An area and distance weighted analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends

PRE-PRINT DRAFT DISCUSSION PAPER

Anthony Watts
President, IntelliWeather, Chico, CA, USA

Evan Jones
IntelliWeather, Chico, CA, USA

Stephen McIntyre, Toronto, Canada

John R. Christy
Department of Atmospheric Science, University of Alabama, Huntsville, AL, USA

[plus additional co-authors that will be named at the time of submission to the journal]

Abstract

In Fall et al, 2011, results from the recently concluded Surface Stations Project surveying the U.S. Historical Climatology Network (USHCN) were presented, using a siting classification system developed by Michel Leroy for Meteofrance in 1999, and employed by the National Oceanic and Atmospheric Administration (NOAA) to develop the U.S. Climate Reference Network (USCRN) in 2002. In 2010, Leroy improved upon this system to introduce a "maintained performance classification" which quantifies the effect of heat sinks and sources within the thermometer viewshed by calculation of the area-weighted and distance-weighted impact of biasing elements such as concrete, asphalt, runways, tarmac, and buildings, creating a new site classification that more accurately reflects the representivity of the station exposure. The new area and distance weighted classification system does a more complete job of siting assessment, particularly when applied retroactively to existing stations, than the original distance weighted classification system described in Leroy (1999) , which performs well for new station siting evaluation, but does not take into account the surface area of heat sinks and sources that may encroach upon a temperature measurement station over its lifetime.

In Fall et al. (2011), using Leroy's 1999 classification system, it was demonstrated that station exposure affects USHCNv2 temperatures, in particular the minimum temperatures, but showed little difference in mean temperature trends used to assess climate variability. Menne et al. (2010), and Muller et al. (2012), both of which also used the older Leroy (1999) classification system, suggested there is little if any mean

temperature trend difference between well and poorly sited stations. Using the new Leroy (2010) classification system on the older siting metadata used by Fall et al. (2011), Menne et al. (2010), and Muller et al. (2012), yields dramatically different results.

Both raw and gridded comparisons were performed on the 30 year trends that were calculated for each surveyed station, using temperature data from USHCNv2. Mean temperature trend is indisputably lower for well sited stations than for poorly sited stations. Minimum temperature trend shows the greatest differences between siting classification while maximum temperature trend shows the smallest.

Well sited stations consistently show a significantly lower trend than poorly sited stations, no matter which class of station is used for a baseline for comparison, and also when using no baseline at all. Well sited stations, using a localized Class 4 (the most common class) baseline show a trend that is 0.09°C per decade lower than poorly sited stations for raw mean temperature trends. Raw mean temperature trends for well sited stations are 0.145°C per decade lower than adjusted mean temperature trends for poorly sited stations, and 0.145°C per decade lower than adjusted mean trend for all stations.

Comparisons demonstrate that NOAA adjustment processes fail to adjust poorly sited stations downward to match the well sited stations, but actually adjusts the well sited stations upwards to match the poorly sited stations. Well sited rural stations show a warming nearly three times greater after USHCNv2 adjustments are applied.

It is also demonstrated that urban sites warm more rapidly than semi-urban sites, which in turn warm more rapidly than rural sites. Since a disproportionate percentage of stations are urban (10%) and semi-urban (25%) when compared with the actual topography of the U.S., this further exaggerates mean temperature trends. Montandon et al (2011) documents this large urban bias in station siting on the Global Historical Climate Network.

These factors, combined with station siting issues, have led to a spurious doubling of U.S. mean temperature trends in the 30 year data period covered by the study from 1979 - 2008.

Keywords: Surface Temperature, Historical Climate Network, U.S. Temperature Trend

1. Introduction

A number of recent studies have addressed the myriad of factors and biases associated with temperature surface measurement in the United States. The identified biases include station moves, changes in instrumentation, localized changes in instrumentation location, changes in observation practices, and evolution of the local and microsite station environment over time. Some of the identified changes have been addressed in previous works such as where land use/cover change are considered (e.g. Asaeda et al.,(1996); Baker,(1975); Karl and Williams,(1987); Karl et al.,(1988); Karl et al.,(1989); Davey and Pielke,(2005); Mahmood et al.,(2006, 2010), Pielke et al.,(2007a and 2007b); Yilmaz et al.,(2008); Christy et al.(2009). It has been described by these and other studies that maximum and minimum temperatures measured at the station are affected in different ways by the changes in the station environment. McNider et al.,(2012) shows that even slight increases in the vertical mixing near the observing site (such as a local change in the surface land use) can result in significant changes in the minimum temperature trend. Such nearby changes in the station environment can create inhomogeneities, which in turn induce artificial trends or discontinuities in long-term temperature time series and can result in erroneous characterization of climate variability (Peterson et al., 1998; Thorne et al., 2005). Thus, even if stations are initially placed at pristine locations, i.e. "well-sited", the station environment can change, altering the characteristics of surface temperature measurements over time. As documented in surveys presented in Watts, (2009), and also in Fall et al.,(2011), the USHCN has a significant portion of stations affected by such changes, with approximately 10% of the USHCN remaining classified as "well-sited" using the Leroy (1999) classification method.

There have also been a number of attempts to address these station inhomogeneities. These include statistical identification methods for detecting, quantifying, and removing discontinuities and various non-climatic biases that affect temperature records have been employed (e.g. Karl et al., 1986; Karl and Williams, 1987; Quayle et al., 1991; Peterson and Easterling, 1994; Imhoff et al., 1997; Peterson et al., 1998; Hansen et al., 2001; Vose et al., 2003; Menne and Williams, 2005; Mitchell and Jones, 2005; Brohan et al., 2006; DeGaetano, 2006; Runnalls and Oke 2006 Reeves et al., 2007; Menne and Williams, 2009; Muller et al, 2012). in order to obtain homogeneous data and create reliable long-term surface temperature time series. Menne et al. (2009) for the United States Historical Climatology Network, Version 2 (USHCNv2), rely exclusively on detecting changes within the unadjusted surface temperature data itself to identify and correct time-varying non-climatic biases. Because of the unreliability of the archived metadata, some recently introduced adjustment approaches, such as that described by Menne et al. (2010), are not fully comprehensive, and are a tradeoff between leaving large undocumented changes uncorrected and inadvertently altering true local climate signals while also failing to detect and correct for other inhomogeneities such as changes in the station siting environment. An example of the incompleteness of their approach is reported, as one example, in Martinez et al (2012), who reported that

"Significant differences in temperature trends based on the surrounding land use were found for minimum temperature and temperature range in the 1970–2009 period indicating that data homogenization of the USHCN temperature data did not fully remove this influence"

The incompleteness by Menne et al. (2010) in correcting for non-climatic effects and non-spatially representative trends can explain the divergence in the multi-decadal temperature trend diagnosed for the surface and the lower troposphere Klotzbach et al. (2009, 2010)

Menne et al. (2010) analyzed the 1979-2008 temperature trends of stations grouped into two categories based on the quality of siting. They found that a trend bias in non-compliant sites relative to compliant sites is consistent with instrumentation changes that occurred in the mid- and late 1980s (conversion from Cotton Region Shelter-CRS to Maximum-Minimum Temperature System-MMTS). The main conclusion of their study is that there is *"no evidence that the CONUS temperature trends are inflated due to poor station siting".*

In Fall et al. (2011), it was it was demonstrated that station exposure affects USHCNv2 temperatures, in particular the minimum temperatures, but showed little difference in mean temperature trends. It was noted however, that there was no century scale trend observed in the diurnal temperature variation.

