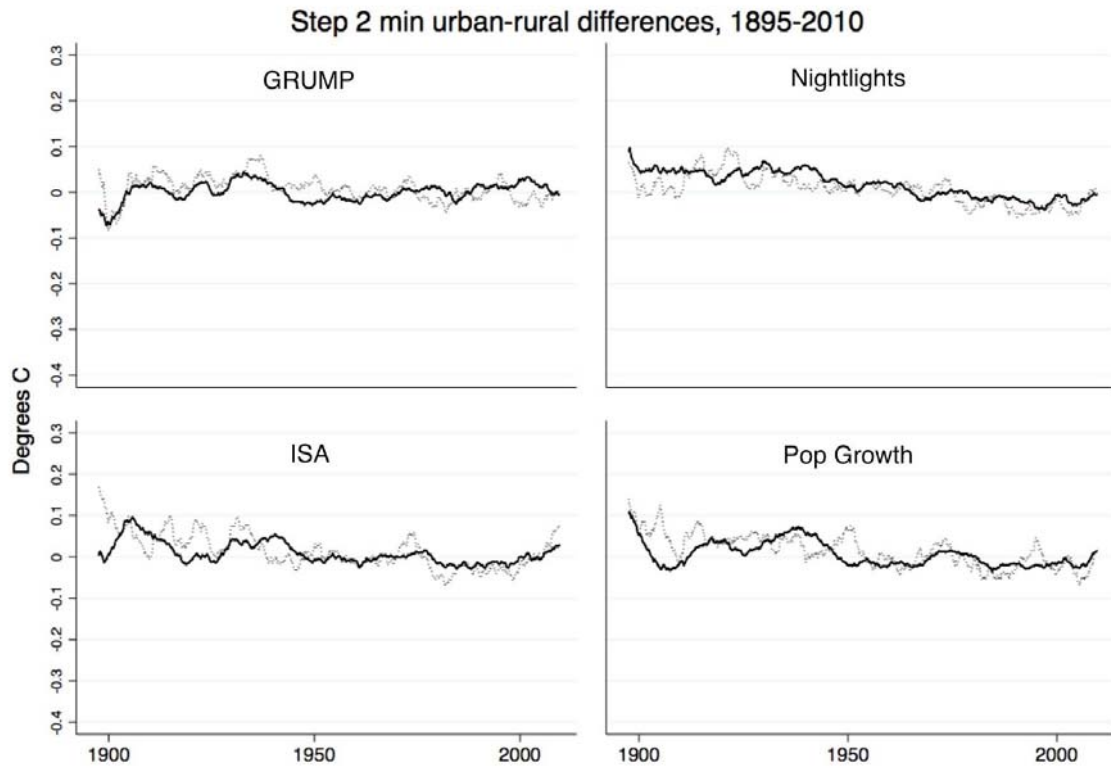
Figure 7: Running 5-year mean of urban and rural differences for Step 2 min USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.
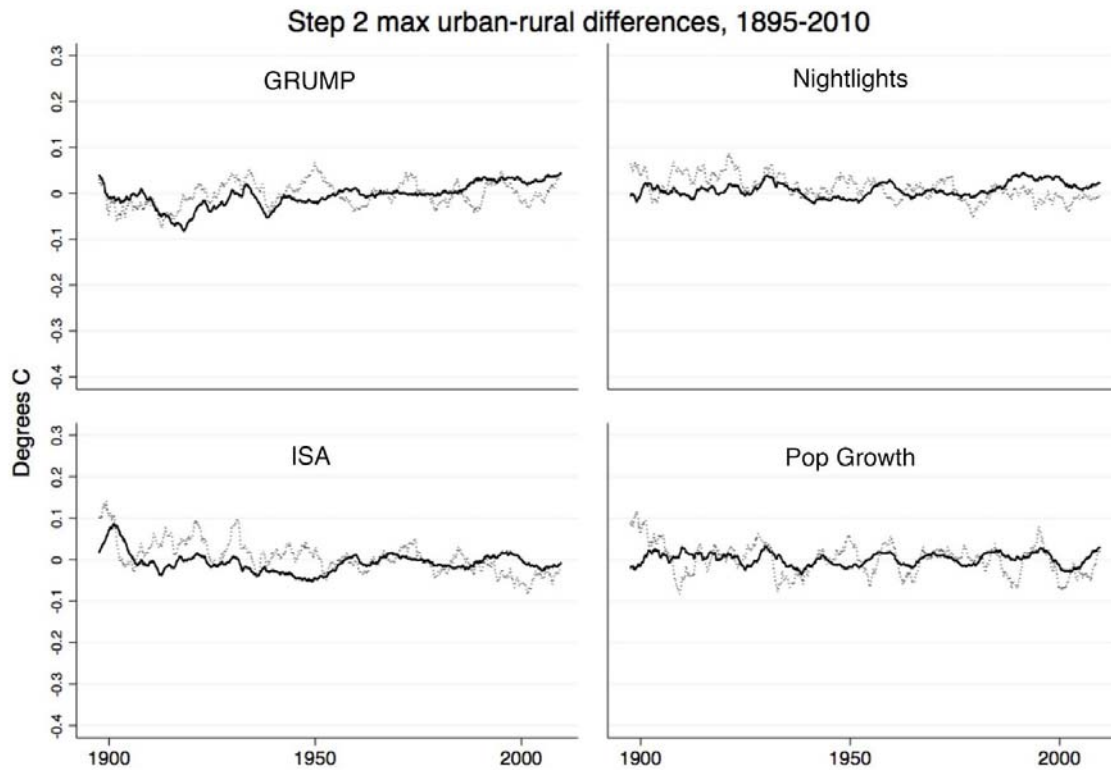
Figure 8: Running 5-year mean of urban and rural differences for Step 2 max USHCN station data from 1895 to 2010, using both station-pair (solid line) and spatial gridding (dashed line) methods for GRUMP, Nightlight, ISA (10%), and Population Growth urbanity proxies.

4.4 Homogenized version 2 data using only Coop neighbors classified as rural (v2-rural neigh)

In all of the urbanity proxies and analysis methods, the differences between urban and rural station minimum temperature trends are smaller in the homogenized data than in the unhomogenized data, which suggests that homogenization can remove much and perhaps nearly all (since 1930) of the urban signal without requiring a specific UHI correction. However, the trends in rural station minimum temperatures are slightly higher in the homogenized minimum temperature data than in the TOB-only adjusted data.   One possible reason for this

is that the PHA is appropriately removing inhomogenities caused by station moves or other changes to rural stations that have had a net negative impact on the CONUS average bias (e.g., many stations now classified as rural were less rural in the past since they moved from city centers to airports or waste water treatment plants). Another possibility is that homogenization is causing nearby UHI-affected stations to "correct" some rural station series in a way that transfers some of the urban warming bias to the temperature records from rural stations. In such a case, a comparison of the homogenized data between rural and urban stations would then show a decreased difference between the two by removing the *appearance* of an urbanization bias without actually removing the bias itself.

