

## Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends

Souleymane Fall,<sup>1</sup> Anthony Watts,<sup>2</sup> John Nielsen-Gammon,<sup>3</sup> Evan Jones,<sup>2</sup> Dev Niyogi,<sup>4</sup> John R. Christy,<sup>5</sup> and Roger A. Pielke Sr.<sup>6</sup>

Received 5 October 2010; revised 26 March 2011; accepted 6 May 2011; published 30 July 2011.

[1] The recently concluded Surface Stations Project surveyed 82.5% of the U.S. Historical Climatology Network (USHCN) stations and provided a classification based on exposure conditions of each surveyed station, using a rating system employed by the National Oceanic and Atmospheric Administration to develop the U.S. Climate Reference Network. The unique opportunity offered by this completed survey permits an examination of the relationship between USHCN station siting characteristics and temperature trends at national and regional scales and on differences between USHCN temperatures and North American Regional Reanalysis (NARR) temperatures. This initial study examines temperature differences among different levels of siting quality without controlling for other factors such as instrument type. Temperature trend estimates vary according to site classification, with poor siting leading to an overestimate of minimum temperature trends and an underestimate of maximum temperature trends, resulting in particular in a substantial difference in estimates of the diurnal temperature range trends. The opposite-signed differences of maximum and minimum temperature trends are similar in magnitude, so that the overall mean temperature trends are nearly identical across site classifications. Homogeneity adjustments tend to reduce trend differences, but statistically significant differences remain for all but average temperature trends. Comparison of observed temperatures with NARR shows that the most poorly sited stations are warmer compared to NARR than are other stations, and a major portion of this bias is associated with the siting classification rather than the geographical distribution of stations. According to the best-sited stations, the diurnal temperature range in the lower 48 states has no century-scale trend.

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

### 1. Introduction

[2] As attested by a number of studies, near-surface temperature records are often affected by time-varying biases. Among the causes of such biases are station moves or relocations, changes in instrumentation, changes in observation practices, and evolution of the environment surrounding the

station such as land use/cover change [e.g., Baker, 1975; Karl and Williams, 1987; Karl et al., 1988, 1989; Davey and Pielke, 2005; Mahmood et al., 2006, 2010; Pielke et al., 2007a, 2007b; Yilmaz et al., 2008; Christy et al., 2009]. Maximum and minimum temperatures are generally affected in different ways. Such inhomogeneities 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]. Even if stations are initially placed at pristine locations, the surrounding region can develop over decades and alter the footprint of these measurements.

[3] To address such problems, climatologists have developed various methods for detecting discontinuities in time series, characterizing and/or removing various nonclimatic biases that affect temperature records in order to obtain homogeneous data and create reliable long-term time series [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,

<sup>1</sup>College of Agricultural, Environmental and Natural Sciences and College of Engineering and Physical Sciences, Tuskegee University, Tuskegee, Alabama, USA.

<sup>2</sup>IntelliWeather, Chico, California, USA.

<sup>3</sup>Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA.

<sup>4</sup>Indiana State Climate Office, Department of Agronomy and Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana, USA.

<sup>5</sup>Department of Atmospheric Science, University of Alabama in Huntsville, Huntsville, Alabama, USA.

<sup>6</sup>CIRES/ATOC, University of Colorado at Boulder, Boulder, Colorado, USA.

2005; Brohan *et al.*, 2006; DeGaetano, 2006; Runnalls and Oke, 2006; Reeves *et al.*, 2007; Menne and Williams, 2009]. Overall, considerable work has been done to account for inhomogeneities and obtain adjusted data sets for climate analysis.