In Muller et al. (2012), there has been considerable new work done to account for known inhomogeneities and obtain adjusted surface temperature datasets for climate analysis using the station siting metadata from Fall et al. (2011). In Muller et al. (2012), a statistical analysis identified a -0.014 ± 0.028 C per century difference between well sited and poorly sited weather stations identified in the Fall et al., 2011, metadata set. Muller et al.,(2012), concluded, *"The absence of a statistically significant difference indicates that these networks of stations can reliably discern temperature trends even when individual stations have nominally poor quality rankings.".*

Independent of the recent finding in Muller et al.,(2012), the National Climatic Data Center (NCDC) has long recognized the need for a climate monitoring network as free as possible from non-climatic trends and discontinuities and has developed the United States Climate Reference Network (USCRN) to fill this need. (NOAA/NESDIS Climate Reference Network-CRN, 2002). Using the method outlined by Leroy,(1999), NOAA USCRN sites were selected based on the consideration of geographic location factors including their regional and spatial representivity, the suitability of each site for measuring long-term climate variability, and the likelihood of preserving the integrity of the site and its surroundings over a long period. The method adopted from Leroy (1999) was appropriate in achieving this goal, because it attempts to quantify the impacts of visible microsite issues for new climatic station sites under consideration for inclusion into the USCRN. The method from Leroy (1999) relies mainly on one observed value, distance from visible heat sinks and heat sources to the thermometer instrumentation, to quantify the station environment as being suitable for deployment of a USCRN climate monitoring site. Having no other published metric by which to gauge station siting and create representative metadata, the resultant siting metadata suggested by Leroy (1999) derived from the Watts (2009) survey, was utilized in Menne et al.,(2010), Fall et al., (2011), and also Muller et al.,(2012). In all cases, station siting effects on mean temperature trends were observed to be small. However, this was metadata derived from the Leroy (1999) siting classification system, which was designed for site pre-selection, rather than retroactive siting evaluation and classification.

The improved Leroy (2010) siting classification system, which included a method for including the surface area of heat sinks and heat sources within the viewshed of thermometer was endorsed by the World Meteorological Organization Commission for Instruments and Methods of Observation Fifteenth session (CIMO-XV, 2010), in September 2010 stating: "*The Commission agreed that the publication of the siting classification as a common WMO-ISO standard would help in assessing and improving the quality of data originating from WMO-owned, cosponsored and non-WMO observing networks. The Commission agreed to further develop this classification as a common WMO-ISO standard.*"

Given that the WMO has endorsed the Leroy (2010) classification system in CIMO-XV (2010) as a WMO-ISO standard, it is suitable for use in re-assessing the station quality issues reported by Watts (2009)., Menne et al.,(2010), Fall et al.,(2011), and Muller et al.,(2012).

The new siting classification system proposed in Leroy (2010) and accepted by CIMO-XV is similar to the Leroy (1999) system, but adds total surface area to the distance measurement as an additional metric for determining station site representivity for thermometers. This resulted in a dramatic and statistically significant improvement in the binning of stations quality ratings as distance alone does not quantify the amount of heat emitted by a source or sink within the thermometer viewshed. As an example, in Lee (1995), it was demonstrated that the design of heat sinks for electronics cooling is highly dependent on the total surface area available to radiate thermal energy away from the surface. The greater the surface area of the heat sink, the more efficient it is at exchanging heat with the fluid medium surrounding it, and in the case of this study, that is the surface layer atmosphere within the thermometer viewshed. Two physical processes are involved with heat sinks and sources within the thermometer viewshed; mass transfer and radiative transfer. Fourier (1822) described the process of mass transfer of heat, such as between a surface and a gas. This process has been observed where wind transport moves heat from nearby artificial surfaces such as asphalt, concrete, and buildings to nearby thermometers, which is the basis for the distance component of the Leroy (1999, 2010) rating systems: to allow adequate mixing of the boundary layer atmosphere, thus minimizing the mass transfer bias before reaching the thermometer. As for radiative transfer, Aseada et al. (1996) reported from measurements and analysis:

"*At the maximum, asphalt pavement emitted an additional 150 W m^{-2} in infrared radiation and 200 W m^{-2} in sensible transport compared to a bare soil surface. Analyses based on a parallel layers model of the atmosphere indicated that most of the infrared radiation from the ground was absorbed within 200 m of the lower atmosphere, affecting air temperature near the ground.*"

It follows that the total amount of infrared radiation and sensible heat released by such artificial surfaces is dependent on the number of square meters of surface area within the thermometer viewshed, thus making the Leroy (2010) rating system, which combines surface area and distance to define the station site rating, more valuable at quantifying the representivity of the station site for temperature measurements than distance alone as was done in Leroy (1999) and the subsequent studies that used that rating system.

Many USHCNv2 stations which were previously rated with the methods employed in Leroy (1999) were subsequently rated differently when the Leroy (2010) method was applied in this study. This simple change in the rating system accounts for the majority of differences in the data and conclusions between this study and Menne et al.,(2010), Fall et al.,(2011), and Muller et al.,(2012). Effectively, the lack of accounting for the surface area of heat sinks and sources using Leroy (1999) methods in Menne et al (2009), Fall et al.,(2010), and Muller et al.,(2012) resulted in binning errors of trends for site representivity, providing what amounted to a pseudo-randomization of the station data in the context of heat sinks and sources, rendering the signal for siting issues into the noise bands of the data. Once the Leroy (2010) site rating system was applied, the binning error was removed, and the signal demonstrating the differences in station trends between siting classes became clear.

2. Data and methods

2.1. USHCNv2 Climate Data

The USHCNv2 monthly temperature data set is described by Menne et al. (2009). The raw and unadjusted data provided by NCDC has undergone the standard quality-control screening for errors in recording and transcription by NCDC as part of their normal ingest process but is otherwise unaltered. The intermediate (TOB) data has been adjusted for changes in time of observation such that earlier observations are consistent with current observational practice at each station. The fully adjusted data has been processed by the algorithm described by Menne et al. (2009) to remove apparent inhomogeneities where changes in the daily temperature record at a station differs significantly from neighboring stations. Unlike the unadjusted and TOB data, the adjusted data is serially complete, with missing monthly averages estimated through the use of data from neighboring stations. The USHCNv2 station temperature data in this study is identical to the data used in Fall et al. (2011), coming from the same data set.

2.2. Station Site Classification

We make use of the subset of USHCNv2 metadata from stations whose sites have been classified by Watts (2009), gathered by the volunteers of the surfacestations.org project using the USCRN site-selection classification scheme for temperature and humidity measurements (NOAA/NESDIS 2002), and originally developed by Leroy (1999). For Watts (2009) and Fall et al. (2011), USCHNv2 site surveys were originally performed between June 2nd, 2007 and Feb 23rd, 2010. For the purpose of this study, the original site rating metadata from Fall et al (2011), also used in Muller (2012), was supplemented with further refinements and additional station surveys inclusive from June 15th, 2011 to July 1st, 2012, followed by application of the Leroy (2010) site survey rating system to both old and new surveys (Table 1) including both a distance and an area rating component. Any known changes in siting characteristics after that period are ignored. A total of 1065 USHCNv2 stations were surveyed, comprising 87.4% of the 1218 station USHCNv2 network. Of those 1065 stations surveyed, 779 were classified per the Leroy (2010) site survey rating system (Figure 1). As a rule, LeRoy (2010) is less “strict” than Leroy (1999). There is a greater number of Class 1, 2, and 3 stations, and fewer Class 4 stations. There are, however, a greater number of Class 5 stations, as well.

In our urban-rural comparisons we use the Urban, Semi-Urban, Rural classifications provided by NASA. We divide the continental contiguous USA into twenty-six 6-degree grid boxes so that the gridding process eliminates distribution bias.

Because the great majority of the station surveys occurred prior to creation the of Leroy (2010) site survey rating system, site surveys previously acquired and used in Fall et al. (2011) and Muller et al (2012) were retroactively resurveyed, and wherever possible, had additional land and aerial photography added, so that surface area measurements required for the Leroy (2010) site survey rating system could be performed. In addition to station ratings, the survey provided an extensive documentation composed of station photographs and detailed survey forms. Because some stations used in Fall et al. (2011) and Muller et al. (2012) suffered from a lack of the necessary supporting photography and/or measurement required to apply the Leroy (2010) rating system, or had undergone recent station moves, there is in a smaller set of station rating metadata (779 stations) than used in Fall et al (2011) and Muller et al. (2012), both of which used the data set containing 1007 rated stations.

