

Urban Heat Island Assessment: Metadata Are Important

THOMAS C. PETERSON AND TIMOTHY W. OWEN

NOAA/NESDIS/National Climatic Data Center, Asheville, North Carolina

(Manuscript received 17 September 2004, in final form 10 January 2005)

ABSTRACT

Urban heat island (UHI) analyses for the conterminous United States were performed using three different forms of metadata: nightlights-derived metadata, map-based metadata, and gridded U.S. Census Bureau population metadata. The results indicated that metadata do matter. Whether a UHI signal was found depended on the metadata used. One of the reasons is that the UHI signal is very weak. For example, population was able to explain at most only a few percent of the variance in temperature between stations. The nightlights metadata tended to classify lower population stations as rural compared to map-based metadata while the map-based metadata urban stations had, on average, higher populations than urban nightlights. Analysis with gridded population metadata indicated that statistically significant urban heat islands could be found even when quite urban stations were classified as rural, indicating that the primary signal was coming from the relatively high population sites. If ~30% of the highest population stations were removed from the analysis, no statistically significant urban heat island was detected. The implications of this work on U.S. climate change analyses is that, if the highest population stations are avoided (populations above 30 000 within 6 km), the analysis should not be expected to be contaminated by UHIs. However, comparison between U.S. Historical Climatology Network (HCN) time series from the full dataset and a subset excluding the high population sites indicated that the UHI contamination from the high population stations accounted for very little of the recent warming.

1. Introduction

Understanding urban heat island (UHI) contamination in the in situ climate record is a complex task because the results are impacted by a wide variety of factors not related to urbanization. For example, temperature observations are impacted by differing observing times, different instrumentation, and different siting practices, each of which may cause inhomogeneities in space when comparing values from several different stations in and around a town and discontinuities in time when examining how the climate is changing (Peterson et al. 1998). Metadata indicating which stations are rural and which are urban are useful, but the manner of categorization varies broadly, with important implications for researchers. Natural location effects confound the analyses even further as “unquestionably, many towns and cities are so located that,

even if we eliminated the man-made features, a microclimatic gradient would still exist between the city and the airport” (Landsberg 1970). Local-scale impacts of urban parks as well as the microscale environment around a station significantly impact observed temperatures (e.g., Spronken-Smith and Oke 1998; Gallo 2005). This paper focuses on the impact of one of these factors, rural/urban classification metadata, to determine the magnitude of the urban heat island signal using different metadata.

2. Metadata

a. Nightlights-based rural/urban metadata

The urban heat island assessment in Peterson (2003) used the rural/urban classification metadata developed by Owen et al. (1998) using nightlights data from the Defense Meteorological Satellite Program Operational Linescan System. Their methodology classified 1 km² grid boxes throughout the United States as urban, rural, or suburban. Nightlights intensity thresholds for urban and rural classifications were determined for the United States based on Geographic Information System (GIS) assessments and correlation with housing

Corresponding author address: Thomas C. Peterson, NOAA/NESDIS/National Climatic Data Center, 151 Patton Ave., Asheville, NC 28801.
E-mail: Thomas.C.Peterson@noaa.gov

density data. The decision tree used to classify station locations was based on both a local ($3\text{ km} \times 3\text{ km}$ or nine 1-km^2 boxes) analysis and regional ($21\text{ km} \times 21\text{ km}$ or 421 1-km^2 boxes) analysis. For a station to be classified as rural, no urban grid boxes are allowed within 3 km and no more than 25% of the regional grid boxes can be classified as urban or suburban. When any 1-km grid box within the station's local $3\text{ km} \times 3\text{ km}$ neighborhood was classified as urban or when more than 50% of the regional grid boxes were classified as urban or suburban, the station was classified as urban. Stations with no urban grid boxes in the local analysis and 25%–50% urban in the $21\text{ km} \times 21\text{ km}$ analysis were classified as suburban. An advantage of night-lights metadata in the United States is that both residential and industrial urbanization are identified. The Owen et al. (1998) methodology was verified using data from the U.S. Census Bureau (see U.S. Census Bureau 2002) with an 84.4% agreement on a cell-by-cell basis between the two metadata.

b. Map-based rural/urban metadata

The map-based metadata was created for use in Gallo et al. (1993). Essentially, each station was located on Defense Mapping Agency Operational Navigation Charts (ONC). If a station was in or associated with a city of 50 000 or greater population, it was classified as urban. As the maps showing urban boundaries were sometimes over a decade out of date, the geographer classifying a station near a city used his judgment, based on knowledge of growth rates and patterns in that region, to subjectively determine whether it should be considered urban or not. Stations not associated with a town of over 10 000 in population were classified as rural. A category of suburban was used for stations between the two classes. A similar ONC-based approach was used by Peterson and Vose (1997) to classify all the stations in the Global Historical Climatology Network (GHCN) as urban, small town, or rural. Peterson (2003) did not use the map-based metadata because of the subjective nature of the analysis and because Owen et al.'s (1998) comparison of the urban boundaries in ONC charts with U.S. Census Bureau urban areas found that the ONC urban area accounted for only 47.4% of the U.S. Census Bureau urban area on a cell-by-cell basis.

c. Gridded population metadata

Population values at a 1 km^2 grid cell resolution were prepared for the conterminous United States (CONUS) for both 1990 and 2000 using two U.S. Census Bureau datasets: The 2000 U.S. Census Bureau 1 km^2 population density grid for CONUS (National Geophysical