[4] However, there is presently considerable debate about the effects of adjustments on temperature trends [e.g., Willmott *et al.*, 1991; Balling and Idso, 2002; Pielke *et al.*, 2002; Peterson, 2003; Hubbard and Lin, 2006; DeGaetano, 2006; Lin *et al.*, 2007; Pielke *et al.*, 2007a, 2007b]. Moreover, even though detailed history metadata files have been maintained for U.S. stations [Peterson *et al.*, 1998], many of the aforementioned changes often remain undocumented [Christy, 2002; Christy *et al.*, 2006; Pielke *et al.*, 2007a, 2007b; Menne *et al.*, 2009]. Because of the unreliability of the metadata the adjustment method for the United States Historical Climatology Network, Version 2 (USHCNv2) seeks to identify both documented and undocumented changes, with a larger change needed to trigger the adjustment when the possible change is undocumented [Menne *et al.*, 2009; Menne and Williams, 2009]. The adjustment of undocumented changes represents a tradeoff between leaving some undocumented changes uncorrected and inadvertently altering true local climate signals.

[5] The National Climatic Data Center (NCDC) has recognized the need for a climate monitoring network as free as possible from nonclimatic trends and discontinuities and has developed the United States Climate Reference Network (USCRN) to fill this need. The USCRN goal is a highly reliable network of climate observing stations that provide “long-term high quality observations of surface air temperature and precipitation that can be coupled to past long-term observations for the detection and attribution of present and future climate change” [National Oceanic and Atmospheric Administration and National Environmental Satellite, Data, and Information Service (NOAA and NESDIS), 2002; Leroy, 1999]. The station sites have been selected based on the consideration of geographic location factors including their regional and spatial representativity, 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.

[6] While the USCRN network, if maintained as planned, will provide the benchmark measurements of climate variability and change within the United States going forward, the standard data set for examination of changes in United States temperature from 1895 to the present is the USHCNv2. USHCNv2 stations were selected from among Cooperative Observer Network (COOP) stations based on a number of criteria including their historical stability, length of record, geographical distribution, and data completeness. The USHCNv2 data set has been “corrected to account for various historical changes in station location, instrumentation, and observing practice” [Menne *et al.*, 2009], and such adjustments are reported to be a major improvement over those used to create the previous version of the USHCN data set [Easterling *et al.*, 1996; Karl *et al.*, 1990]. Nonetheless, the stations comprising the USHCNv2 data set did not undergo the rigorous site selection process of their USCRN counterparts and do not generally have redundant temperature sensors that permit intercomparison in the event of instrument changes.

[7] Prior to the USCRN siting classification system, there existed the NOAA “100 foot rule” (NOAA Cooperative Observer Program, Proper siting, <http://web.archive.org/web/20020619233930/http://weather.gov/om/coop/standard.htm>, 2002; NOAA, Cooperative Observer Program, Proper siting: Temperature sensor siting, <http://www.nws.noaa.gov/os/coop/standard.htm>, 2009, accessed 30 September 2010) which stated: “The sensor should be at least 100 feet from any paved or concrete surface.” This was to be applied to all NOAA Cooperative Observer Program stations (COOP), which includes the special USHCN station subset. The genesis of this specification is rooted in the Federal Standard for Siting Meteorological Sensors at Airports [Office of the Federal Coordinator for Meteorological Services and Supporting Research, 1994, chap. 2, p. 4], which states that “The sensors will be installed in such a position as to ensure that measurements are representative of the free air circulating in the locality and not influenced by artificial conditions, such as large buildings, cooling towers, and expanses of concrete and tarmac. Any grass and vegetation within 100 feet (30 meters) of the sensor should be clipped to height of about 10 inches (25 centimeters) or less.” Prior to that, siting issues are addressed in the National Weather Service Observing Handbook No. 2 [National Weather Service (NWS), 1989, p. 46], which states that “The equipment site should be fairly level, sodded, and free from obstructions (exhibit 5.1). It should be typical of the principal natural agricultural soils and conditions of the area... Neither the pan nor instrument shelter should be placed over heat-absorbing surfaces such as asphalt, crushed rock, concrete slabs or pedestals. The equipment should be in full sunlight during as much of the daylight hours as possible, and be generally free of obstructions to wind flow.” One purpose of these siting criteria is to eliminate artificial temperature biases from man-made surfaces, which can have quite a large effect in some circumstances [e.g., Yilmaz *et al.*, 2008].