For each site in this study, ground and or aerial photography was obtained, distance measurements of visible encroachments were made, and a calculation was done to determine the percentage of area within the different radii (3m, 5m, 10m, 30m, and 100m) surrounding the thermometer per Leroy (2010), containing heat sinks and/or heat sources. The distance and area values were applied to the final rating for each station. Quality control checks were routinely done to ensure that the proper station was identified, that it matched descriptions in metadata provided by NCDC, that it was consistent with the latitude and longitude given for the station, and that the equipment seen in photography and described in survey reports matched the equipment description aaccording to NCDC metadatabase. Where discrepancy existed, interviews were conducted with the station curator when possible to resolve such discrepancy and to ensure the location of the thermometer in some aerial photos that had marginal resolution. Where such discrepancies could not be resolved, or it was determined from photographs, metadata, or curator interviews that the station had been closed or moved after 2002, and prior location could not be established, that station was excluded from consideration and not included in this study. Since the site metadata is either incomplete or cannot be verified for those stations that were excluded, it became impossible to bin them into their siting classes for use in this study. Examples of problems that caused exclusion include but are not limited to; recent station moves that made a station previously identifiable now unidentifiable, obscuration of the thermometer viewshed in aerial photos preventing a full distance and area measurement are; low resolution aerial photography that made it impossible to identify the exact location of the thermometer for measurements, no usable aerial photographic coverage at all, and inability to contact the site curator for verification of details not clearly visible in aerial and ground photography.

The best sites (compliant per Leroy, 2010) consist of 160 stations classified as either Class 1 (48 stations) or Class 2 (112 stations) doubling the number of compliant stations used in Fall et al. 2011 (80 stations), where the Leroy (1999) site survey rating system was applied. The worst (non-compliant per Leroy 2010) sites, of Classes 3, (247 stations) 4, (277 stations) and 5 (95 stations), comprise the majority of the USHCNv2 network with 619 stations at 79.5% (Table 2). The distribution of the best and poorest sites is displayed in Figure 1. Because Leroy (2010) considers both Class1 and Class 2 sites to be acceptably representative for temperature measurement, with no associated measurement bias, these were combined into the single “compliant” group with all others, Class, 3,4,and 5 as the “non-compliant” group. In contradiction to Leroy (1999) and Leroy (2010) publicly available review papers for Muller et al. (2012), showed they used grouping of Classes 1,2,3 as compliant sites, and Classes 4&5 as non-compliant sites. In addition to the lack of class binning using surface area by applying Leroy (2010) site classifications, this may also have contributed to Muller et al. (2012) finding no discernible trend differences between station classes.

As in Fall, et al (2011), Menne (2010), and Muller (2012), only the heat source/sink proximity and area ratings from Leroy 2010 and are do consider ground-level vegetation or shade.

Shade (a cooling bias) will inevitably affect poorly sited stations more than those that are well sited: The poorer sited stations are often shaded by nearby structures which result in their poor rating in the first place. Therefore, if anything, not accounting for shade would most likely lessen the differences between the better and poorer sites rather than increase them. Ground vegetation (a warming bias), on the other hand, affects the better sites, particularly stations located in rural areas, rather than the poorer and urban sites. Therefore, not accounting for vegetation may well lessen the differences between good and bad sites rather than increase them. Therefore we can be reasonably certain that excluding these factors will not bias this study in ways that will exaggerate the differences between well and poorly sited stations.

In any event, with the resources currently available, we are unable to rate either shade or ground cover adequately. Perhaps this will be addressed in a future study (including factors such as terrain and altitude). We can, however, quite accurately determine heat sink coverage by use of satellite and aerial imagery and in particular, that of Google Earth aerial photography and its distance measurement tool.

2.3. Methods of Analysis

The purpose of this study is to determine whether and to what extent regional and national-scale temperatures and temperature trends estimated from poorly-sited stations differ from those estimated from well-sited stations, by building on what was learned from Menne et al. (2010), Fall et al. (2011, and Muller et al. (2012) and by applying the new Leroy (2010) rating system against the stations surveyed by Watts (2009). The analysis involves aggregating USHCNv2 monthly station data into regional and national averages and comparing values obtained from different population groups of stations.

The process is started by computing monthly anomalies relative to a 30-year baseline period, in this case 1979-2008, to be consistent for comparison with previous works of Menne et al. (2010), Fall et al. (2011), and Muller et al. (2012). We then average the monthly anomalies across all stations in a particular Leroy (2010) class or set of classes within each of the nine NCDC-defined climate regions shown in Figure 2. In Figure 2, without separating any classes of stations to provide a baseline for the CONUS, the raw data average of all rated stations in each region shows a positive trend ranging from 0.173°C/decade in the Northwest region to 0.380 °C/decade in the Southwest region, with a continental United States (CONUS) gridded value of 0.231°C/decade.

Further investigations include separating stations by classes, and then examining the effect on the trends of the difference between classes for Tmin, Tmax, and Tmean, including examinations of rural and urban stations, stations at airports versus the general station population, and differences in station equipment. Finally, an overall average value for the (CONUS) is computed as a gridded, area-weighted mean of the regional averages for each of the station siting classes and subsets of siting classes, examining rural and urban, airport and non-airport stations, and equipment differences between stations using Cotton Region Shelters (CRS) and Maximum-Minimum Temperature System (MMTS) electronic thermometers.

The multiple regional analyses presented are designed to account for the spatial variations of the background climate and the variable number of stations within each region, so that the national analysis is not unduly influenced by data from an unrepresentative but data-rich corner of the United States. Figure 3 shows station distributions by class in the CONUS.

Menne et al. (2010) used a gridded analysis approach for the CONUS, as in our study. However, compared to the Menne et al. (2010) results, as well as the Muller (2012) results, both of which found very little difference between well sited and poorly sited stations in the CONUS, our gridded results based on the Leroy (2010) site ratings yields national trend values for all well sited (compliant classes 1&2) stations of 0.155°C /decade trend, while the poorly sited (non-compliant classes 3,4,5) stations show a 0.248°C/decade trend. Even greater and more significant differences are seen in the regional, environmental, class, and station type specific analyses we completed.

The results of the analysis suggest that these differences may be due specifically to the station siting characteristics or be due to other characteristics that covary with station siting, such as instrument type. Siting differences directly affect temperature trends if the poor siting compromises trend measurements or if changes in siting have led to artificial discontinuities. In what follows, to the extent that significant differences are found among classes, the well sited stations will be assumed to have more accurate measurements of temperature and temperature trends than poorly sited stations.

3. Results

3.1. Regional trend analysis

Figure 4 shows regional decadal trends in the CONUS for 1979-2008 as calculated with USHCNv2 data from all stations and all classes of stations. Clear statistically significant differences between Class 1&2 (compliant) and Class 3,4,5 (non-compliant) stations are indicated in the bar graphs. Without exception, in each region, compliant stations have a lower decadal scale trend than non-compliant stations. In the most striking example of this difference, one region, the SE, a slight negative trend exists for compliant stations of -0.02°C/decade while non-compliant stations have a positive trend of 0.223°C/decade. For the entire CONUS, the average of all regions shows the compliant Class 1&2 stations have a decadal scale trend of 0.155°C/decade while non-compliant Class 3,4,5 stations have a 0.248 °C/decade trend. Fully adjusted USHCNv2 data for the entire CONUS (all classes of stations) has a 0.309 °C/decade trend.

When USHCNv2 stations located at airports are considered, such differences between poor and well sited stations were observed to grow even larger. Figure 5 shows that when airport stations are excluded for the CONUS analysis, compliant stations have a 0.124°C/decade trend while non-compliant stations are almost double the trend at 0.246°C/decade. The difference in the SE region grew even larger with compliant stations having a -0.131°C/decade trend while non-compliant stations have a 0.219 °C/decade trend for a difference of 0.350°C/decade. Again, for all classes of stations, in all nine regions considered, compliant stations have a lower decadal scale trend than non-compliant stations. Conversely when only USHCNv2 stations sited at airports are considered these differences are not as strong as seen in Figure 6. Part of the differences may be attributed to the way equipment is deployed, sited, and maintained at airports. May airports, due to the weather stations being placed on grassy areas in between runways, are rated as "compliant" by both Leroy (1999) and Leroy (2010) rating systems. However, the data from airport stations is logged with aviation monitoring systems known as ASOS, from OFCM, (1994), and it has been demonstrated by error reports, such as in the Senate testimony of Snowe (1998) stating *"The ASOS systems in Maine have been very unreliable, The station in Houlton recorded more than 1400 mistakes in one year"* that the ASOS system has significant reporting problems particularly with the HO-83 hygrothermometer used in the ASOS system. Problems with temperature biases in the HO-83 hygrothermomter were first reported in Gall et al. (1992) in connection with large errors in the Tucson ASOS station. They report that in Tucson, an all-time maximum temperature record was set of 114°F, along with numerous daily records during the time this ASOS station was in use, many of these records having been set while no other records were broken within 1000 miles of Tucson.