Data Center 2002) and tabular U.S. census county data (U.S. Census Bureau 2002). These datasets were merged in two steps. First, the 1-km^2 2000 gridded population density dataset for CONUS was merged with a 1 km^2 gridded dataset that defined the spatial extent of all counties. This combined information was then used to estimate the spatial distribution of population within each county for 1990 and 2000. Weights for each grid cell, summing to 1.0 for each county, were computed. Then these weights were multiplied by the county total for 1990 and 2000. Finally, values were adjusted for rounding so that the sum of each county's grid cells precisely matched the published census values. These metadata were previously used in Owen and Gallo (2000) to augment estimated populations for U.S. HCN stations with incomplete population records.

d. Station location information

Oke (2001) indicated that local- and microscale impacts can dominate over mesoscale UHIs with local-scale being on the order of an urban park and microscale referring to the house or garden around the observing station. Gallo et al. (1996), based on information provided by station observers, found that the microscale (100 m) land use around weather observing stations had a stronger influence on the observed temperature than the land use at local- and larger scales of 1 and 10 km. Having population metadata on a 1-km grid would imply that we may have the ability to evaluate the local-scale environment. Unfortunately, the station latitude and longitude was only resolved to degrees and minutes. Thus, one minute of latitude represents 1.9 km and, at 40°N , one minute of longitude represents 1.4 km. So a latitude–longitude box defined by the station location resolution covers 2.6 km^2 , which would include at least parts of from four to nine $1\text{ km} \times 1\text{ km}$ grid boxes. Ergo, in most cases, the station would not be located in the actual $1\text{ km} \times 1\text{ km}$ grid box containing the official latitude and longitude. Therefore micro- and local-scale analyses are not likely to be accurate and care should be taken when assessing any results associated with small radii around stations.

3. Temperature data

The data and the data adjustments used in this analysis were described in detail in Peterson (2003). The data are 3-yr averages of annual average mean temperature data (1989–91) at 289 stations located in 40 clusters around the contiguous United States, see Fig. 1. Ending the period in December 1991 allowed the analysis to avoid the confounding influence of the Automated Surface Observing System (ASOS) deployment, which started in 1992. Three years was determined to be long



FIG. 1. Map of stations and clusters used in these analyses. Temperature data meeting the requirements (complete monthly data for the three years analyzed from standard instrumentation located with standard siting) were not available from all of the stations; however, these stations were still used in metadata comparisons (Peterson 2003).

enough to produce representative means. A longer period would increase the problem of missing data as to be used, stations were required to be serially complete on a monthly basis and not have nonstandard siting (e.g., rooftop locations). The data were adjusted on a cluster basis to account for and remove the mean differences caused by differences in elevation, latitude, instrumentation, and time of observation. These data are referred to in the text as adjusted data.

To determine the latitude adjustment, for example, the United States was divided into 2° latitude by 1° longitude grid boxes. Over 5000 National Weather Service (NWS) Cooperative Network station latitudes and mean 1989–91 temperatures were converted to anomalies from the means of all the stations in their grid boxes and a regression line through all the data points was calculated to determine the mean change in temperature per degree of latitude. The efficacy of an adjustment using a constant rate has been questioned in Gallo (2005) based on analysis of one year of data at five pairs of stations. In three out of the five pairs, the more northerly station is warmer and in four out of five pairs, the higher elevation station is warmer. Due to the significant but random impact of microclimatic features around a station, analysis of 5000 stations is likely to produce a more robust assessment of the impact of latitude on temperature than an analysis of five station pairs. Furthermore, because air moving along the surface from a lower elevation station to a higher elevation station expands, which according to the First Law of

Thermodynamics would cool the air if no other factors were involved, to understand the true magnitude of the microclimatic influences one would have to take the elevation effect into consideration even if the higher elevation station were warmer. The analysis presented in Peterson (2003) not only took these effects into con-

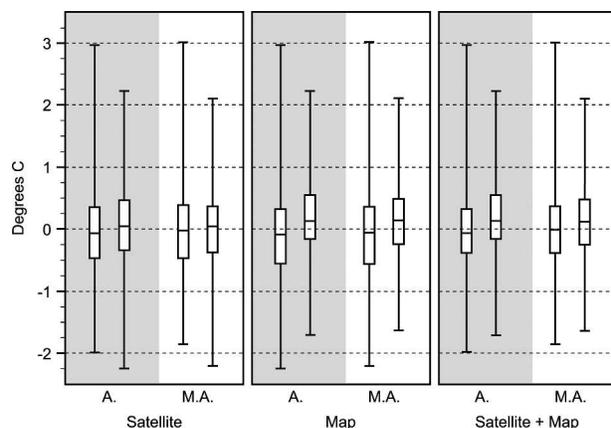


FIG. 2. Mean annual temperature anomalies relative to each cluster's mean temperature for (left) rural and (right) urban stations using both adjusted (gray background) and modified (white background) data and based on three different metadata classifications: satellite nightlights, map-based metadata, and both satellite- and map-based metadata. The box and whiskers indicate the median value (centermost line), the 25th and 75th percentiles (edges of the box), and extreme values (end of the whiskers).