[8] The interest in station exposure impacts on temperature trends has recently gained momentum with the completion of the USHCNv2 station survey as part of the Surface Stations Project [Watts, 2009]. The survey was conducted by more than 650 volunteers who visually inspected the USHCNv2 stations and provided site reports that include an extensive photographic documentation of exposure conditions for each surveyed station. The documentation was supplemented with satellite and aerial map measurements to confirm distances between sensors and heat sources and/or sinks. Based on these site reports, the Surface Stations Project classified the siting quality of individual stations using a rating system based on criteria employed by NOAA to develop the USCRN.

[9] This photographic documentation has revealed wide variations in the quality of USHCNv2 station siting, as was first noted for eastern Colorado stations by Pielke *et al.* [2002]. It is not known whether adjustment techniques satisfactorily compensate for biases caused by poor siting [Davey and Pielke, 2005; Vose *et al.*, 2005a; Peterson, 2006; Pielke *et al.*, 2007b]. A recent study by Menne *et al.* [2010] used a preliminary classification from the Surface Stations Project, including 40% of the USHCNv2 stations. Approximately one third of the stations previously classified as good exposure sites were subsequently reevaluated and

**Table 1.** Climate Reference Network Classification for Local Site Representativity<sup>a</sup>

Class	Description
1	Flat and horizontal ground surrounded by a clear surface with a slope below 1/3 (<19°). Grass/low vegetation ground cover <10 centimeters high. Sensors located at least 100 meters from artificial heating or reflecting surfaces, such as buildings, concrete surfaces, and parking lots. Far from large bodies of water, except if it is representative of the area, and then located at least 100 meters away. No shading when the Sun elevation >3 degrees.
2	Same as Class 1 with the following differences. Surrounding Vegetation <25 centimeters. Artificial heating sources within 30 m. No shading for a Sun elevation >5°.
3	(error 1°C) - Same as Class 2, except no artificial heating sources within 10 meters.
4	(error ≥2°C) - Artificial heating sources <10 meters
5	(error ≥5°C) - Temperature sensor located next to/above an artificial heating source, such a building, roof top, parking lot, or concrete surface

<sup>a</sup>The errors for the different classes are estimated values that represent the associated uncertainty levels. Source is Climate Reference Network, 2002.

found to be poorly sited. The reasons for this reclassification are explained in section 2. Because so few USHCNv2 stations were actually found to be acceptably sited, the sample size at 40% was not fully spatially representative of the continental USA. Menne et al. analyzed the 1980–2008 temperature trends of stations grouped into two categories based on the quality of siting. They found that a trend bias in poor exposure sites relative to good exposure ones 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 [Menne et al., 2010, p. 1] “no evidence that the CONUS temperature trends are inflated due to poor station siting.”

[10] In this study, we take advantage of the unique opportunity offered by the recently concluded survey with near-complete characterization of USHCNv2 sites by the Surface Stations Project to examine the relationship between USHCNv2 station siting and temperatures and temperature trends at national and regional scales. In broad outline, for both raw and adjusted data, we compare the maximum, minimum, mean, and diurnal range temperature trends for the United States as measured by USHCN stations grouped according to CRN site ratings. A secondary purpose is to use the North American Regional Reanalysis [NARR] [Mesinger et al., 2006] as an independent estimate of surface temperatures and temperature trends with respect to station siting quality.