To help determine the relative merits of these two explanations, the PHA was run separately allowing only rural- and only urban-classified Coop stations to be used as neighbors in calculating the PHA corrections for USHCN stations. In Figure 9, the spatially averaged U.S minimum temperature anomalies for rural stations are shown for the four different datasets: the unhomogenized (TOB-adjusted only); the version 2 (all-Coop-adjusted; v2) data; the homogenized dataset adjusted using only coop stations classified as rural; and the homogenized dataset adjusted using only urban coop stations.
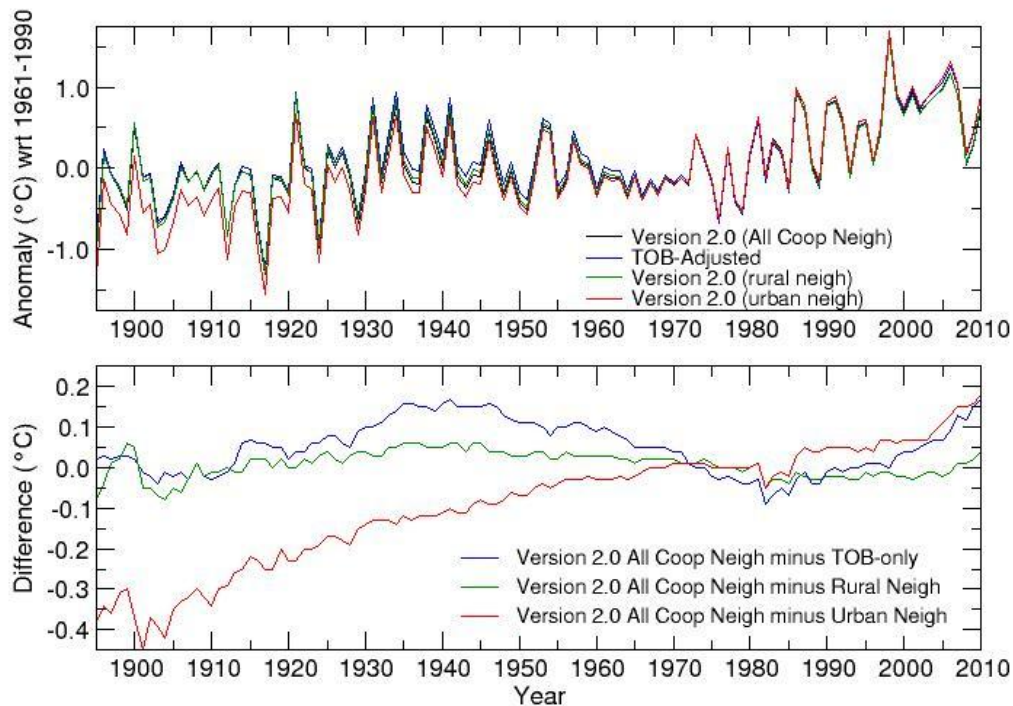
Figure 9: Comparison of spatially gridded minimum temperatures for the TOB-only adjusted USHCN data, v2 USHCN data (homogenized using all Coop station series as reference series), USHCN data homogenized using series from Coop stations only classified as rural according to the impervious surface method, and USHCN data homogenized using series from Coop stations only classified as urban (according to the impervious surface method). Top Panel: CONUS average anomalies for the four versions of the USHCN data. Bottom Panel: the differences between the USHCN v2 data homogenized with all Coop station series and: data adjusted only for the tob-bias (blue); data homogenized using only rural station series (green); and, data homogenized using only urban station series (red).

The large difference in the trends between the urban-only adjusted and the rural-only adjusted datasets suggests that when urban coop station series are used exclusively as reference series for the USHCN some of their urban-related biases can

be transferred to USHCN station series during homogenization. However, the fact that the homogenized all-coop-adjusted minimum temperatures are much closer to the rural-station-only adjustments than the urban-only adjustments suggests that the bleeding effect from the ISA classified urban stations is likely small in the USHCN version 2 dataset. This is presumably because there are a sufficient number of rural stations available for use as reference neighbors in the Coop network to allow for the identification and removal of UHI-related impacts on the USHCN temperature series. Furthermore, as the ISA classification shows the largest urban-rural difference in the TOB data, it is likely that greater differences between rural-station-only-adjusted and all-coop-adjusted series using stricter rural definitions result from fewer identified breakpoints due to less network coverage, and not UHI-related aliasing. Nevertheless, it is instructive to further examine the rural-only and urban-only adjustments to assess the consequences of using these two subsets of stations as neighbors in the PHA.

Figure S.I.2 shows the cumulative impact of the adjustments using the rural-only and urban-only stations as neighbors to the USHCN. In this example, the impermeable surface extent was used to classify the stations. The cumulative impacts are shown separately for adjustments that are common between the two runs (i.e., adjustments that the PHA identified for the same stations and dates) versus those that are unique to the two separate urban-only and rural-only reference series runs. In the case of both the common and unique adjustments, the urban-only neighbor PHA run produces adjustments that are systematically larger (more positive) than the rural-only neighbor run. The magnitude of the resultant

systematic bias for the adjustments common to both algorithm versions is shown in black. The reason for the systematic differences is likely that UHI trends or undetected positive step changes pervasive in the urban-only set of neighboring station series are being aliased onto the estimates of the necessary adjustments at USHCN stations. This aliasing from undetected urban biases becomes much more likely when all or most neighbors are characterized by such systematic errors.

Figure S.I.3. provides a similar comparison of the rural-only neighbor PHA run and the all-Coop (v2) neighbor run. In this case, the adjustments that are common to both the rural-only and the all-Coop neighbor runs have cumulative impacts that are nearly identical. This is evidence that, in most cases, the Coop neighbors that surround USHCN stations are sufficiently "rural" to prevent a transfer of undetected urban bias from the neighbors to the USHCN station series during the homogenization procedure. In the case of the adjustments that are unique to the separate runs, the cumulative impacts suggest that the less dense rural-only neighbors are missing some of the negative biases that occurred during the 1930 to 1950 period, which highlights the disadvantage of using a less dense station network. In fact, the all-Coop neighbor v2 dataset has about 30% more adjustments than the rural-only neighbor PHA run produces.   Results using the other three station classification approaches are similar and are provided as Figures S.I.3 – S.I.8.