TABLE 1. Evaluation of different groups of stations based on metadata. The station data, adjusted to account for location, instrumentation, and observation time differences, were turned into anomalies to the respective cluster mean. A multiresponse permutation test (Mielke 1991) was used to determine the probability that, given all the data points, two groups more different could occur by random chance alone. Temperature differences with probabilities less than 5% are given in bold. As the adjustments to remove instrumentation bias left statistically significant residuals, modifications were made also to the adjustments to ensure that the residual for each instrument type was zero.

	Mean urban minus rural temperature difference ($^{\circ}\text{C}$)		Number of		
	Using adjusted data	Using modified adjusted data	Clusters of stations	Rural stations	Urban stations
Satellite-derived metadata	0.03	-0.00	28	59	110
Map-derived metadata	0.31	0.25	32	119	68
Both satellite- and map-derived metadata	0.22	0.15	25	54	54

sideration but also thoroughly documented how well each adjustment worked.

Evaluation of the adjustments, as shown in Peterson (2003), indicated that latitude, elevation, and time of observation adjustments adequately removed these biases on a whole network basis. However, the instrumentation adjustments left statistically significant residual biases in the data. After the adjustments, data from stations with Maximum-Minimum Temperature Systems (MMTS) were $\sim 0.25^{\circ}\text{C}$ too cold compared to the stations with different instrumentation, while the adjusted HO-83 and liquid-in-glass thermometers in Cotton Region Shelters data were a little too warm. To

account for and remove this residual bias, the instrumentation adjustments were further modified to ensure that the networkwide average difference in temperature between instruments was eliminated based on the data themselves. These data are referred to as the *modified* adjusted data. The problem with modifying the data in this manner is that certain types of thermometers are more likely to be used in urban areas than other types of thermometers. So the modification has the potential for removing some urban heat island effect as well. On the other hand, not fully removing the instrumentation bias can allow an instrumentation adjustment problem to appear as an UHI effect. In the analysis presented here, the results from both versions of the data will be presented. In all cases, the most robust answers will be those where analyses of the two versions of the data agree.

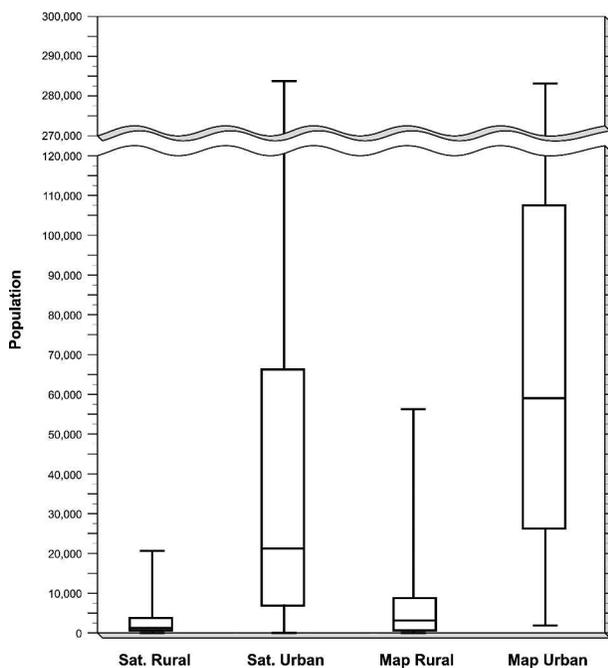


FIG. 3. Total population within 6 km of the stations based on the gridded 1-km population dataset.

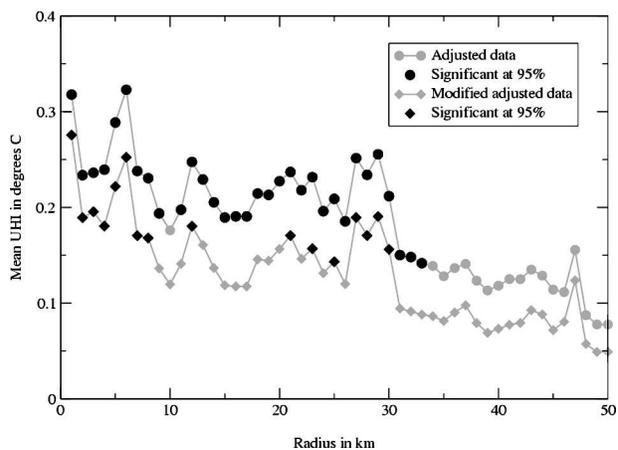


FIG. 4. Mean urban minus rural difference calculated in the same way as in Fig. 2 but the rural/urban designations were based on whether the population within a given radius was above or below the cluster mean population.