## 2. Data and Methods

### 2.1. Climate Data

[11] The USHCNv2 monthly temperature data set is described by Menne et al. [2009]. The raw or unadjusted (unadj) data has undergone quality control screening by NCDC but is otherwise unaltered. The intermediate (tob) data has been adjusted for changes in time of observation such that historical observations are consistent with current observational practice at each station. The fully adjusted

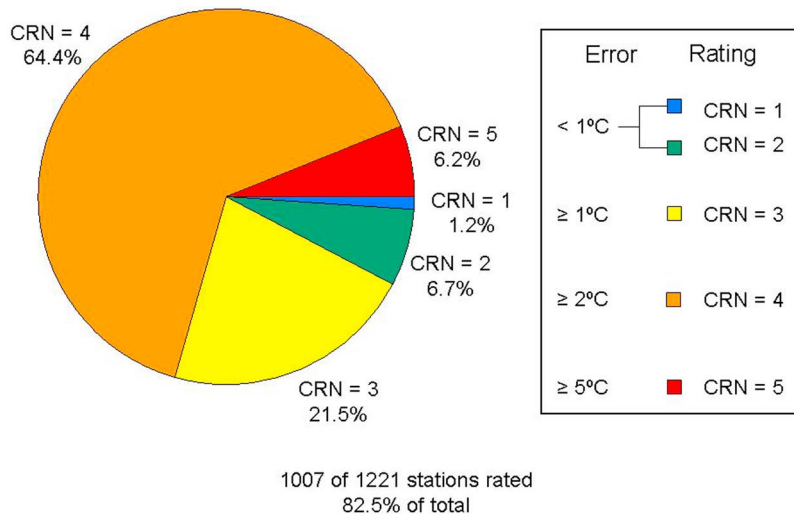
(adj) data has been processed by the algorithm described by Menne and Williams [2009] to remove apparent inhomogeneities where changes in the temperature record at a station differ significantly from those of its neighbors. Unlike the unadj and tob data, the adj data is serially complete, with missing monthly averages estimated through the use of data from neighboring stations.

### 2.2. Station Site Classification

[12] We make use of the subset of USHCNv2 data from stations whose sites were initially classified by Watts [2009] and further refined in quality control reviews led by two of us (Jones and Watts), using the USCRN site selection classification scheme for temperature and humidity measurements [NOAA and NESDIS, 2002], originally developed by Leroy [1999] (Table 1). The site surveys were performed between 2 June 2007 and 23 February 2010, and 1007 stations (82.5% of the USHCN network) were classified (Figure 1). Any known changes in siting characteristics after that period are ignored.

[13] In the early phase of the project, the easiest stations to locate were near population centers (shortest driving distances), this early data set with minimal quality control had a disproportionate bias toward urban stations and only a handful of CRN1/2 stations existed in that preliminary data set. In addition, the project had to deal with a number of problems including (1) poor quality of metadata archived at NCDC for the NWS managed COOP stations; (2) no flag for specific COOP stations as being part of the USHCN subset; (3) some station observers not knowing whether their station was USHCN or not; and (4) NCDC-archived metadata often lagging station moves (when a curator died for example) as much as a year. As a result, the identification of COOP stations was difficult, sometimes necessitating resurveying the area to get the correct COOP station that was part of the USHCN network. Whenever it was determined that a station had been misidentified, the survey was done again. In January 2010, NCDC added a USHCN flag to the metadata description, making it easier to perform quality control checks for station identification. NCDC has also now archived accurate metadata GPS information for station coordinates, making it possible to accurately check station placement using aerial photography and Google Earth imagery. Three quality control passes to ensure station identification, thermometer placement, and distances to objects and heat sinks were done by a two person team. The two quality control team members had to agree with their assessment, and with the findings of the volunteer for the station. If not, the station was assigned for resurvey and then included if the resurvey met quality control criteria. At present, the project has surveyed well in excess of 87% of the network, but only those surveys that met quality control requirements are used in this paper, namely 82.5% of the 1221 USHCN stations.