5. Conclusions

According to all four proxy measures used to identify station environments that are currently urban, there is consistent evidence that urban stations have a systematic bias relative to rural stations throughout the USHCN period of record. This bias has led to an apparent urban warming signal in the unhomogenized data that accounts for approximately 14 to 21 percent of total rise in USHCN minimum temperatures averaged over the CONUS for the period since 1895, and 6 to 9 percent of the rise over the past 50 years. Homogenization of the monthly temperature data via NCDC's Pairwise Homogenization Algorithm (PHA) removes the majority of this apparent urban bias, especially over the last 50 to 80 years. Moreover, results from the PHA using the full set of Coop station series as reference series and using only those series from stations currently classified as rural are broadly consistent, which provides strong evidence that the reduction of the urban warming signal by homogenization is a consequence of the real elimination of an urban warming bias present in the raw data rather than a consequence of simply forcing agreement between urban and rural station trends through a spreading of the urban signal to series from nearby stations.

As noted in the introduction, one of the challenges in quantifying the UHI signal in land surface air temperature records is that changes affecting urban stations can occur at both the micro and meso-scales. Changes at the micro-scale (e.g., small station moves, growth of a tree) are not necessarily of interest in evaluations of the UHI signal because they are highly localized and may have no relevance to the broader land use changes associated with urbanization that can affect the mesoscale temperature signal. For this reason, micro-scale changes

reasonably can be included in the list of inhomogeneities that should be corrected for via homogenization (along with instrument changes and time of observation changes). In contrast, it may be desirable to preserve changes in the in the meso-scale signal because these changes encompass a broader footprint and are arguably more likely to be related to larger-scale land use changes. Unfortunately, it may not be possible to distinguish (at least automatically) changes occurring at the micro-scale from changes at the meso-scale, especially if only one station record is available to sample the meso-scale signal. Whatever the cause, when any station series exhibits a sustained change relative to highly correlated surrounding stations, the change is likely to be identified by the PHA as uniquely local, and its impact on that stations temperature trend will be removed with a bias adjustment. This happens whether the USHCN station is from a rural or urban environment, which means that the same challenge that exists for identifying UHI impacts also exists for identifying the impacts of other types of (non-urban) land use changes.

Nevertheless, the pairing of urban and rural stations in a manner that controls for instrument type and time of observation changes reveals larger trends at urban stations, which is consistent with the understanding that land use changes associated with urbanization lead to larger historic temperature trends at urban stations. However, that this larger trend signal is effectively removed through homogenization suggests that the urban environments characterized by larger trends do not have large spatial scales that allow them to be sampled by a number of Coop stations (or that the urban temperature signal is heterogeneous) and thus the local urban signal is being effectively removed via homogenization.

Because homogenization is largely successful in removing urban bias in the USHCN temperature data, it appears that only about 5% of the period-of-record USHCN version 2 minimum temperature trends across the CONUS can be attributed to local urban influences and, further, that most of this contribution is coming from data for years prior to 1930. This residual urban bias for the earlier years in the record may be a consequence of the reduced station density of the Coop network in the early part of the twentieth century, which limits the number of pairs available for detecting inhomogenities some of which may be related to urbanization.

NASA GISS's (GISTEMP) "Step 2" nightlight-based UHI adjustments effectively remove the remaining urban-rural differences during this early period, suggesting that the additional UHI-specific adjustment is achieving the goal of forcing agreement between urban and rural temperature trends. Nevertheless, the recently released USHCN version v2.5 data (homogenized with the PHA algorithm version "52i as shown in figure S.I.1) improves the pre-1930 period considerably vis-à-vis v2.0 (except in the case of GRUMP), which may also mean that homogenization procedures may be able to more fully account for urban-related biases in the future, at least in areas with sufficient station density. In any case, at present, the net effect of urban-correlated biases on the version 2.5 adjusted data is evidently small, accounting for less than 5% of the trend since 1895 (and between 0 and 2% since 1960).   While it would likely be worthwhile to further characterize the uncertainty in UHI-related warming in datasets like the USHCN (e.g., by exploring a range of cutoffs for classifying a station as urban with the various proxies or by quantifying more site-specific aspects of a stations environment), UHI

does not appear to represent a significant contributing factor in the CONUS-average temperature signal over the past 50-80 years.

References

Arnfield A.J. (2003), Two decades of urban climate research: a review of turbulence, exchanges of energy and water, and the urban heat island. *Int J Climatol.*, 23:1–26. DOI:10.1002/joc.859.

Arnold, C.L. and Gibbons, C.J. (1996), Impervious surface coverage. *American Planning Association Journal*, *62,* 243-258.

Balk, D., F. Pozzi, , G. Yetman, , U. Deichmann, and A. Nelson (2004), The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents. Working Paper, CIESIN, Columbia University. Palisades, NY, pp 31.

Barnes, N., and D. Jones (2011), Clear Climate Code: Rewriting legacy science code for clarity, *IEEE*, 28, 36-42.

Baugh, K., C. Elvidge, T. Ghosh, and D. Ziskin (2010), Development of a 2009 Stable

   Lights Product using DMSP-OLS data, Proceedings of the 30th Asia-Pacific

   Advanced Network Meeting, 114-130.

Boehm, R., (1998), Urban bias in temperature series—A case study for the city of

   Vienna, *Climate Change*, 38, 113–128.

Center for International Earth Science Information Network-CIESIN, Columbia

   University; International Food Policy Research Institute-IFPRI; The World

   Bank; and Centro Internacional de Agricultura Tropical-CIAT (2004), Global

   Rural-Urban Mapping Project (GRUMP), Alpha Version. Palisades, NY:

   Socioeconomic Data and Applications Center (SEDAC), Columbia University.

   Available at http://sedac.ciesin.columbia.edu/gpw.

DeGaetano, A.T. (2006), Attributes of several methods for detecting discontinuities

   in mean temperature series. *J. Climate,* **19,** 838–853.

Elvidge, C.D.,B.T. Tuttle, ,P.C. Sutton, , K.E. Baugh, A.T. Howard, C. Milesi, B.L. Bhaduri,

   R. Nemani (2007), Global Distribution and Density of Constructed

   Impervious Surfaces, *Sensors*, 7, 1962-1979.

Gallo, K.P., T.W. Owen, D.R. Easterling, and P.F. Jamason (1999), Temperature trends

   of the U.S. Historical Climatology Network based on satellite designated land

   use/land cover, *J. Climate*, 12, 1344-1348, doi: 10.1175/1520-

   0442(1999)012<1344:TTOTUS>2.0.CO;2.

Gallo, K.P. and T.W. Owen (2002), A sampling strategy for satellite sensor-based

   assessments of the urban heat-island bias, *Int. J. Remote Sensing*, 23, 1935-

   1939, doi: 10.1080/01431160110097259.

Karl, T.R, C. N. Williams Jr., P. J. Young, and W. M. Wendland (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States. *J. Climate Appl. Meteor.,* **25,** 145–160.

Karl, T.R., H.F. Diaz, and G. Kukla (1988), Urbanization: its detection and effect in the United States climate record, *J. Climate*, 1, 1099-1123.