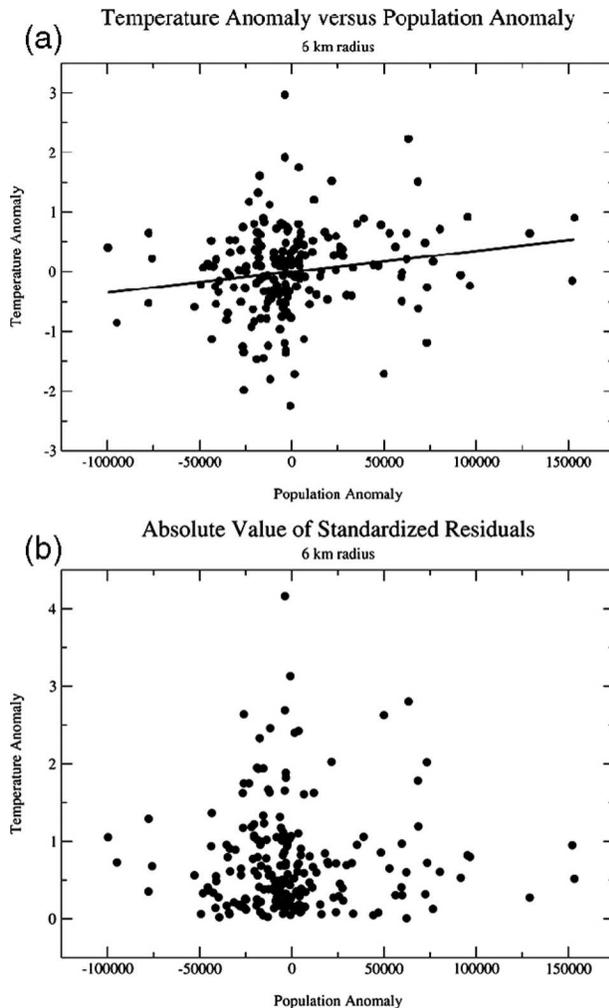


FIG. 5. (a) Temperature anomaly plotted against population anomaly, both anomalies with respect to the cluster means. Linear regression line is statistically significantly different than 0, but the data do not have a linear pattern. (b) The absolute value of the standardized residuals from the regression line shown in (a) vs population anomaly. The heteroscedasticity revealed by this plot indicates that a linear model is not a good representation of the data.

4. UHI analysis using satellite- and map-based metadata

For the analysis, station temperatures were converted to anomalies from the average of their respective clusters by subtracting each cluster's mean value from each station's 1989–91 mean temperature. Monthly and seasonal analyses were not performed, only the 3-yr mean temperature was used, as monthly and seasonal adjustments were determined not to be as robust as annual adjustments. Rural and urban groups of stations were defined using metadata. Only clusters with at least one station in each group were used in the

respective analysis. The null hypothesis that the temperature anomalies of the stations in these two groups were not significantly different was tested using a multiresponse permutation test (MRPP; Mielke 1991). The results of the analysis using the satellite nightlights-based rural/urban classifications are shown in Fig. 4 and in Fig. 5 of Peterson (2003). The mean urban minus rural difference is 0.03°C using adjusted data and -0.00 with the modified adjusted data. The first result differs slightly from the 0.04°C reported in Peterson (2003) with the difference due to correcting a processing error in the metadata assignments at a few of the stations. In both cases, these small differences are not statistically significant.

A similar analysis with the map-based rural/urban metadata as well as a combination of satellite nightlights- and map-based metadata are also shown in Fig. 2. As indicated in Table 1, with the map-based metadata, urban station temperatures were statistically significantly higher (0.31° and 0.25°C) than rural stations using adjusted and modified adjusted data. An additional analysis was performed using what should be only the “most rural” stations and the “most urban” stations defined by the station being classified as rural or urban by both metadata sets. Interestingly, using only the most rural and most urban stations, as shown in Table 1, the mean urban – rural temperature difference was fairly large but no longer statistically significant, perhaps because of the smaller sample size and variability. The large difference is also an artifact of the limited (47.4%) agreement in urban classification between the two approaches.

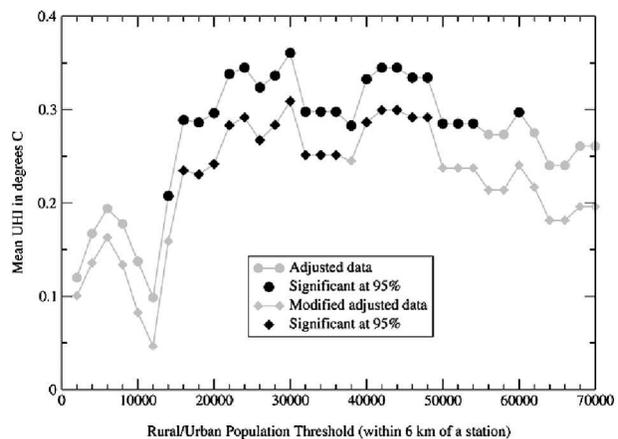


FIG. 6. UHI analysis where the station rural (urban) classification was based on whether it was below (above) a given population threshold. The first statistically significant UHI is detected at 14 000 people within 6 km. This classified 42% of the stations as urban. The peak value at 30 000 classifies 30% as urban. And the final statistically significant threshold at 60 000 within 6-km radius classifies only 18% of the stations as urban.