In response to issues raised by Gall et al. (1992), ASOS hygrothermometers were redesigned. In Jones and Young, 1995 they reported:

"Examination of differences between the two instruments found that the original version of the HO-83 read approximately 0.6 deg C warmer than the redesigned instrument. Significant changes in the differences between the two instruments were noted between winter and summer. It is suggested that for stations with climatology similar to the ones used in this study monthly mean temperatures reported by the original version of the HO-83 be adjusted by adding -0.4 deg C to June, July August and Sept observations and by adding -0.7 deg C for the remainder of the year."

Karl et al. (1995) noted issues with the HO-83 hygrothermometer in Chicago in relation to reporting temperatures during a summer heat wave. In Karl and Knight, (1996) it was further discussed:

"Karl et al. (1995) show that, on average, the HO-83 increased the maximum temperature by about 0.5°C relative to the HO-63 instrument and also increased the minimum but only by 0.1°C. Much larger effects have been noted in Tucson, for example (Gall et al. 1992), and Jones and Young (1995) also find a consistent positive bias at

several stations they examined in the southern and central plains. This suggests that the trends of maximum T in Chicago are biased warm not only due to increased urbanization but by the introduction of the HO-83 instrument in 1986."

In the Snowe, 1998 testimony before the Senate, concerns over ASOS station data reliability were great enough to cause this amendment to be added to the bill being discussed:

"The administrator of the Federal Aviation Administration shall not terminate human weather observers for Automated Surface Observation System Stations until (1) The secretary of transportation determines that the system provides consistent reporting of changing meteorological conditions and notifies the Congress in writing of that determination; and (2) 60 days have passed since the report was submitted to the Congress."

The issues of reliability, combined with known historical problems with airport ASOS station instrumentation introducing positive temperature biases into the record, taken along with our findings that airport stations add a warm bias to our own siting analysis, suggests that airport weather stations utilizing ASOS may produce artificially high and uncorrected temperature records, and thus may not be suitable for inclusion into long term climate data without detailed retroactive examinations of instrument maintenance and calibration records and corrections applied to the daily data.

In the data shown in this study, airport stations clearly have a less definitive siting bias signal. This should not be taken as a suggestion that airport stations have better siting overall, as Watts 2009 demonstrated that many airport ASOS stations were near runways and tarmac, but that due to errors and inconsistencies in the ASOS temperature instrumentation, the temperature data may not accurately reflect the station siting bias issues due to being swamped by the larger errors of ASOS instrumentation.

Further analysis of the USHCNv2 data, taking into account rural stations, and excluding airports demonstrates even stronger bias magnitudes between compliant and non-compliant stations. In figure 7, the CONUS Class 1&2 trend for rural stations without airports is observed to be an even lower value at 0.108°C/decade, with Class 3,4,5 non-compliant stations having more than double that value at 0.228°C/decade. The Class difference in the SE region is -0.100 for compliant stations, with non-compliant stations at 0.157°C/decade for a difference of 0.257°C/decade between compliant and non-compliant stations. This is in stark contrast to figure 6, using airport stations only, where the SE region negative shows a positive trend of 0.181°C/decade. These findings further suggest that airports are not representative recorders of regional climatic trends.

The regional examination of classes that demonstrate the lowest decadal scale trend of all subsets, that of rural MMTS stations, excluding airports, reveals some of the most significant differences in siting biases between compliant and non-compliant stations.

Figure 8 shows that rural MMTS stations, excluding small rural airports that are sometimes equipped with MMTS equipment, have the lowest decadal trends of all classes and subsets of stations. The difference between compliant at -0.207°C/decade and non-compliant stations at 0.113°C/decade, in the SE region grows to 0.310°C/decade, with two additional regions, ENC and WNC, now showing negative decadal scale trends of -0.125°C/decade and -0.055°C/decade respectively with each showing large differences with their non-compliant station counterparts. The ENC region now records the largest regional scale difference between compliant and non-compliant stations in the entire USHCNv2 dataset at 0.365°C/decade.

The gridded average of all compliant Class 1&2 stations in the CONUS is only slightly above zero at 0.032°C/decade, while Class 3,4,5 non-compliant stations have a trend value of 0.212°C/decade, a value nearly seven times larger. NOAA adjusted data, for all classes of rural non-airport stations has a value of 0.300°C/decade nearly ten times larger than raw data from the compliant stations.

These large differences demonstrated between regional and CONUS trends accomplished by removal of airports and choosing the rural subset of stations to remove any potential urbanization effects suggests that rural MMTS stations not situated at airports may have the best representivity of all stations in the USHCNv2.

3.2. Temperature bias analysis by site classification and equipment type.

Significant decadal trend differences were observed between compliant CRS stations and compliant MMTS stations, with MMTS stations generally being cooler, confirming what was observed in Menne et al (2010). But, this effect is swamped by the larger effect of siting bias in the non-compliant stations, particularly in the trends of the Tmin, suggesting a sensitivity to heat sinks within the thermometer viewshed, which is the basis of the Leroy classification system. In Watts 2009 it was observed that with the introduction of the MMTS electronic thermometers in the NOAA COOP network starting in 1983, difficulties in trenching past obstacles (sidewalks, driveways, roadways, etc.) due to cabling, placed MMTS thermometers closer to offices and domiciles of the COOP observers. Our findings confirm this to have a real effect across all classes, with non-compliant MMTS stations having warmer trends. Additionally, it was observed that the Tmax trends of compliant CRS stations was significantly higher, suggesting that maintenance issues, such as paint deterioration over time and differences as discussed in Watts (2009), and seen in figure 9 darkened the wood, and lowered the surface albedo of the CRS equipped stations, making them more susceptible to solar insolation effects near the time of Tmax.

3.2.1 Comparison by site classifications

For the CONUS, we compare the average temperature of each Leroy (2010) class with the average of each of the other classes within each grid. This results in these baseline comparisons.

Figure10: Class 4 Comparisons (with each other Class ratings within each grid, then all results are averaged). Figure11: Class 3 Comparisons Figure12: Class 1 & 2 Comparisons.

The results are listed in order of robustness: There are more Class 4 stations than any other rating, so the Class 4 comparisons are examined first, followed by Class 2, then Class 1&2 stations. There is insufficient CONUS grid box coverage of Class 5 stations to use them as a baseline for a gridded comparison.

In figure 10, the columns represent the following measurements:

Class 1&2 compared with Class 4 within each grid box. The resulting differences for each Class 1&2 station are then averaged. Class 3 is compared with Class 4,

Class 4 compared with Class 4 (the baseline, so the result will be 0.), Class 5 compared with Class 4 and all lower classes.

Note that the well sited stations (Class 1 & 2) show a substantial difference in the Tmean trend compared with poorly sited stations. As reported in Fall et al. (2010), the difference is most significant in terms of Tmin. Tmax shows a very similar pattern to Tmin, although the differences are smaller.

Note also that while all classes of stations higher than Class 1&2 demonstrate higher trends than nearby Class 3&4 stations, Class 5 stations appear to be overwhelmed with waste heat which appears to be masking the trend. Note also that for Class 5 stations we observe a reversal of the decadal trend for Tmax and Tmin compared to all other classes of stations. We posit that this reflects the thermal latency of nearby heat sinks and sources for Class 5 stations that are applying a dampening effect on the surface layer thermometer, limiting its sensitivity to the surface layer atmosphere diurnal range. Such an effect would be demonstrated by a reversal of trends as heat sinks in the immediate proximity of Class 5 stations, such as concrete, asphalt, and buildings, dump stored heat from daytime solar insolation into the nighttime Tmin period, buffering the minimum temperature. Conversely, during the day, a large area of nearby heat sinks can act as solar radiation absorbers, buffering the ability of the local surface atmosphere to reach a representative Tmax compared to nearby stations. The overall result would be higher absolute temperatures, but, at the same time, lower temperature trends.

Both of these observations

1.) Poorly sited stations show greater trend results than well sited stations.
2.) Class 5 stations show smaller increases in trend results, which effect is possibly due to overwhelming by waste heat.

Will be either supported or disputed by the many various comparisons which follow.
In Figure 11, gridded with a Class 3 baseline, we see the same pattern as in Figure 10 observing that Tmean trend is indisputably higher for well sited stations than for poorly sited stations. Tmin shows the greatest differences between station classes, while Tmax shows the smallest.