Kukla, G., J. Gavin, and T.R. Karl, (1986), Urban Warming, *J. Clim. Applied. Meteor.*, 25, 1265-1270.

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

Hansen, J.E., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl (2001), A closer look at United States and global surface temperature change. *J. Geophys. Res.*, **106**, 23947-23963.

Hansen, J.E., R. Ruedy, M. Sato, and K. Lo (2010), Global Surface Temperature Change, *Rev. Geophys.*, 48, RG4004, doi:10.1029/2010RG000345.

Imhoff, M. L., W. T. Lawrence, D. C. Stutzer, and C. D. Elvidge (1997), A technique for using composite DMSP-OLS "city lights" satellite data to map urban area, *Remote Sens. Environ.*, 61, 361–370, doi:10.1016/S0034-4257(97)00046-1.

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

Jones, P.D. and D.H. Lister (2010), The Urban Heat Island in Central London and urban-related warming trends in Central London since 1900. *Weather*, 64, 323-327, DOI: 10.1002/wea.432.

Lawrimore, J.H., M.J. Menne, B.E. Gleason, C.N. Williams, D.B. Wuertz, R.S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network Monthly Mean Temperature Dataset, Version 3. *Journal of Geophysical Research-Atmospheres*, 116, D19121, doi:10.1029/2011JD016187.

McCarthy, M.P., M. J. Best, and R. A. Betts (2010), Climate change in cities due to global warming and urban effects. *Geophys. Res. Lett.*, 37, L09705, doi:10.1029/2010GL042845, 2010.

Menne, M.J., and C.N. Williams (2009), Homogenization of temperature series via pairwise comparisons, *Journal of Climate*, 22, 1700-1717.

Menne, M.J., C.N. Williams Jr., and R.S. Vose (2009), The United States Historical Climatology Network monthly temperature data–Version 2. *Bulletin of the American Meteorological Society*, 90, 993-1007.

Menne, M. J., C. N. Williams, Jr., and M. A. Palecki (2010), On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research*, 115, D11108, doi:10.1029/2009JD013094.

Mitchell, J. M. (1953), On the causes of instrumentally observed secular temperature trends, *J. Meteorol.*, 10, 244–261.

National Geophysical Data Center/NESDIS/NOAA (2002), Land Use Change Project Data: 2000 U.S. Census Population Grids (http://dmsp.ngdc.noaa.gov/html/download.html).

National Climatic Data Center/NOAA (2012), Historical Observing Metadata

   Repository (http://www.ncdc.noaa.gov/land-based-station-data/station-

   metadata)

Oke, T.R., (1982), The energetic basis of the urban heat island, *Quarterly Journal of*

   the Royal Meteorological Society, 108, 1–24.

Parker, D.E., (2006): A demonstration that large-scale warming is not urban. *J.*

   Clim., 19, 4179-4197, DOI:10.1175/JCLI3730.1.

Parker, D. E. (2010), Urban heat island effects on estimates of observed climate

   change, Wiley Interdiscip. Rev. Clim. Change, 1, 123–133,

   doi:10.1002/wcc.21.

Peterson, T.C., K.P. Gallo, J.H. Lawrimore, A. Huang, D.A. McKittrick (1999), Global

   rural temperature trends. *Geophys. Res. Lett.*, 26, 329-332, DOI:

   10.1029/1998GL900322.

Peterson, T.C. (2003), Assessment of Urban Versus Rural In Situ Surface

   Temperatures in the Contiguous United Stats: No Difference Found*. J.*

   Climate, 16:18, 2941-2959.

Peterson, T. C., and T.W. Owen (2005), Urban Heat Island Assessment: Metadata Are

   Important. J. Climate, 18:14, 2637-2646.

Potere, D., A. Schneider, S. Angel, and D.L. Civco, D.L (2009), Mapping urban areas on

   a global scale. *International Journal of Remote Sensing,* 30(24),6531-6558.

Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes (1991), Effects of recent

   thermometer changes in the Cooperative Station Network, *Bull. Amer.*

Meteorol. Soc., 72, 1718–1723, doi:10.1175/1520-

0477(1991)072<1718:EORTCI>2.0.CO;2.

Quinlan, F. T., T. R. Karl, and C. N. Williams Jr. (1987), United States Historical

Climatology Network (HCN) serial temperature and precipitation data. NDP-

019, Carbon Dioxide Information Analysis Center, Oak Ridge National

Laboratory, U.S. Department of Energy, Oak Ridge, TN.

Schueler, T.R., (1994), The importance of imperviousness*. Watershed Protection

Techniques*, *1(3),* 100-111.

Schaal, L. A., and R. F. Dale (1977), Time of observation bias and "climatic change," *J.

Appl. Meteorol.*, 16, 215–222, doi:10.1175/1520-

0450(1977)016<0215:TOOTBA>2.0.CO;2.

Spronken-Smith, R. A., and T. R. Oke (1998), The thermal regime of urban parks in

two cities with different summer climates. *Int. J. Remote Sens.,* **19,** 2085–

2104.

U.S. Census Bureau, 2002: Population of States and Counties of the United States:

1790-2000, Washington, D.C., 226 pp.

Venema, V. K. C., et al. (2012), Benchmarking monthly homogenization algorithms,

Clim. Past Discuss., 7, 2655–2718.

Vose, R.S., C.N. Williams, T.C. Peterson, T.R. Karl, and D.R. Easterling, 2003: An

evaluation of the time of observation bias adjustment in the US Historical

Climatology Network. *Geophysical Research Letters*, 30 (20), 2046,

doi:10.1029/2003GL018111.

Vose, R.S., D. Arndt, V. F. Banzon; D.R. Easterling; B.E. Gleason; B. Huang; E. Kearns;
J.H. Lawrimore; M.J. Menne; T.C. Peterson; R.W. Reynolds; T.M. Smith; C.N.
Williams; D.L. Wuertz, 2012: NOAA's Merged Land-Ocean Surface
Temperature Analysis. *Bulletin American Meteorological Society*, in press.

Williams, C.N., M.J. Menne, and P.W. Thorne (2012a), Benchmarking the
performance of pairwise homogenization of surface temperatures in the
United States. *J. Geophys. Res.*, 117, D5, doi:10.1029/2011JD016761.

Williams, C.N, M.J. Menne and J.H Lawrimore (2012b), Modifications to Pairwise
Homogeneity Adjustment software to address coding errors and improve
run-time efficiency. NOAA Technical Report NCDC No. GHCNM-12-02
(available at
http://www1.ncdc.noaa.gov/pub/data/ghcn/v3/techreports/Technical%20
Report%20NCDC%20No12-02-3.2.0-29Aug12.pdf).