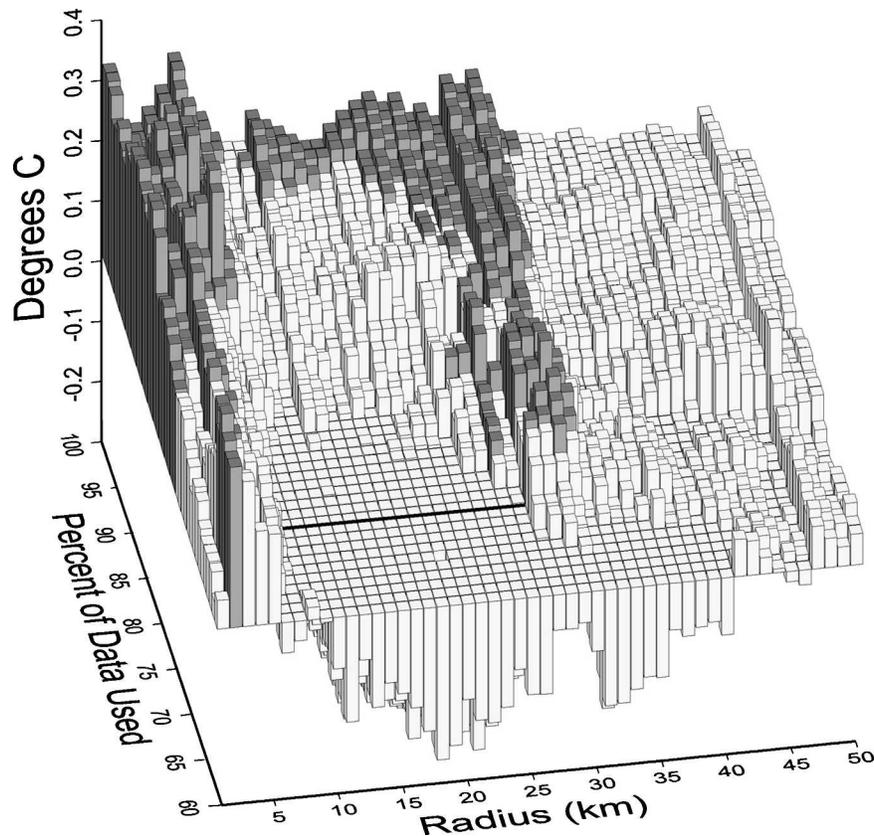


FIG. 7. (a) UHI analysis performed as described in Fig. 3 except using only a portion of the data as the highest population stations were removed prior to performing the analysis. Assessments significant at the 95% level are shaded. The x axis is the radius around the station location used to determine the population. The y axis is the percent of data used (100 minus the percent of highest populated stations removed). The heavy line indicates the threshold for removing the highest populated 30% of the data. The z axis is the mean high population minus low population temperature. The data used were the homogeneity adjusted data.

5. Assessment of the rural/urban metadata

To better understand the causes of the different results, the 1-km gridded population dataset was used. Figure 3 shows the population of the two different rural and urban station sets. The analysis used the 1990 population grid boxes with centers within a 6-km radius of each of the stations. The reason for using 6 km is provided in the following section. The results were found to be essentially the same for all the other radii (from 1 to 50 km) evaluated. Basically, rural stations classified using satellite nightlights are associated with smaller populations than map-based metadata rural stations. This should be expected from the station numbers: the nightlights metadata had more restrictive rural criteria, which classified only 29% of the stations as rural while the map-based metadata classified 62% as rural. This is also partially an artifact of the saturation and extension (“blooming”) of higher light frequencies

beyond the edge of cities (Imhoff et al. 1997). Similarly, the map-based rural/urban metadata had more restrictive urban criteria, which classified only 38% of the stations as urban compared to 69% urban for nightlights metadata.

6. UHI analysis using gridded population metadata

The 1-km gridded population metadata set provides the opportunity to perform a variety of additional UHI analyses based on population thresholds, different radii around the stations, and with actual population values rather than the binary rural/urban categorization metadata. The validity of using population data as a surrogate of urbanized land cover was established with strong correlation to housing density (Owen et al. 1998). During the process of evaluating the accuracy of homogeneity adjustments to account for, for example, the difference in elevation, Peterson (2003) divided the

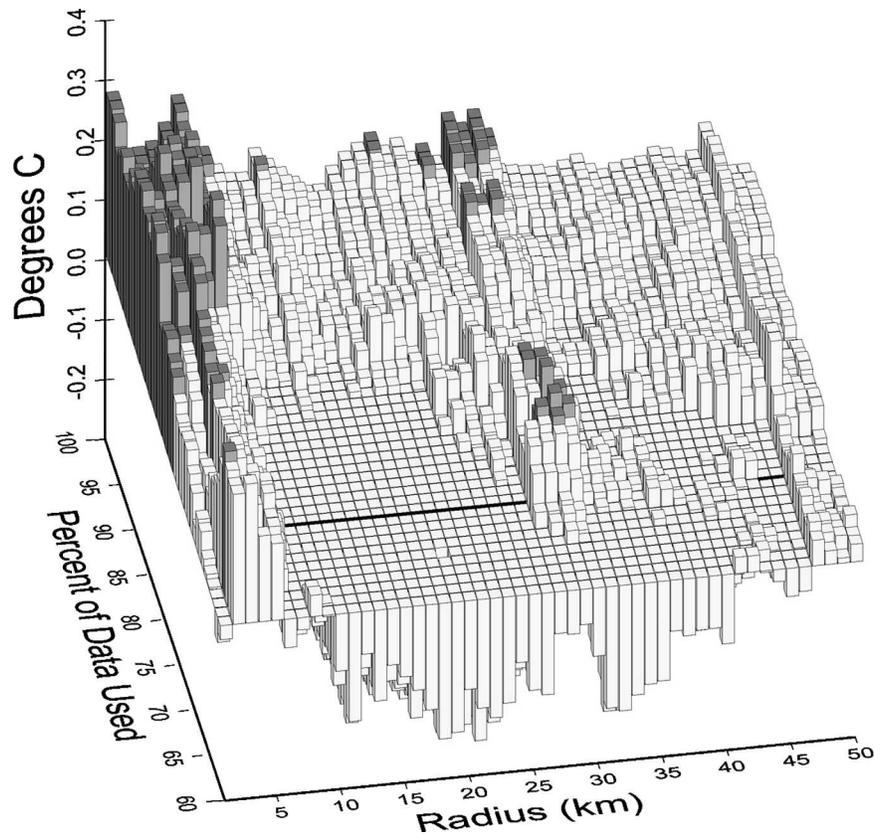


FIG. 7. (b) The same analysis as (a) but using modified adjusted data.