In figure12, Class 1 & 2 gridded comparisons, we observe that all remaining non-compliant classes of stations, and the non-compliant grouped class3\4\5 have higher decadal trends than the compliant stations of Class 1&2. As in figures 10 and 11, Tmin shows the greatest differences between station classes, while Tmax shows the smallest.

3.2.2 Equipment Comparisons

We next examine whether these differences are an artifact of equipment or whether they hold true for both MMTS and CRS stations.

The USHCNv2 ground level photographic survey of Watts (2009), plus subsequent re-surveys and new station surveys for the purpose of this study reveal that the great majority of USHCNv2 stations consists of either CRS or MMTS equipped stations. There is a smaller number of airport based ASOS/AWOS stations and a very small population of non-standard equipment, such as consumer grade weather stations approved for use at a private station by the local National Weather Service Office COOP manager. The population of USHCN stations equipped with consumer grade instrumentation is too small to provide a statistically meaningful comparison and is ignored for the purposes of this study.

For purposes of evaluation in this study, we are classifying as MMTS any station which converted to MMTS prior to 1995, (and/or has been MMTS for a plurality of the study period), and the same applies for ASOS/AWOS. We classify as CRS any station which converted to MMTS (or other non-CRS equipment) in 1995 or later.

Comparing equipment alone, we observe in figure 13 that ASOS stations equipped with electronic hygrothermometers, such as the problematic HO-83, have the highest raw (ungridded) Tmean trends at 0.277 °C/decade, followed by CRS equipped stations at 0.265 °C/decade, and MMTS equipped stations at 0.192 °C/Decade. MMTS equipped stations are observed to have significantly lower Tmean trends than the two other equipment types.

This is of particular importance, considering that ASOS/AWOS systems are by far the better sited systems. 57% of rated ASOS/AWOS systems are Class 1\2, as opposed to 23% of CRS stations and a mere 14% of MMTS.

In order to demonstrate that these differences are a result of equipment bias and not actually a sign that poorly sited stations tend to show a smaller Tmean warming trend, we examine Class 1&2, plus Class 3, 4, and 5 stations for MMTS and CRS equipped stations. There is not a significant enough population of ASOS equipped USHCN stations for a statistically significant gridded comparison and that comparison is not done for that reason.

The following set of figures shows the gridded comparisons of each, calculated using the same method as for figures 10 through 12.

In figures 14 and 15 above, showing gridded comparison of CRS and MMTS equipped stations, respectively, we observe the same basic Tmean pattern for both sets of equipment. The only difference is that Class 5 CRS stations have a lower comparative difference than do Class 5 MMTS stations. In the case of both MMTS and CRS, well sited (Class 1&2) stations show a significantly smaller trend compared with poorly sited (Class 3,4,5) stations. Furthermore, in the case of MMTS stations (the most prevalent station type), the difference is about twice as great as for the CRS stations.

Another question that arises is whether microsite differences are masked by mesosite considerations of rural vs. urban environment.

To examine this question, we first look at overall mesosite trends for all stations, and then for Class 1&2, Class 3, Class 4, and Class 5 stations. For purposes of mesosite classification, we use the terms provided by NASA Goddard Institute for Space Studies for their GISTEMP database: Urban, Semi-Urban, Rural. Shown in Figure 16 is a six panel comparison showing comparisons for Urban, Semi-Urban, Rural stations with raw and adjusted data for all stations, raw and adjusted data for Class 1&2 stations, raw and adjusted data for Class 3,4,5 stations, raw and adjusted data for Class 3 stations, raw and adjusted data for Class 4 stations, and raw and adjusted data for Class 5 stations.

We observe that for the Tmin value identified in Fall et al. (2011) as being the most affected by siting issues, significant differences exist in Tmin raw data between urban and rural compliant Class 1&2 stations and between urban and rural non-compliant Class 3,4,5 stations. Rural Class 1&2 stations have a Tmin raw trend of 0.127°C/decade while urban stations have a Tmin raw trend of 0.278°C/decade. Rural Class 3,4,5 stations have a Tmin raw trend of 0.278°C/decade, while urban Class 3,4,5 stations have a Tmin raw trend of 0.420°C/decade, the highest in the dataset. This suggests that no matter what the microsite level issues of siting, urban sited stations are proportionately more affected in the Tmin by the mesoscale heat sinks and sources that make up urbanity. When looking at the Tmin USHCNv2 adjusted data for rural stations, we observe that it is adjusted higher in value, from 0.127°C/decade to 0.249°C/decade, effectively doubling the trend, and with that adjustment very nearly matches the rural Class 3,4,5 Tmin adjusted value of 0.265°C/decade. This suggests that USHCNv2 data homogenization methods are erroneously adjusting pristine Tmin data from rural Class 1&2 stations to be similar to that of rural Class 3,4,5 stations, effectively eliminating the preferred station representivity defined by Leroy (2010) .

In order to demonstrate that microsite considerations prevail, regardless of mesosite condition, we examine Class 1&2, Class 3, Class 4, and Class 5 averages (ungridded) for Rural, Semi-urban, and Urban environments in figure 17

This confirms that the microsite conditions we are seeing remain consistent in Rural and Semi-urban settings. In urban settings (10% of all stations), the differences are somewhat masked, especially in the case of Class 5 stations.

This is consistent with the hypothesis that artificially heated areas tend to overwhelm microsite considerations after a certain point. Note that urban Class 5 stations have the lowest trend, and that rural Class 4 stations have a lower trend than urban Class 4 stations as they are beginning to be overwhelmed by heat sink/source effects as well. This is further supported by the observation that the behavior of Class 5 stations in non-urban settings parallels the behavior of Class 4 stations in urban settings.

3.2.3 Discussion of Adjustments

Finally, just to confirm our overall findings, we present in figure 18 the USHCNv2 raw and adjusted gridded average for all stations. We do this by simply averaging each Class of station within each of our 26 grid boxes seen in figure 19 and then we average all the boxes for each Class of station. This removes the distribution bias, and is standard procedure for calculating temperature trends.

For all stations, Tmean trends are adjusted upwards from 0.23 °C per decade to 0.31°C per decade, an increase of 35%.

One will note that the adjusted Tmean trends “correct” the inequities caused by microsite quality – not by adjusting the poorly sited station trends down, to match the well sited stations, but by adjusting the well sited station trends upward by 92% to match the poorly sited stations. The poorly sited stations are adjusted warmer by 23%, as well.

After these adjustments, Tmean trends from poorly and well sited stations match almost exactly. This suggests that much of the representivity for well sited stations defined by Leroy (2010) are being discarded in adjustment processes.

In figure 20, the differences in regional and gridded CONUS decadal scale trends between all compliant, all non-compliant, and final NOAA USHCNv2 adjusted data for the CONUS are illustrated. The compliant thermometers (Class 1&2) have a trend value of 0.155°C/decade, the non-compliant thermometers (Class 3,4,5) have a trend value of 0.248°C/decade, and the NOAA final adjusted USHCNv2 data have a trend value of 0.309°C/decade, nearly double that of all compliant thermometers in the CONUS.

This disparity suggests that a combination of siting issues and adjustments are creating a spurious doubling of the U.S. surface temperature record for the 30 year period of this study. When rural, non-airport stations are considered, the CONUS trend is almost one third that of the NOAA adjusted record.

3.2.4 Statistical Significance Testing

In order to separately assess the effects of ratings, urbanization, equipment, max-min and region, a random effects model was constructed using the R-package lme4 by Pinheiro et al., (2012) as follows:

(1) trend~(1|ratings)+(1|Type)+(1|equipment)+(1|max)+(1|Grid)

where ratings is a factor with two classes: “compliant”= Class 1-2 and “non-compliant”= Class 3-5; Type is a factor for urbanization with three classes: R(ural); S(mall); U(rban); equipment is a factor with three classes: MMTS, CRS, and ASOS, max is a factor with three classes: max, min, mean; and Grid is a factor with 26 classes each representing a geographic region.

The base model considered the network of 779 stations with valid metadata (as defined above), less four stations with “other” equipment, reducing the base network slightly to 775 stations. Trends were calculated using “raw” USHCN v2 data.

The base model was compared to random effects models leaving each random effect out one by one using an anova test. Each random effect was highly significant as summarized in table 3.