Supplementary Information
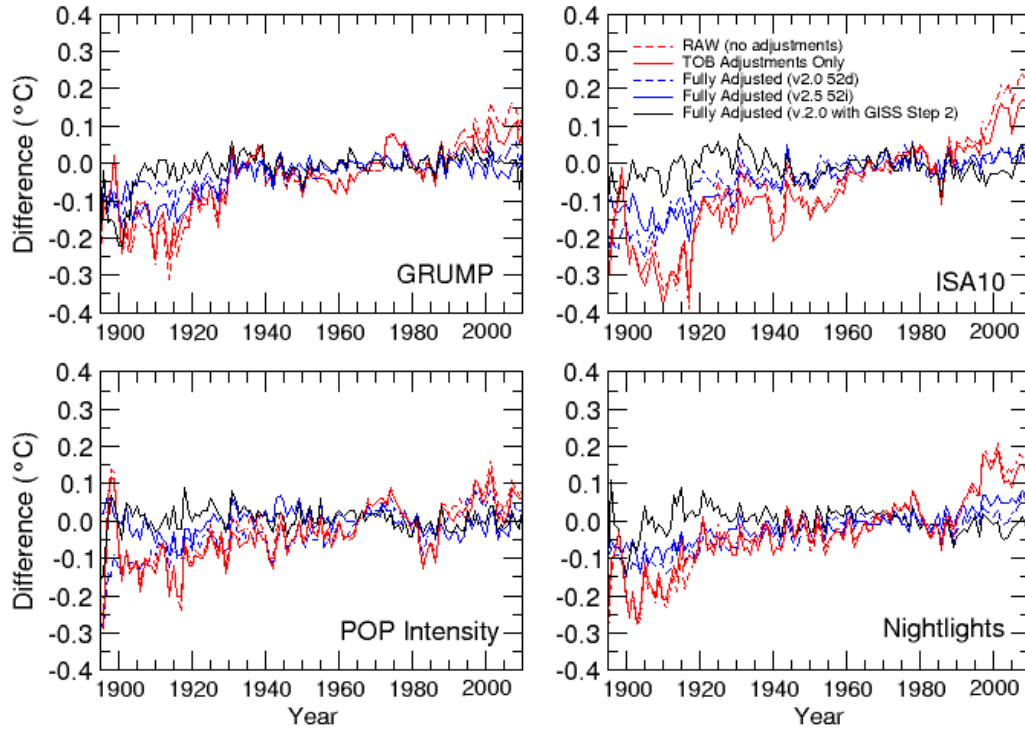
Figure S.I. 1. CONUS average urban and rural minimum temperature differences for five different versions of the USHCN station data and four different station classifications. The five different dataset versions are "Raw" (no bias adjustments-dashed red); TOB-only adjusted (solid red); v2 fully adjusted data homogenized using the pairwise algorithm version "52d" (dashed blue); fully adjusted v2 data homogenized using the pairwise algorithm version "52i" (solid blue); and fully adjusted v2 data homogenized using algorithm version "52d" with the NASA GISS (GISTEMP) "Step 2" UHI correction. CONUS averages were computed as described in *Menne et al.* [2009, 2010].

Table SI.1. Minimum and maximum trends in CONUS temperatures for specified sets of stations using the Menne et al [2009] spatial gridding method for time of observation-adjusted (TOB), homogenized v2.0 (52d), and homogenized v2.5 (52i) series.