clusters into two groups according to whether the station was above or below the mean cluster elevation. In a similar way, an UHI temperature signal assessment was done by dividing the stations in each cluster based on whether the population associated with a station was above or below the mean clusterwide population.

The results of this analysis for populations within radii from 1 to 50 km of the stations are presented in Fig. 4. Results from analyses using both the adjusted data and the modified adjusted data indicate that the mean urban minus rural temperature differences were significant at the 95% level (presented in bold). The gridded population information (Fig. 4) can be used to divide the stations into significantly different groups. An $\sim 0.2^{\circ}$ – 0.3° C UHI temperature signal was found using adjusted data. The signal was less for adjusted data that were also modified to account for inadequately adjusted instrumentation. When the radius around the station exceeds ~ 30 km, the temperature differences are no longer significant. Indeed, with the modified adjusted data only a few radii beyond 8 km show statistically significant differences. The largest UHI for both the adjusted dataset and the modified adjusted dataset is detected using populations within a radius of 6 km.

The urban heat island effect may be further examined with a scatterplot of the temperature anomaly versus the population anomaly (Fig. 5a), which used the metadata indicating the strongest UHI effect found in the analyses (i.e., the adjusted dataset using populations in the 6-km radius). A linear trend indicates a change of about 1° C across the range of population anomalies in the clusters. This trend is statistically significantly different from zero but that does not mean it represents the data well. Figure 5b shows absolute value of the standardized residuals to the regression line versus population anomaly. The heteroscedasticity (i.e., the situation in which the variability of the residuals is not constant) revealed by this plot indicates that a linear model is not a good representation of the data (von Storch and Zwiers 1998). It should not be surprising from the scatter in Fig. 5a that the correlation r of temperature anomaly from the dataset having the strongest UHI signal and population anomaly at the radius with the strongest urban temperature signal is only 0.17, indicating that the population anomaly only explains 3% of the variance in temperature anomalies.

An analysis was also performed at the 6-km radius using fixed population thresholds for a rural/urban cut-

off. Figure 6 shows the results where stations above the threshold were classified as urban and below were classified as rural with the analysis incrementing at intervals of populations of 2000. The first statistically significant UHI result came at a threshold of 14 000 people within 6 km. The peak UHI effect came from classifying stations above (below) 30 000 people as urban (rural). The last statistically significant result came with a classification of 60 000. It is hard to interpret exactly how the population within 6 km translates into towns and cities. But it does offer insights into the mix of stations: a 14 000 threshold classifies 42% of the stations as urban, 30 000 classifies 30%, and 60 000 classifies 18%.

Few people would consider 60 000 people living within 6 km of a station a rural site, yet we found statistically significant UHI results with that as the rural/urban cutoff. The threshold results suggest two things. The first is that, if a significant UHI signal can be found when stations with fewer than 60 000 people living near them are considered rural, the UHI signal must be dominated by the most urban of the stations. Second, an appropriate rural/urban cutoff is likely somewhere between 14 000 and 60 000. Perhaps the peak value at 30 000 should be interpreted as indicating an appropriate threshold. From this perspective, theoretically only the most urban 30% of the stations would have to be removed from this dataset to avoid statistically significant urban heat island influences.

To test this hypothesis, analyses similar to that shown in Fig. 3 were repeated with the most urban stations removed. That is, at each cluster rural and urban were defined as those above and below the mean population of the cluster but, prior to performing the analysis, the highest population stations at that radius from an entire dataset standpoint were removed. Figures 7a and 7b show the results in three dimensions: the radius around the station, the percent of data kept in the analysis, and the magnitude of the UHI signal. Urban minus rural temperature differences that were statistically significant at the 95% level are shaded. Figure 7a uses adjusted data while Fig. 7b is from adjusted data modified to remove lingering instrumentation biases.

The highest urban minus rural temperature difference is where all the data were used and the radius around the station in which the population was calculated was small. When 10%–20% of the highest population stations were removed prior to performing the analyses, the observed urban minus rural difference is smaller. A heavy line indicates where the 70% threshold is, below which almost no statistically significant urban heat signal is found. In fact, the urban minus rural temperature difference is actually negative for

many of the radii when 30% or more of the highest population location stations are not used.

7. Discussion

Generally speaking, the satellite nightlights classification may have erred on the side of classifying stations as rural only if they were deemed fairly pristine (Owen et al. 1998). No significant UHI effect was found with satellite nightlights metadata, primarily because these pristine stations' temperatures were not significantly different from temperature at the large number of somewhat higher population stations. Only when classifying fewer stations as urban (i.e., only the most urban), which the map-based metadata did, was a significant UHI signal found. This suggests that the UHI impact may not be pervasive in the U. S. in situ temperature record as it is primarily in the highest population locations.