[14] In addition to station ratings, the surveys provided an extensive documentation composed of station photographs and detailed survey forms. The best and poorest sites consist of 80 stations classified as either CRN 1 or CRN 2 and 61 as CRN 5 (8% and 6% of all surveyed stations, respectively). The geographic distribution of the best and poorest sites is displayed in Figure 2 and sites representing each CRN class are shown in Figure 3. Because there are so few CRN 1

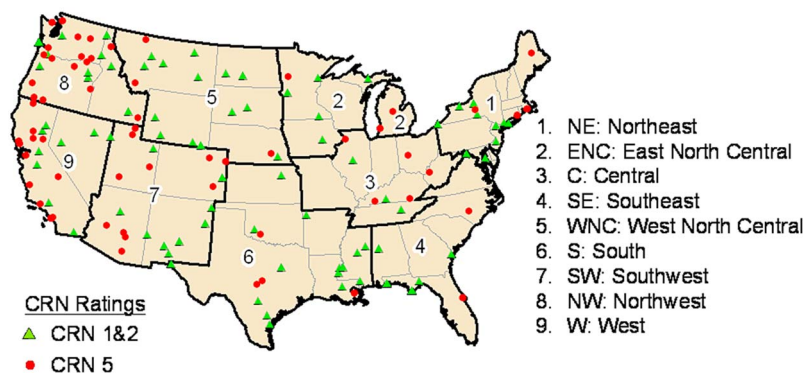


**Figure 1.** Surveyed USHCN surface stations. The site quality ratings assigned by the Surface Stations Project are based on criteria utilized in site selection for the Climate Reference Network (CRN). Temperature errors represent the additional estimated uncertainty added by siting [Leroy, 1999; NOAA and NESDIS, 2002].

sites, we treat sites rated as CRN 1 and CRN 2 as belonging to the single class CRN 1&2. These would also be stations that meet the older NOAA/NWS “100 foot rule” (~30 m) for COOP stations.

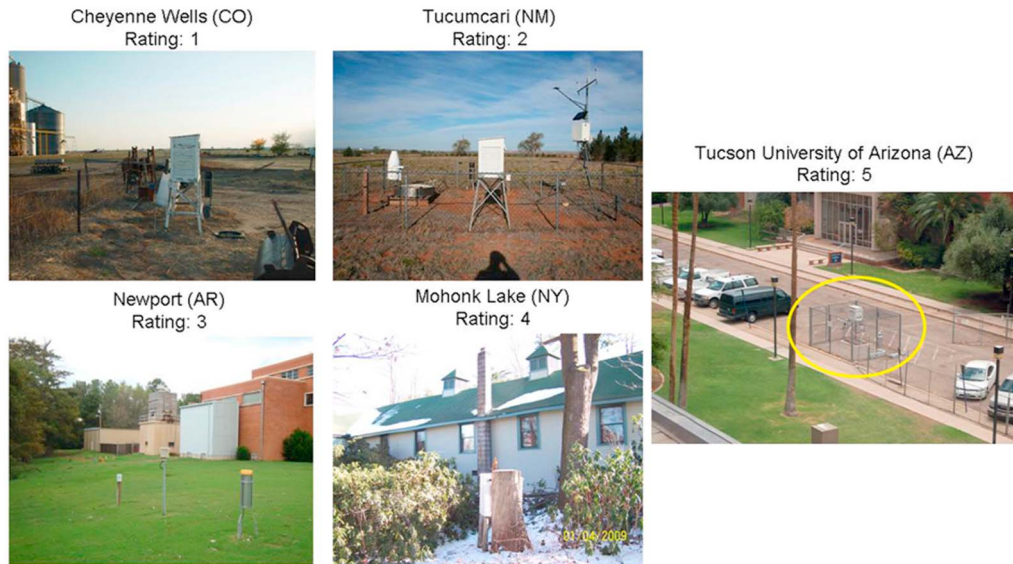
[15] The CRN 1&2 and CRN 5 classes are not evenly distributed across the lower 48 states or within many individual climate regions. In order to test the sensitivity of results to this uneven distribution, we create two sets of “proxy” stations for the CRN 1&2 and CRN 5 stations. The proxy stations are the nearest CRN 3 or CRN 4 class stations to the CRN 1&2 and CRN 5 stations, except that proxies must be within the same climate region and cannot simultaneously represent two CRN 1&2 or two CRN 5 stations. The proxy stations thus mimic the geographical distribution of the stations they are paired with. The CRN 1&2 proxies have a slightly greater proportion of CRN 3 stations than do the CRN 5 proxies (31% versus 26%), but this difference in siting characteristics is expected to be too small to affect the analyses.