| Stations	Dates	Series	TMIN Trend	TMAX Trend
All Stations	1895-2010	TOB	0.074	0.028
All Stations	1895-2010	52d	0.075	0.061
All Stations	1895-2010	52i	0.070	0.056
GRUMP Rural	1895-2010	TOB	0.060	0.027
GRUMP Rural	1895-2010	52d	0.068	0.053
GRUMP Rural	1895-2010	52i	0.068	0.056
ISA Rural	1895-2010	TOB	0.064	0.026
ISA Rural	1895-2010	52d	0.072	0.060
ISA Rural	1895-2010	52i	0.070	0.060
Nightlight Rural	1895-2010	TOB	0.062	0.025
Nightlight Rural	1895-2010	52d	0.069	0.056
Nightlight Rural	1895-2010	52i	0.068	0.056
Pop Growth Rural	1895-2010	TOB	0.064	0.025
Pop Growth Rural	1895-2010	52d	0.076	0.058
Pop Growth Rural	1895-2010	52i	0.071	0.059
All Stations	1960-2010	TOB	0.255	0.127
All Stations	1960-2010	52d	0.248	0.189
All Stations	1960-2010	52i	0.236	0.196
GRUMP Rural	1960-2010	TOB	0.234	0.106
GRUMP Rural	1960-2010	52d	0.242	0.184
GRUMP Rural	1960-2010	52i	0.237	0.190
ISA Rural	1960-2010	TOB	0.240	0.119
ISA Rural	1960-2010	52d	0.247	0.193
ISA Rural	1960-2010	52i	0.234	0.198
Nightlight Rural	1960-2010	TOB	0.233	0.113
Nightlight Rural	1960-2010	52d	0.244	0.185
Nightlight Rural	1960-2010	52i	0.234	0.194
Pop Growth Rural	1960-2010	TOB	0.236	0.120
Pop Growth Rural	1960-2010	52d	0.248	0.190
Pop Growth Rural	1960-2010	52i	0.236	0.193
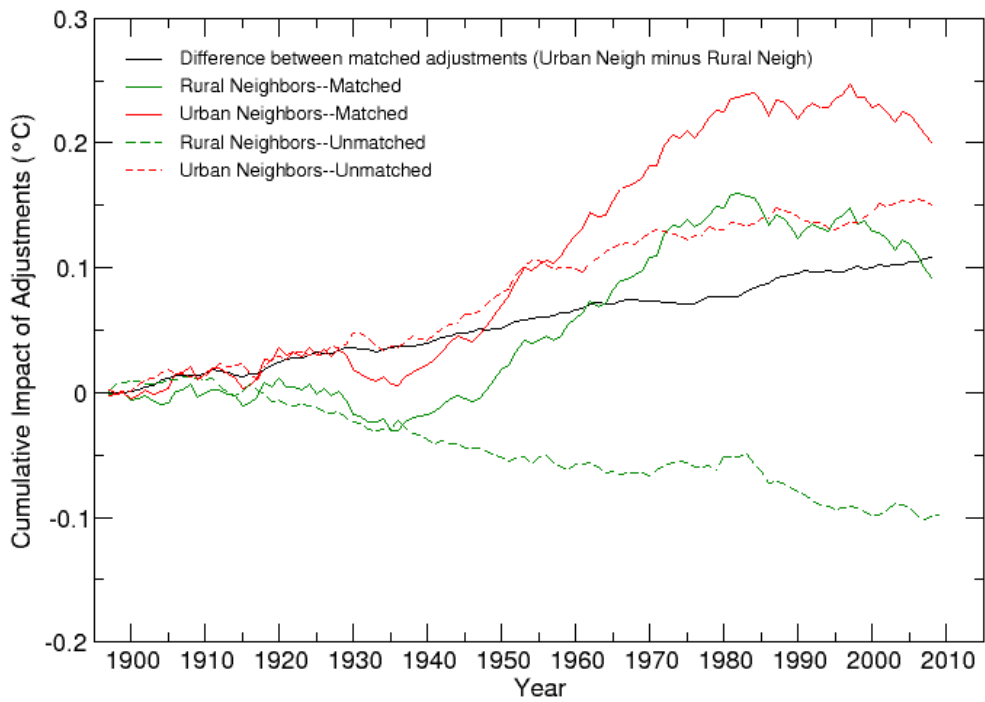
Figure S.I. 2. Cumulative average of PHA-derived minimum temperature adjustments using Coop station reference series classified as urban only (red lines) and as rural only (green lines) according to the impermeable surface area (ISA10) classification method. The cumulative average of the adjustments that are common to both datasets are shown as solid lines and those that are unique are shown as dashed lines.
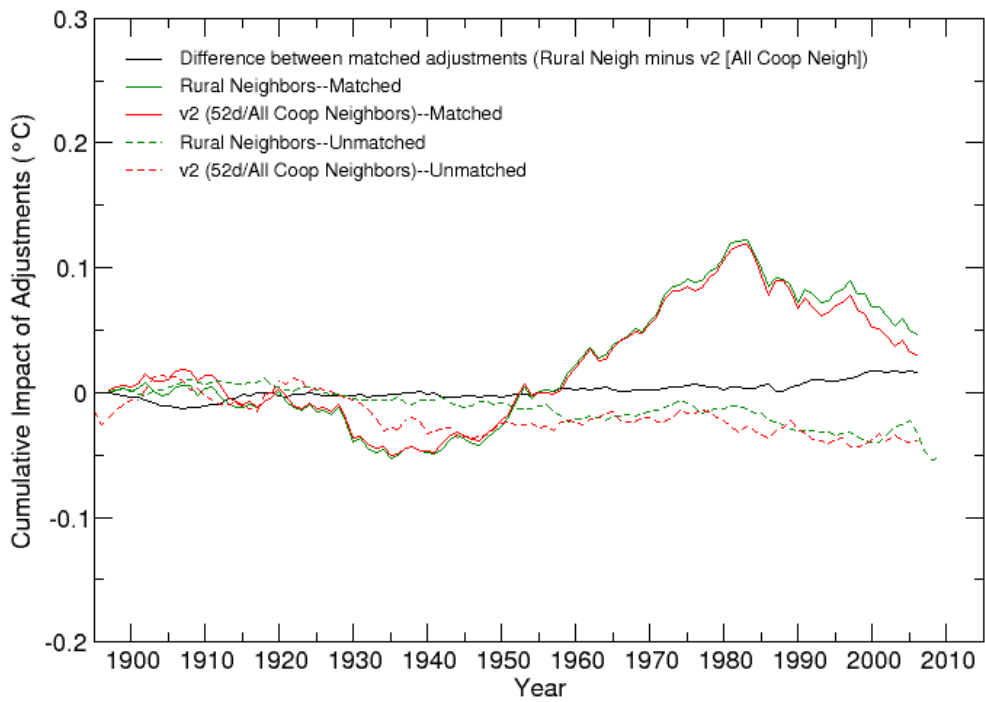
Figure S.I. 3. Cumulative average of PHA-derived minimum temperature adjustments using all Coop station series as reference series-v2-"52d" (red lines) and classified as rural only (green lines) according to the impermeable surface classification method. The cumulative average minimum temperature adjustments that are common to both datasets are shown as solid lines and those that are unique are shown as dashed lines.
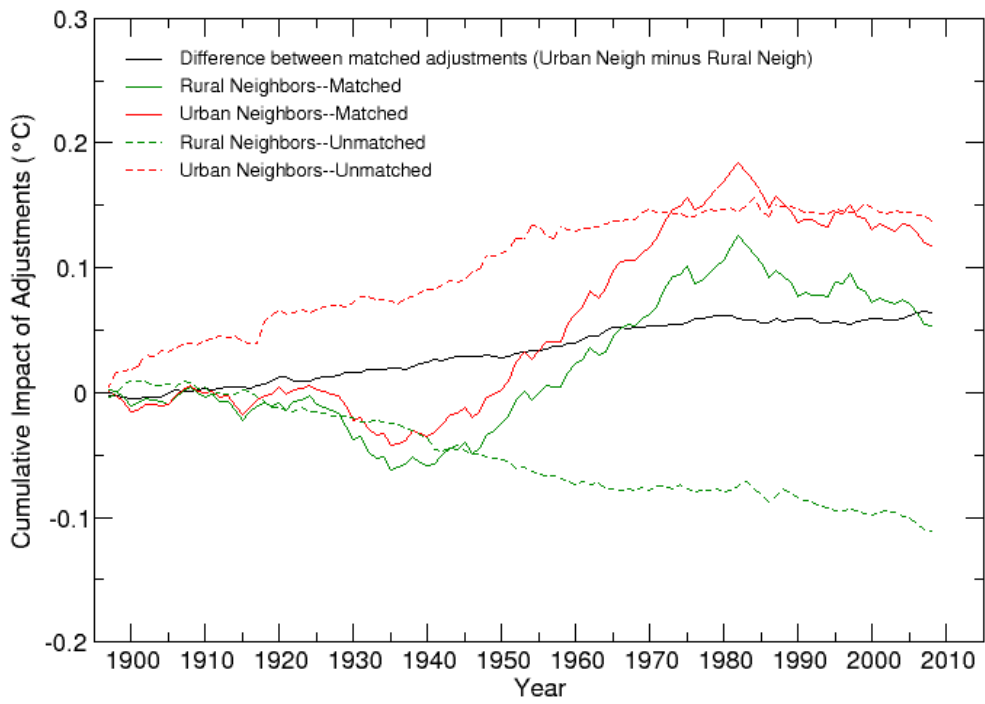
Figure S.I. 4. As in Fig. S.I.2 but from stations classified using GRUMP.
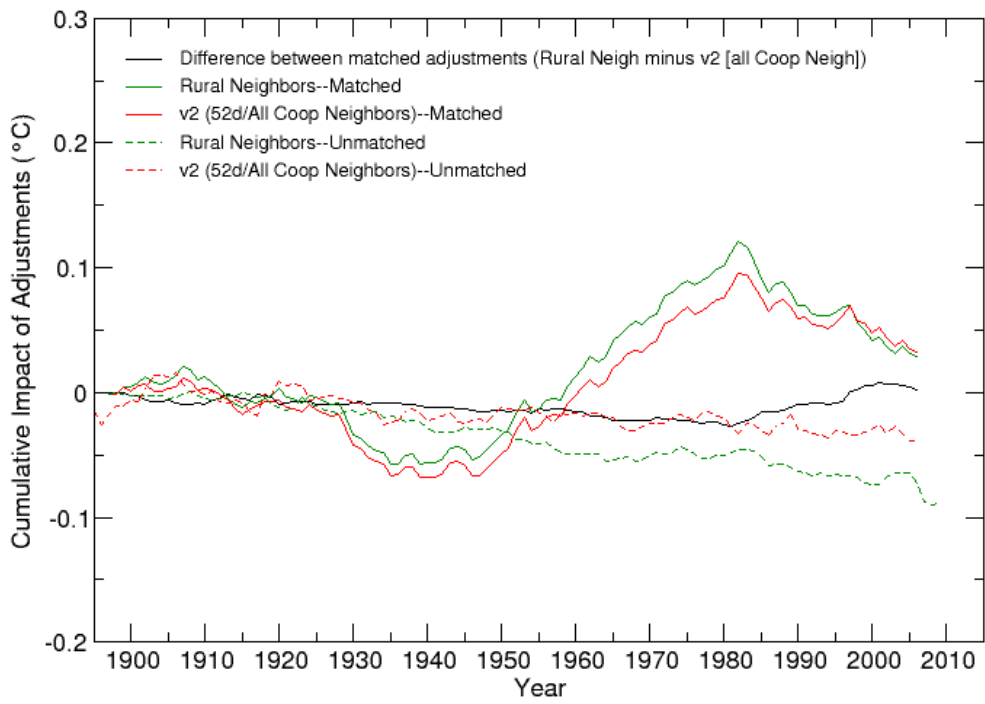
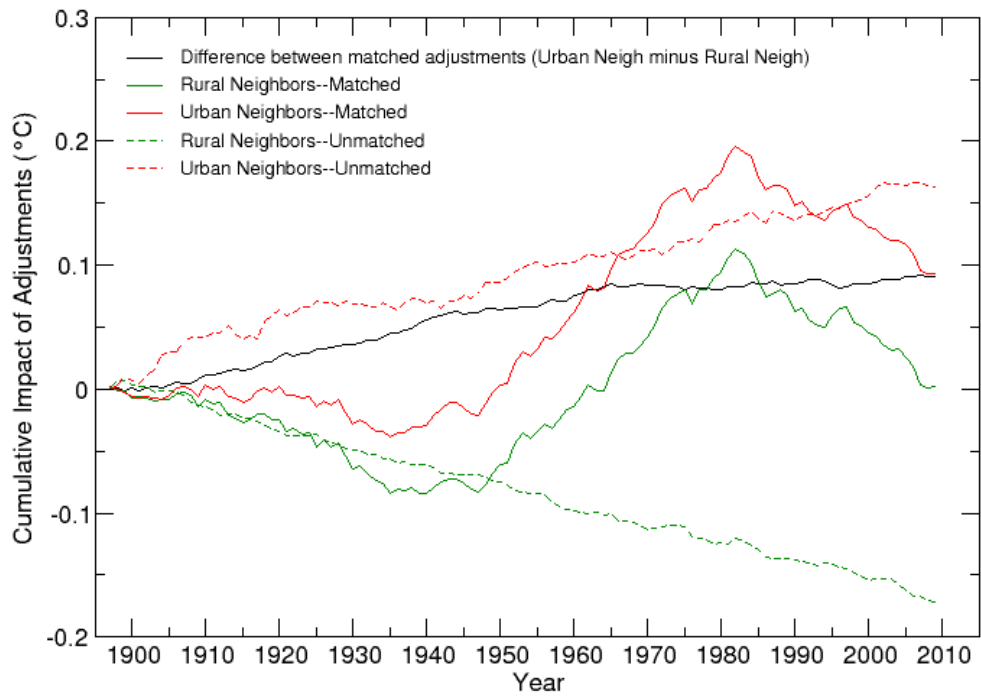Figure S.I. 5. As in Fig. S.I.3 but from stations classified using GRUMP.