The results shown in Figs. 7a and 7b suggest that removing the 30% of the most populated stations from an UHI analysis essentially removes the UHI signal from U.S. data. As stated earlier, the 30% most urban in this analysis refers to stations with 30 000 or more people within 6 km. The U.S. HCN (Easterling et al. 1996) is the most widely used in situ network for long-term temperature change analyses in the United States. Figure 8 shows the percent of U.S. HCN stations with populations in excess of various thresholds within 6 km calculated from the 1 km gridded population data for 2000 rather than 1990 as used previously. It turns out that 16% of the U.S. HCN stations are at or above the 30 000 population within the 6-km threshold. As the

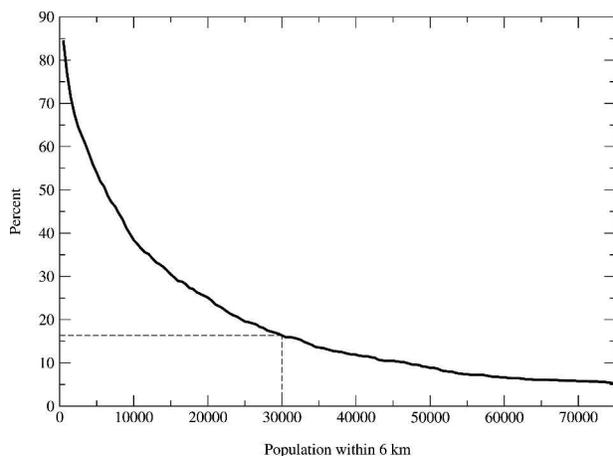


FIG. 8. Percent of U.S. HCN stations exceeding various 6-km radius population values. Population data are from the 2000 census. Dashed line indicates that 16% of the stations have populations within 6 km of over 30 000.

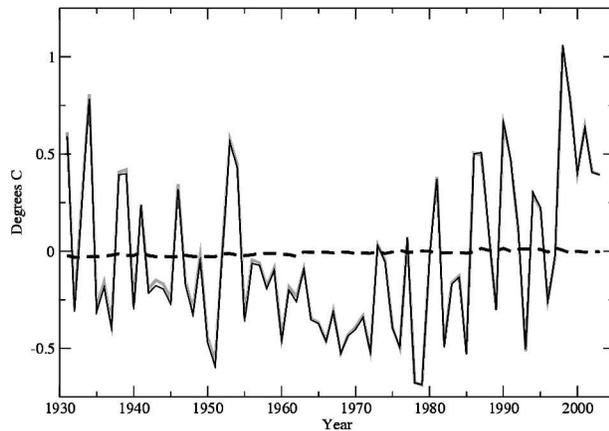


FIG. 9. Anomaly time series of the full U.S. HCN data (black) and the U.S. HCN data without the 16% of the stations with populations of over 30 000 within 6 km in the year 2000 (gray). The full dataset values minus the set without the most urban stations are shown with a dashed line. Clearly the warming in recent years is not caused by increasing UHI effects at the most urban stations.

heteroscedasticity discussed earlier indicates that the regression relationship was not reliable enough for determining an UHI adjustment, the implications of this analysis is that, rather than adjust U.S. HCN data to account for potential UHI impacts on a wholesale basis, efforts should be made to use lower population sites for U.S. HCN stations situated in high population locations. It is reassuring that $\sim 85\%$ of the U.S. HCN data are in low enough population locations that they are unlikely to have UHI contamination. As 30 000 people within 6 km is such a high threshold, it should be possible to find replacement stations where necessary. This, of course, reflects the care that went into selecting the U.S. HCN sites from the mesoscale perspective available to researchers in the 1980s (Easterling et al. 1996).

Comparison of time series of a subset of the U.S. HCN that excludes any stations with populations of over 30 000 within 6 km of the station in the year 2000 with the full U.S. HCN dataset (Fig. 9) indicates that UHI contamination from these higher population stations can explain very little of the recently observed warming. The U.S. HCN data used includes all homogeneity adjustments except that for urbanization. To minimize the differences in the time series caused by the subset not observing part of the country, the time series shown in Fig. 9 start in 1931, the first year the subset had data in all 120 2.5° latitude \times 3.5° longitude grid boxes normally used in the National Climatic Data Center's analysis of U.S. HCN. A linear regression of the difference series created by subtracting the more rural subset from the full dataset has a value of 0.048°C

century $^{-1}$. This compares well with the Jones et al. (1990) determination that the impact of urbanization on hemispheric temperature time series was, at most, 0.05°C century $^{-1}$.

The population within very small radii around the stations seems to have the most significant results shown in Figs. 7a and 7b. This implies that the most important part of the urban heat island effect might not be the mesoscale UHI but rather smaller-scale influences. Gallo (2005) indicated that microclimatic elements associated with station siting can have large effects on the observed temperature. Theoretically, if there is a large population within close proximity of a station, it would likely limit the potential siting options and therefore microclimatic characteristics. It could also have a direct microclimatic influence along the line of the classic UHI causes involving asphalt and buildings instead of trees and grass as well as sky view geometry. Unfortunately, the precision of station location information defines a 2.6 km^2 area, making it difficult to draw firm conclusions from the very nearby $1\text{ km} \times 1\text{ km}$ population metadata.