The difference between trends for “compliant” and “non-compliant” stations was 0.105°C/decade; between rural and urban stations was 0.066°C/decade; between min and max measurements was 0.090°C/decade (max has lower trend) and between non-MMTS and MMTS approximately 0.06°C/decade (MMTS cooler), as shown in the figure 21

When a similar analysis was carried out on USHCN v2 adjusted data, the random effects for rating urbanization and equipment were completely eliminated; none were statistically significant, as seen in Figure 22. The sign of the max-min random effects was reversed. The fixed effect for adjusted data was 0.31°C/decade (as compared to 0.25°C/decade for raw data.)

Our interpretation of these results is that the USCHNv2 adjustment method from Menne et al (2009) is over-homogenizing the data and, in the process, removing statistically significant and important information. Because of the interaction between max-min, urbanization and rating, the following variation of the above was used to illustrate the interaction (see Figure 23):

(1) trend~(1|max:ratings:Type)+(1|equipment)+(1|Grid)

The left panel shows trends for Class 1-2 (“compliant”) stations by urbanization class for trends for max, min and mean temperatures. The right panel shows the same information for Class 3-5 (non-compliant) stations. The trends for maximum and minimum temperatures for compliant stations are virtually identical for each of the three urbanization classes, with a difference of about 0.092°C/decade between rural and urban stations. In contrast, non-compliant stations show a dramatic difference between trends of maximum and minimum temperatures of approximately 0.14°C/decade, in accordance with previous observations.

3.2.5 Reconciliation to NOAA and NASA GISS

The trend (mean temperatures) for “compliant” rural stations is 0.16°C/decade, substantially less than corresponding trends for the corresponding periods for the continental U.S. as calculated by NOAA (0.31°C/decade) and by GISS (0.31°C/decade). These values are identical to the fixed effect using adjusted USHCN data, also 0.31°C/decade as noted above. Both NOAA and GISS indices use adjusted USHCN data in their calculations. Both NOAA and GISS estimates more or less correspond to trends from non-compliant stations. Berkeley (BEST), Muller et al. (2012) adjustment methodology is substantially similar to USHCN adjustment methodology and accordingly yields almost identical results to NOAA.

GISS formerly (prior to the present version) ran noticeably cooler in the continental U.S. than NOAA (or CRU). This was because their prior methodology did not use USHCN adjusted data; GISS instead established trends from a network of "rural" stations (as defined by nightlights) using less processed USHCN data. This method (as noted in online discussions at the time) yielded trends more similar to that from "compliant" stations in the surface stations study. GISS' adoption of USHCN adjusted data therefore appears to be a retrogression in their analysis.

Within "compliant" stations, the effect of urbanization is as expected and ranges from 0.11 to 0.14°C/decade. Similarly, the effect of ratings on rural stations is directionally as expected at the outset of the surface stations project but with a marked interaction with max-min: the effect of ratings is much stronger with minimum temperatures (0.15°C/decade) than for maximum temperatures (only 0.03°C/decade), in line with the emphasis of Christy et al (2008) on maximum temperatures as an indicator.

By way of comparison, the University of Alabama Huntsville (UAH) Lower Troposphere CONUS trend over this period is 0.25°C/decade and Remote Sensing Systems (RSS) has 0.23°C/decade, the average being 0.24°C/decade. This provides an upper bound for the surface temperature since the upper air is supposed to have larger trends than the surface (e.g. see Klotzbach et al (2011). Therefore, the surface temperatures should display some fraction of that 0.24°C/decade trend. Depending on the amplification factor used, which for some models ranges from 1.1 to 1.4, the surface trend would calculate to be in the range of 0.17 to 0.22, which is close to the 0.155°C/decade trend seen in the compliant Class 1&2 stations.

4. Discussion and Conclusions

The analysis demonstrates clearly that siting quality matters. Well sited stations consistently show a significantly cooler trend than poorly sited stations, no matter which class of station is used for a baseline, and also when using no baseline at all.

Statistically significant differences between compliant and non-compliant stations exist, as well as urban and rural stations. We have demonstrated evidence that USCHNv2 adjustments are over-homogenizing the data and, in the process, removing statistically significant and important information.

It is demonstrated that stations with poor microsite (Class 3, 4, 5) ratings have significantly higher warming trends than well sited stations (Class 1, 2): This is true for, in all nine geographical areas of all five data samples. The odds of this result having occurred randomly are quite small.

It is demonstrated that stations with poor mesosite (airports and urbanized areas) show an increase in temperature trends of both well and poorly microsited stations, alike. Over a third of all stations are located in a poor mesosite environment. This is extremely unrepresentative of the topography the stations purport to represent. Poor mesosite has its greatest effect on Class 1, 2 stations (over 40% spurious exaggeration of trend), as so many of them are located in airports.

Well sited stations, using a localized Class 4 (the most common class) baseline show a trend of 0.09°C per decade lower than poorly sited stations for raw Tmean trends. The Raw Tmean trend for well sited stations is 0.14°C per decade lower than adjusted Tmean trend for poorly sited stations.

Not only does the NOAA USCHNv2 adjustment process fail to adjust poorly sited stations downward to match the well sited stations, but actually adjusts the well sited stations upwards to match the poorly sited stations.

In addition to this, it is demonstrated that urban sites warm more rapidly than semi-urban sites, which in turn warm more rapidly than rural sites. Since a disproportionate percentage of stations are urban (10%) and semi-urban (25%) when compared with the actual topography of the U.S., this further exaggerates Tmean trends.

NOAA adjustments procedure fails to address these issues. Instead, poorly sited station trends are adjusted sharply upward (not downward), and well sited stations are adjusted upward to match the already-adjusted poor stations. Well sited rural stations show a warming nearly three times greater after NOAA adjustment is applied. We have shown that the site-classification value is a clear factor in the calculation of the trend magnitude. We are investigating other factors such as Time-Of-Observation changes which for the adjusted USHCNv2 is the dominant adjustment factor during 1979-2008.

Future investigations could test to see if the siting issue is broader. Given that USHCN stations overlap and are a part of the GHCN, the siting issue should be examined for all of the GHCN and BEST sites used in Muller (2012).

Class 5 sites, even more so than Class 3 and 4, have a multitude of major non-climatic effects and local microclimate which result making it difficult, if not impossible, to explain the behavior of its trend signal. This includes shading from buildings and trees, cooling of dry bulb temperatures by evaporation from grasses around the site in otherwise dry vegetation areas, their location on roof tops with more wind ventilation, etc. There is also the likelihood of more evaporation of water vapor into the air such as from water treatment plants and non-representative nearby vegetation such as lawns and shrubs.

In future analyses, the assessment of moist enthalpy trends could provide more insight. As shown in Pielke et al (2004), Davey et al (2006), Fall et al (2010), and Peterson et al (2011) concurrent long term trends in the absolute humidity of the surface air make the interpretation of the dry bulb temperature trend more difficult. However, it is the combined effect of dry bulb temperature and absolute humidity that are the true measure of heating and cooling.

As shown in Figure 11 in Pielke et al (2007), for example, the hottest time of the day in the dry bulb temperature is not the hottest in the physics unit of heat (i.e. Joules per kg of air). It could be that in the urban area the added water vapor from those sites could be resulting in really warm conditions in terms of Joules per kg, but the dry bulb temperature is suppressed. This certainly could be true around sites at water treatment plants, of which a significant population exists in the USHCN.

There is the further issue of equipment inhomogeneity. Modern MMTS sensors show a significantly lower warming trend than the obsolete CRS shelters. Yet rather than lowering the trends of CRS stations, the trends of MMTS stations are sharply adjusted upwards. It is difficult, however, to be certain of the true effect thanks to the relatively small number of Class 1,2, rural, non-airport stations.

Taken *in toto*, these factors identified in this study have led to a spurious doubling of U.S. Tmean trends from 1979 - 2008.

Acknowledgments

The authors wish to acknowledge the many cooperative observers that unselfishly carry out COOP observations which are the backbone of climate monitoring. We also acknowledge the many volunteers that made the surfacestations.org project possible with their personal time and efforts in gathering the nationwide survey. Special thanks are given to these prominent volunteers who expended special efforts and expertise in metadata collection and collation: Gary Boden, Kristen Byrnes, Don and Liz Healy, Eric Gamberg, John Goetz, Don Kostuch, Steven McIntyre, John Slayton, Ted Semon, Russell and Ellen Steele, and Barry Wise. Acknowledgement is given to former California State Climatologist James Goodridge, who was inspirational with surveys he made of California COOP stations during his tenure.

Special thanks are given to Dr. Roger Pielke Sr. for inspiration, advice, and technical proofreading of this study.