Figure S.I. 6. As in Fig. S.I.2 but from stations classified using Nightlights.
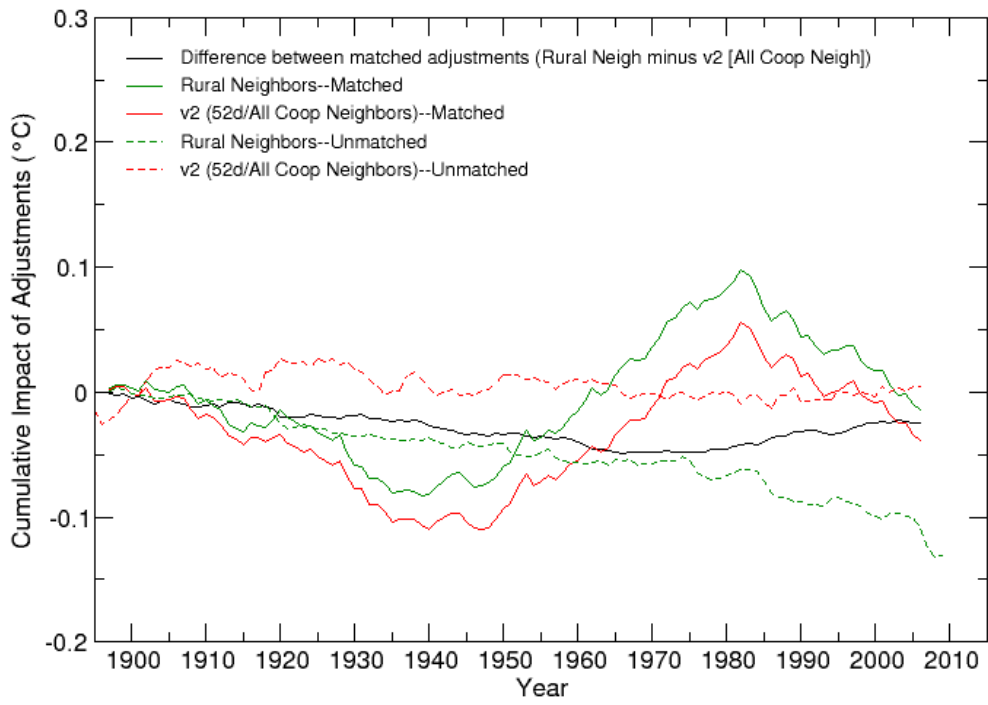
Figure S.I. 7. As in Fig. S.I.3 but from stations classified using Nightlights.
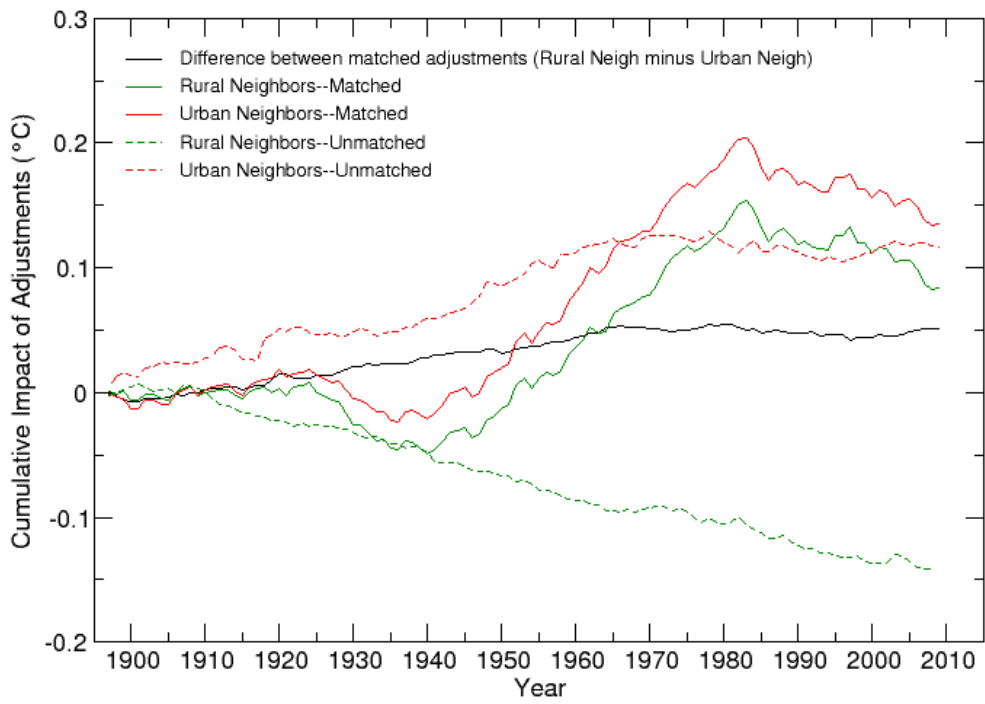
Figure S.I. 8. As in Fig. S.I.2 but from stations classified using population growth.
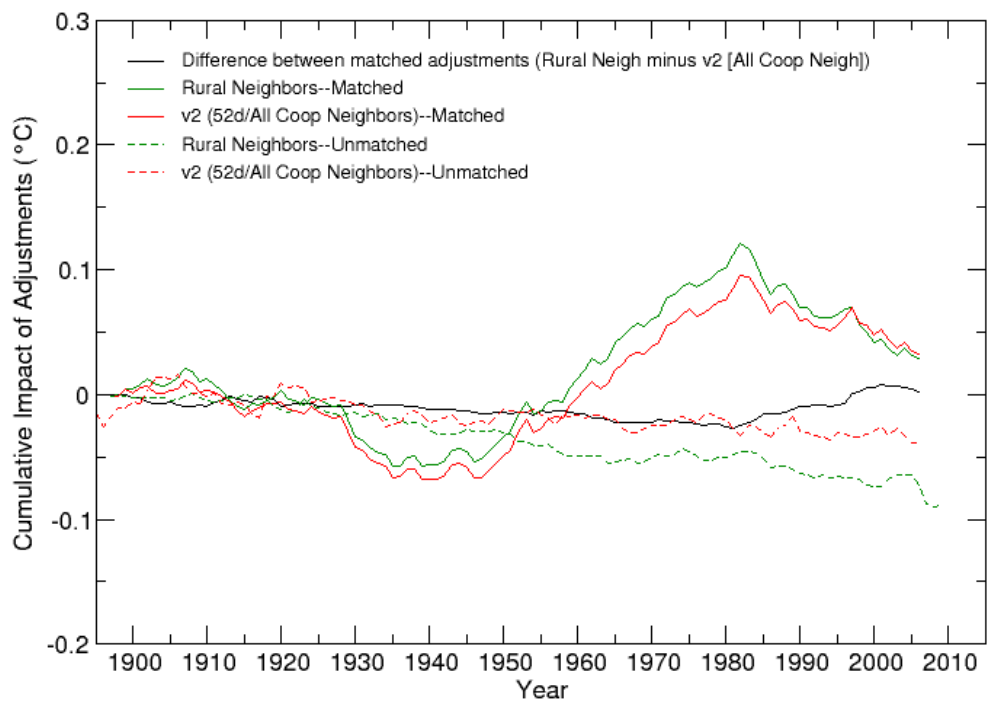
Figure S.I. 9. As in Fig. S.I.3 but from stations classified using population growth.