[16] A match between temperatures or trends calculated from CRN 1&2 proxies and the complete set of CRN 3 and 4 stations implies that the irregular distribution of CRN 1&2 stations does not affect the temperature or trend calculations. Conversely, if the calculations using CRN 1&2 stations and CRN 1&2 proxy stations differ in the same manner from calculations using CRN 3 and 4 stations, geographical distribution rather than station siting characteristics is implicated as the cause of the difference between CRN 1&2 and CRN 3/CRN 4 calculations. Similar comparisons may be made between CRN 5 and CRN 3/CRN 4 using the CRN 5 proxies. Differences between CRN 1&2 and CRN 5 temperature and trend estimates are likely to be due to poor geographical sampling if their proxies also produce different temperature and trend estimates, while they are likely to be due to siting and associated characteristics if estimates from their proxies match estimates from the complete pool of CRN 3 and CRN 4 stations.



**Figure 2.** Distribution of good exposure (Climate Reference Network (CRN) rating = 1 and 2) and bad exposure (CRN = 5) sites. The ratings are based on classifications by Watts [2009] using the CRN site selection rating shown in Table 1. The stations are displayed with respect to the nine climate regions defined by NCDC.





**Figure 3.** U.S. Historical Climate Network (USHCN) station exposure at sites representative of each CRN class: CRN 1, a clear flat surface with sensors located at least 100 m from artificial heating and vegetation ground cover <10 cm high; CRN 2, same as CRN 1 with surrounding vegetation <25 cm and artificial heating sources within 30 m; CRN 3, same as CRN 2, except no artificial heating sources within 10 m; CRN 4, artificial heating sources <10 m; and CRN 5, sensor located next to/above an artificial heating source.

### 2.3. Methods of Analysis

[17] We are interested in whether and to what extent national-scale temperatures and temperature trends estimated from poorly sited stations differ from those estimated from well-sited stations. The analysis involves aggregating station data into regional and national averages and comparing values obtained from different populations of stations.

[18] We begin the aggregation process by computing monthly anomalies relative to the 30 year baseline period 1979–2008 except where noted. Small differences are obtained in unadj and tob by using a different baseline period, due to missing data. We then average the monthly anomalies across all stations in a particular CRN class or set of classes within each of the nine NCDC-defined climate regions shown in Figure 2. Finally, an overall average value for the contiguous 48 states is computed as an area-weighted mean of the regional averages.

[19] The regional analysis is 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. Note that there are at least two stations rated as CRN 1&2 and CRN 5 in each climate region.

[20] *Menne et al.* [2010] use a different aggregation approach, based on gridded analysis that accomplishes the same objective. When using Menne et al.’s station set, ratings, and normals period, our aggregation method yields national trend values that differ from theirs on average by less than 0.002°C/century.

[21] We further examine the relationship between station siting and surface temperature trends by comparing observed and analyzed (reanalysis) monthly mean temperatures. Fol-

lowing the initial work of *Kalnay and Cai* [2003] and *Kalnay et al.* [2006], recent studies have demonstrated that the National Center for Environmental Prediction (NCEP) reanalyses (Global Reanalysis and NARR) can be used as an independent tool for detecting the potential biases related to station siting [*Pielke et al.*, 2007a, 2007b; *Fall et al.*, 2010]. This approach, which is referred to as the “observation minus reanalysis” (OMR) method [*Kalnay and Cai*, 2003; *Kalnay et al.*, 2006], relies on the fact that land surface temperature observations are not included in the data assimilation process of some reanalyses such as the NCEP reanalyses which, therefore, are entirely independent of the USHCNv2 temperature observations. Moreover, as mentioned by *Kalnay et al.* [2008], this method separates surface effects from the greenhouse warming by eliminating the natural variability due to changes in atmospheric circulation (which are included in both surface observations and the reanalysis). As a result, the comparison between observation and reanalysis can yield useful information about the local siting effect on observed temperature records.