References

Asaeda T., Ca V. T., Wake A., 1996; Heat storage of pavement and its effect on the lower atmosphere. *Atmospheric environment*. CUTEST '92 : conference on the urban thermal environmental studies in Tohwa, Fukuoka, JAPON (07/09/1992) 1996, vol. 30, n^{o} 3 (185 p.) (30 ref.), pp. 413-427

Baker, D. G., 1975: Effect of observation time on mean temperature estimation. *Journal of Applied Meteorology*, **14**, 471-476.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850, *Journal of Geophysical Research*, **111**, D12106, doi:10.1029/2005JD006548.

Christy, J. R., 2002: When was the hottest summer? A State Climatologist struggles for an answer. *Bulletin of American Meteorological Society,* **83,** 723-734.

Christy, J. R., W. Norris, K. Redmond, and K. Gallo, 2006: Methodology and results of calculating central California surface temperature trends: Evidence of a human induced climate change. *Journal of Climate,* **19,** 548-563.

Christy, J.R., W.B. Norris, and R.T. McNider, 2009: Surface Temperature Variations in East Africa and Possible Causes. *Journal of Climate*, **22**, 3342–3356.

Davey, C.A., and R.A. Pielke Sr., 2005: Microclimate exposures of surface-based weather stations - implications for the assessment of long-term temperature trends. *Bulletin of American Meteorological Society*, **86**, 497-504.

Davey, C.A., R.A. Pielke Sr., and K.P. Gallo, 2006: Differences between near-surface equivalent temperature and temperature trends for the eastern United States - Equivalent temperature as an alternative measure of heat content. *Global and Planetary Change*, **54**, 19–32.

DeGaetano, A.T., 2006: Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate*, **19**, 838-853.

Easterling, D. R., T. R. Karl, E.H. Mason, P. Y. Hughes, and D. P. Bowman, 1996: *United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data*. ORNL/CDIAC-87, NDP-019/R3. Carbon Dioxide Information

Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

Easterling, D. R., B. Horton, P. D. Jones, T. C. Peterson, T. R. Karl, D. E. Parker, M. J. Salinger, V. Razuvayev, N. Plummer, P. Jamason, C. K. Folland, 1997: Maximum and minimum temperature trends for the globe. *Science*, **277**, 364-367.

Fall, S., D. Niyogi, R. A. Pielke Sr., A. Gluhovsky, E. Kalnay and G. Rochon, 2009: Impacts of land use land cover on temperature trends over the continental United States: assessment using the North American Regional Reanalysis, *International Journal of Climatology*, DOI: 10.1002/joc.1996.

Fall, S., N. Diffenbaugh, D. Niyogi, R.A. Pielke Sr., and G. Rochon, 2010: Temperature and equivalent temperature over the United States (1979 – 2005). *Int. J. Climatol.*, DOI: 10.1002/joc.2094.

Fall, S., Watts, A., Nielsen-Gammon, J. Jones, E. Niyogi, D. Christy, J. and Pielke, R.A. Sr., 2011, Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends, *Journal of Geophysical Research*, **116**, D14120, doi:10.1029/2010JD015146, 2011

Fourier, J. B., 1822, Theorie analytique de la chaleur, Paris; Freeman, A., 1955, translation, Dover Publications, Inc, NY. Available online at: http://www.math.umass.edu/~lr7q/ps_files/finalfourier.pdf

Gall, R, K. Young, R. Schotland, and J. Schmitz , 1992. *The Recent Maximum Temperature Anomalies in Tueson: Are They Real or an Instrumental Problem?* Journal of Climate Volume 5, Issue 6 (June 1992) pp. 657-665

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, *Journal of Geophysical Research*, **106**, 23947–23963.

Hubbard K.G., and X. Lin, 2006. Reexamination of instrument change effects in the U.S. Historical Climatology Network. *Geophysical Research Letters* **33**: L15710, DOI: 10.1029/2006GL027069.

Imhoff, M.L., W. T. Lawrence, D. C. Stutzer, and C. D. Elvidge, 1997. A Technique for Using Composite DM SP/OLS "City Lights" Satellite Data to Map Urban Area. *Remote Sensing of Environment* **61**, 361-370.

Jones, Cyrus G. and Young, Kenneth C., An Investigation of Temperature Discontinuities Introduced by the Installation of the HO-83 Thermometer *Journal of Climate* Volume 8, Issue 5 (May 1995) pp. 139-140

Kalnay, E. and M. Cai, 2003: Impact of urbanization and land-use change on climate. *Nature* **423**: 528– 531.

Kalnay, E., M. Cai, H. Li, and J. Tobin, 2006. Estimation of the impact of land-surface forcings on temperature trends in eastern United States. *Journal of Geophysical Research* **111,** (D06106) 1-13.

Kalnay, E., M. Cai, M. Nunez, and Y. Lim, 2008. Impacts of urbanization and land surface changes on climate trends. *International Association for Urban Climate* **27**, 5–9.

Karl, T. R., G. Kukla, and J. Gavin, 1984: Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. *J. Climate Appl. Meteor.*, **23**, 1878-1886.

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, *Journal of Climate and Applied Meteorology*, **25**, 145-160.

Karl, T.R., and C.N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology*, **26**, 1744-1763.

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

Karl, T. R., J. D. Tarpley, R. G. Quayle, H. F. Diaz, D. A. Robinson, and R. S. Bradley, 1989: The recent climate record: What it can and cannot tell us, *Rev. Geophys.*, **27**, 405–430, doi:10.1029/RG027i003p00405

Karl, T.R., C.N. Williams, Jr., F.T. Quinlan, and T.A. Boden, 1990: *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data*, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.

Karl, T. R., and Coauthors, 1993: Asymmetric trends of daily maximum and minimum temperature. *Bull. Amer. Meteor. Soc.*, **74**, 1007-1023.

Karl, T.R., and Coauthors, 1995: Critical issues for long-term climate monitoring. *Climate Change*, **31**, 185-221.

Karl, Thomas R., Knight, Richard W. 1997: The 1995 Chicago Heat Wave: How Likely Is a Recurrence?. *Bull. Amer. Meteor. Soc.*, **78**, 1107–1119.

Klotzbach, P.J., R.A. Pielke Sr., R.A. Pielke Jr., J.R. Christy, and R.T. McNider, 2009: An alternative explanation for differential temperature trends at the surface and in the lower troposphere. *J. Geophys. Res.*, 114, D21102, doi:10.1029/2009JD011841.

Klotzbach, P.J., R.A. Pielke Sr., R.A. Pielke Jr., J.R. Christy, and R.T. McNider, 2010: Correction to: "An alternative explanation for differential temperature trends at the surface and in the lower troposphere. J. Geophys. Res., 114, D21102, doi:10.1029/2009JD011841", *J. Geophys. Res*., 115, D1, doi:10.1029/2009JD013655.

Leroy, M., 1999: *Classification d'un site*. Note Technique no. 35. Direction des Systèmes d'Observation, Météo-France, 12 pp.

Leroy, M., 2010: Siting Classification for Surface Observing Stations on Land, Climate, and Upper-air Observations *JMA/WMO Workshop on Quality Management in Surface*, Tokyo, Japan 27-30 July 2010

Lee, Seri, 1995 : Optimum design and selection of heat sinks, *Semiconductor Thermal Measurement and Management Symposium*, 1995. SEMI-THERM XI., Eleventh Annual IEEE, 7-9 Feb 1995

Lin, X., R.A. Pielke Sr., K.G. Hubbard, K.C. Crawford, M. A. Shafer, and T. Matsui, 2007: An examination of 1997-2007 surface layer temperature trends at two heights in Oklahoma. *Geophysical Research Letters*, 34, L24705, doi:10.1029/2007GL031652.

Mahmood, R. L., S. A. Foster, and D. Logan, 2006: The GeoProfile metadata, exposure of instruments, and measurement bias in climatic record revisited. *International Journal of Climatology*, **26,** 1091-1124.