Data Sources

Urbanity Proxies

For the satellite nightlights, the Global Radiance Calibrated Nighttime Lights data-set year 2006, satellite F16 was used:

http://www.ngdc.noaa.gov/dmsp/download_radcal.html

For population growth and low population proxies, Gridded 1 km Population Estimates for the Conterminous U.S., 1930-2000 were used:

http://www.ncdc.noaa.gov/oa/climate/research/population/popdata.html

For impermeable surfaces, Global Distribution and Density of Constructed Impervious Surfaces was used:

http://www.ngdc.noaa.gov/dmsp/download_global_isa.html

For urban boundaries, Global Rural-Urban Mapping Project (GRUMP) data was used: http://sedac.ciesin.columbia.edu/gpw/documentation.jsp

Station and Temperature Data

USHCN TOB and version 2.0 maximum and minimum monthly temperature data:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013-suppinfo/data/ushcn-temps-9641C_201207.zip

USHCN v2.0 monthly maximum and minimum temperature data homogenized using rural only and urban only data are available here:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013-suppinfo/data/ ushcn-urban-rural.tar.gz

USHCN v2.0 monthly maximum and minimum homogenized temperature data with the additional NASA GISS (GISTEMP) step 2 urban correction applied are available here:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013-suppinfo/data/ushcn-with-step2.tar.gz

USHCN rural/urban proxy classifications, instrument type and transition dates:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013-suppinfo/metadata/ushcn-proxy-metadata.csv

Source Code

Code used for both the station pairing and spatial gridding analysis is available for the statistical software STATA here:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013-suppinfo/code/stata

A Java version of the code is also available:

ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/papers/hausfather-etal2013suppinfo/code/java

Fortran code (spheremap.f) used in the gridding method described in *Menne et al.* [2009; 2010] was obtained via anonymous ftp from climate.geog.udel.edu and is described here:

http://climate.geog.udel.edu/~climate/publication_html/Pdf/WRP_Am_Cart_85.pdf

Code used to produce both fully homogenized and rural-homogenized data via the Pairwise Homogenization Algorithm in *Menne and Williams* [2009] is available here:

ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/software/52i/

Code used to apply the NASA GISS Step 2 adjustment is available here:

http://code.google.com/p/ccc-gistemp/

Numerical values for trends and confidence intervals for Figures 1 and 2 are available here: ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/ papers/hausfather-etal2013suppinfo /trends/paired-and-gridded-trends-and-cis.csv