It is also worth noting the differences between the adjusted data, which contained instrumentation-related biases and the modified adjusted data, which took the residual biases into account. In almost all cases, the calculated UHI was less using the modified adjusted data than with the adjusted data. The results are likely to be most robust where they agree. So care should be taken when drawing conclusions from only one of these datasets, as each of them are likely to contain slight biases—just different biases. Care should also be taken when drawing any conclusions where the results are not statistically significant as, for example, the negative UHI results at 6-km radius using 70% of the data would be hard to interpret otherwise.

8. Summary and conclusions

Urban heat island analyses for CONUS were performed using three different forms of metadata: nightlights-derived metadata, map-based metadata, and gridded U.S. Census Bureau population metadata. The results indicated that metadata are important: Whether a UHI signal was found or not depended on which metadata were used. The UHI signal is apparently very weak, for example, population was able to explain at most only a few percent of the variance between weather observing station temperature data. The nightlights metadata tended to have the lower population rural stations compared to map-based metadata while the map-based metadata urban stations had, on average, higher populations than nightlights urban. Analysis

with gridded population metadata indicated that statistically significant urban heat islands could be found when even quite urban stations were classified as rural, indicating that the primary signal was coming from the relatively high population sites. If ~30% of the highest population stations were removed from the analysis, no statistically significant UHI could be found.

The implications of this work on U.S. climate change analyses is that, if the highest population stations are avoided, the analysis should not be expected to be contaminated by urban heat islands. However, analysis of time series of the full U.S. HCN dataset and a more rural subset that excludes these high population stations indicates that UHI contamination from these high population stations can explain very little of the recent warming (only $0.048^{\circ}\text{C century}^{-1}$). This agrees with the work of Jones et al. (1990), which concluded that the impact of urbanization on hemispheric temperature time series was, at most, $0.05^{\circ}\text{C century}^{-1}$ as well as the recent work of Parker (2004), which demonstrated that urban warming has not introduced significant biases into estimates of recent global warming by comparing trends in temperatures from calm nights when UHI should be enhanced with windy nights when UHIs are reduced.

Acknowledgments. The authors would like to acknowledge the useful discussions with Ned Guttman and Russell Vose during various stages of this research. The excellent suggestions by Phil Jones and an unnamed reviewer were greatly appreciated. This work is supported by NOAA's Climate and Global Change Climate Change Data and Detection and Office of Biological and Environmental Research, U.S. Department of Energy.

REFERENCES

- Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, D. P. Bowman, R. C. Daniels, and T. A. Boden, 1996: United States Historical Climatology Network (U.S. HCN) monthly temperature and precipitation data. Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, Publication 4500, 83 pp. + appendixes.
- Gallo, K. P., 2005: Evaluation of temperature differences for paired stations of the U.S. Climate Reference Network. *J. Climate*, **18**, 1631–1638.
- , A. L. McNab, T. R. Karl, J. F. Brown, J. J. Hood, and J. D. Tarpley, 1993: The use of NOAA AVHRR data for assessment of the Urban Heat Island effect. *J. Appl. Meteor.*, **32**, 899–908.
- , D. R. Easterling, and T. C. Peterson, 1996: The influence of land use/land cover on climatological values of the diurnal temperature range. *J. Climate*, **9**, 2941–2944.
- Imhoff, M. L., W. T. Lawrence, D. C. Stutzer, and C. D. Elvidge, 1997: A techniques for using composite DMSP/OLS 'city lights' satellite data to map urban areas. *Remote Sens. Environ.*, **61**, 361–370.
- Jones, P. D., P. Ya. Groisman, M. Coughlan, N. Plumier, W.-C. Wang, and T. R. Karl, 1990: Assessment of urbanization effects in time series of surface air temperature over land. *Nature*, **347**, 169–172.
- Landsberg, H., 1970: Meteorological observations in urban areas. *Meteorological Observations and Instrumentation, Meteor. Monogr.*, No. 11, Amer. Meteor. Soc., 91–99.
- Mielke, P. W., 1991: The application of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Sci. Rev.*, **31**, 55–71.
- National Geophysical Data Center, cited 2002: Land use change project data: 2000 U.S. census population grids. [Available online at <http://dmisp.ngdc.noaa.gov/html/download.html>.]
- Oke, T. R., 2001: Climate and the urban environment. *Tech. Conf. on Climate Services for the Twenty-First Century*, Geneva, Switzerland, WMO Commission for Climatology.
- Owen, T. W., and K. P. Gallo, 2000: Updated population metadata for the United States Historical Climatology Network stations. *J. Climate*, **13**, 4028–4033.
- , —, C. D. Elvidge, and K. E. Baugh, 1998: Using DMSP-OLS light frequency data to categorize urban environments associated with U.S. climate observing stations. *Int. J. Remote Sens.*, **19**, 3451–3456.
- Parker, D. E., 2004: Large-scale warming is not urban. *Nature*, **432**, 290.
- Peterson, T. C., 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. *J. Climate*, **16**, 2941–2959.
- , and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- , and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517.
- 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. U.S. Census Bureau, Washington, DC, 226 pp.
- von Storch, H., and F. W. Zwiers, 1998: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.