Mahmood, R., R.A. Pielke Sr., K.G. Hubbard, D. Niyogi, G. Bonan, P. Lawrence, B. Baker, R. McNider, C. McAlpine, A. Etter, S. Gameda, B. Qian, A. Carleton, A. Beltran-Przekurat, T. Chase, A.I. Quintanar, J.O. Adegoke, S. Vezhapparambu, G. Conner, S. Asefi, E. Sertel, D.R. Legates, Y. Wu, R. Hale, O.W. Frauenfeld, A. Watts, M. Shepherd,

C. Mitra, V.G. Anantharaj, S. Fall, R. Lund, A. Nordfelt, P. Blanken, J. Du, H.-I. Chang, R. Leeper, U.S. Nair, S. Dobler, R. Deo, and J. Syktus, 2010: Impacts of land use land cover change on climate and future research priorities. *Bulletin of the American Meteorological Society*, 91, 37–46, DOI: 10.1175/2009BAMS2769.1

Martinez, C.J., Maleski, J.J., Miller, M.F, 2012: Trends in precipitation and temperature in Florida, USA. *Journal of Hydrology*. volume 452-453, issue , year 2012, pp. 259 – 281

McNider, R.T., G.J. Steeneveld, B. Holtslag, R. Pielke Sr, S. Mackaro, A. Pour Biazar, J.T. Walters, U.S. Nair, and J.R. Christy, 2012: Response and sensitivity of the nocturnal boundary layer over land to added longwave radiative forcing. *J. Geophys. Res.*, doi:10.1029/2012JD017578, in press.

Montandon, L.M., S. Fall, R.A. Pielke Sr., and D. Niyogi, 2011: Distribution of landscape types in the Global Historical Climatology Network. *Earth Interactions*, 15:6, doi: 10.1175/2010EI371

Menne, M.J., and C.N. Williams, Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, **18**, 4271- 4286.

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 American Meteorological Society*, **90**, 993-1007, doi: 10.1175/2008BAMS2613.1.

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

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

Mitchell, T.D. and P.D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**, 693-712

Muller, R.A., Curry, J., Groom, D. Jacobsen, R.,Perlmutter, S. Rohde, R. Rosenfeld, A., Wickham, C., Wurtele, J., 2012: Earth Atmospheric Land Surface Temperature and Station Quality in the United States. *JGR Special Publication The Third Santa Fe Conference on Global and Regional Climate Change manuscript number 12JD018146* http://berkeleyearth.org/pdf/berkeley-earth-station-quality.pdf

NOAA/NESDIS, 2002: *Climate Reference Network Site Information Handbook*. National Oceanic and Atmospheric Administration, NOAA-CRN/OSD-2002-0002R0UD0. 19 pp. Available online at: ftp://ftp.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X030FullDocumentD0.pdf

OFCM, 1994: Office of the Federal Coordinator for Meteorology, *Federal Standard for Siting Meteorological Sensors at Airports*, FCM-S4-1994, Washington, D.C.August 1994 Chap. 2- Sensor Exposure, Section 2.6, Temperature and Dew Point Sensors, Available online at: http://www.ofcm.gov/siting/text/a-cover.htm

Peterson, T.C., and D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series, *International Journal of Climatology*, **14**, 671–679.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Bohm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Foland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker,D. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18**: 1493-1517

Peterson, T.C., 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. *Journal of Climate*, **16**, 2941–2959.

Peterson, T.C., 2006: Examination of potential biases in air temperature caused by poor station locations. *Bull. Amer. Meteor. Soc.*, **87**, 1073-1080.

Peterson, T. C., K. M. Willett, and P. W. Thorne (2011), Observed changes in surface atmospheric energy over land, *Geophys. Res. Lett.*, **38**, L16707, doi:10.1029/2011GL048442

Pielke, R.A., Sr., T. Stohlgren, L. Schell, W. Parton, N. Doesken, K. Redmond, J. Money, T. McKee, and T.G.F. Kittel., 2002: Problems in evaluating regional and local trends in temperature: An example from eastern Colorado, USA. *International Journal of Climatology*, **22**, 421–434.

Pielke, R.A. Sr., K. Wolter, O. Bliss, N. Doesken, and B. McNoldy, 2007: The July 2005 Denver heat wave: How unusual was it? *Nat. Wea. Dig*., **31**, 24-35.

Pielke Sr., R.A., C. Davey, and J. Morgan, 2004: Assessing "global warming" with surface heat content. *Eos*, **85**, No. 21, 210-211.

Pielke Sr., R.A. J. Nielsen-Gammon, C. Davey, J. Angel, O. Bliss, N. Doesken, M. Cai, S. Fall, D. Niyogi, K. Gallo, R. Hale, K.G. Hubbard, X. Lin, H. Li, and S. Raman, 2007a: Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bulletin of American Meteorological Society*, **88,** 913-928.

Pielke Sr., R.A., C. Davey, D. Niyogi, S. Fall, J. Steinweg-Woods, K. Hubbard, X. Lin, M. Cai, Y.-K. Lim, H. Li, J. Nielsen-Gammon, K. Gallo, R. Hale, R. Mahmood, S. Foster, R.T. McNider, and P. Blanken, 2007b: Unresolved issues with the assessment of multi-decadal global land surface temperature trends. *Journal of Geophysical Research*, **112,** D24S08, doi:10.1029/2006JD008229.

Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar and the R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104. Available online at: http://cran.r-project.org/web/packages/nlme/

Quayle, R.G., D.R. Easterling, T.R. Karl, and P.Y. Hughes, 1991, Effects of recent thermometer changes in the Cooperative Station Network, *Bulletin of American Meteorological Society*, 72, 1718– 1723.

Reeves, J., J. Chen, X.L. Wang, R. Lund, and Q.Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **46**, 900-914.

Runnalls, K.E. and Oke, T.R. 2006. A technique to detect microclimatic inhomogeneities in historical records of screen-level air temperature. *Journal of Climate*, **19**, 959-978.

Snowe, Olympia J., Senator from Maine, Testimony before the U.S. Senate on Automated Surface Observations System Stations, Congressional Record, V. 144, Pt. 15, September 22, 1998 to September 26, 1998, page 158, Accessible online at : http://books.google.com/books?id=MfrwH4TB93QC&lpg=PA158&ots=kEEsz2_h9k&dq=ASOS%20errors%20Congressional%20record&pg=PA158#v=onepage&q&f=false

Stone, D. A. and A. J. Weaver , 2003: Factors contributing to diurnal temperature range trends in the twentieth and twenty first century simulations of the CCAma coupled model, *Climate Dynamics*, **20**, 435-445.

Stott P. and P. Thorne; 2010: Proposals for surface-temperature data bank now open for scrutiny *Nature*, **466**, 1040-1040

Thorne, P.W., D.E. Parker, J.R. Christy, and C.A. Mears, 2005: Uncertainties in climate trends: Lessons from upper-air temperature records, *Bulletin of American Meteorological Society*, **86**, 1437-1442.

Vose, R.S., C.N. Williams Jr., T.C. Peterson, T.R. Karl, and D.R. Easterling, 2003: An evaluation of the time of observation bias adjustment in the U.S. Historical Climatology Network, *Geophysical Research Letters*, **30**, 2046, doi:10.1029/2003GL018111.

Vose, R. S., D. R. Easterling, T. R. Karl, and M. Helfert, 2005a: Comments on "Microclimate exposures of surface-based weather stations." *Bulletin of American Meteorological Society*, **86,** 504–506.

Vose, R. S., D. R. Easterling, and B. Gleason, 2005b: Maximum and minimum temperature trends for the globe: An update through 2004. *Geophys. Res. Lett.*, **32**, L23822, doi:10.1029/2005GL024379.

Walters, J. T., R. T. McNider, X. Shi, and W. B. Norris, 2007: Positive surface temperature feedback in the stable nocturnal boundary layer. *Geophysical Research Letters*, doi:10.1029/2007GL029505.

Watts, A., 2009: Is the U.S. surface temperature record reliable? The Heartland Institute, Chicago, IL. 28 pp.

Willmot, C.J., S.M. Robeson, and J.J. Feddema, 1991: Influence of spatially variable instrument networks on climatic averages, *Geophysical Research Letters*, **18**, 2249–2251.

World Meteorological Organization Commission for Instruments and Methods of Observation, Fifteenth session, (CIMO-XV, 2010) WMO publication Number 1064, available online at: http://www.wmo.int/pages/prog/www/CIMO/CIMO15-WMO1064/1064_en.pdf

Yilmaz, H., Toy, S., Irmak, M.A., Yilmaz, S., Bulit, Y., 2008: Determination of temperature differences between asphalt concrete, soil, and grass surfaces of the City of Ezurum, Turkey. *Atmosfera*, **21**, 135-146