[22] Because surface data is not assimilated in the reanalysis, diurnal variations in near-surface temperatures in the reanalysis are largely controlled by model turbulent and radiative parameterizations. Such parameterizations, especially in the coarsely resolved nocturnal boundary layer, may have errors [*Walters et al.*, 2007] which may impact mean surface temperatures. In fact, *Stone and Weaver* [2003] and *Cao et al.* [1992] indicate that models have a difficult time replicating trends in the diurnal temperature range. Thus, while the reanalysis does not have surface siting contaminations, it may not have the true temperature trend. However, the NARR 2 m temperatures have generally smaller biases and more accurate diurnal temperature cycles

than previous reanalysis products, as shown by *Mesinger et al.* [2006] and in recent studies [e.g., *Pielke et al.*, 2007a].

[23] Because there may be systematic biases in both NARR temperatures and NARR temperature trends, the NARR temperatures are used here in a way that minimizes or removes the effect of such biases. The only assumption made is that any NARR temperature biases or temperature trend biases at USHCNv2 station locations are independent of the microscale siting characteristics of the USHCNv2 stations. This seems plausible: since USHCNv2 temperature data is not ingested into NARR, there is no way that NARR can be directly affected by USHCNv2 microscale siting characteristics.

[24] We bilinearly interpolate the NARR gridded mean temperatures for every month within the period 1979–2008 to USHCNv2 locations and subtract them from the USHCNv2 values of maximum, average, and minimum temperature prior to computation of monthly anomalies and aggregation. The resulting anomaly values are then aggregated to the lower 48 states in the same manner as described above for the USHCNv2 trends. This procedure effectively produces an estimate of the average temperature across the United States based upon particular classes of USHCNv2 observations, using the NARR mean temperatures as a first guess field. The NARR mean temperatures are not defined in the same way as USHCNv2 average temperatures. However, systematic differences among CRN classes in analyzed observed minus NARR temperatures will be due to the surface station characteristics, not to the NARR temperature computation. The possibility of random NARR errors producing a false signal is addressed through the Monte Carlo resampling described below, and the possibility of regional variations of NARR errors producing a false signal is addressed through comparisons with proxy stations.

[25] The statistical significance of differences in temperatures and temperature trends of stations within a particular target CRN class relative to CRN 1&2 is estimated using Monte Carlo resampling. The assignments of stations to the target class and to class 1&2 are permuted randomly, under the constraint that at least two stations of each class must remain in each NCDC climate region, and values recomputed. For example, there are 80 CRN 1&2 stations and 61 CRN 5 stations. These stations are randomly rearranged to produce random groups of 80 and 61 stations. The two groups are checked to see if there are at least two stations of each group within each climate region. If so, calculation of means and/or trends proceeds as described above for the true CRN classes. This procedure is carried out until there are 10,000 realizations of means and/or trends, and the null hypothesis that differences are independent of CRN classification is rejected at the 95% confidence level if the two-sided  $p$  value of the observed difference is less than 0.05.

[26] Differences that depend upon CRN classification 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. We plan to investigate

the various possible covarying causes for the dependence of climate data quality on siting classification in a separate paper.

### 3. Trend Analysis

[27] Figure 4 shows the ordinary least squares linear trends across the contiguous 48 states for 1979–2008 as estimated with data from the various classes of USHCNv2 stations. Also shown is the trend computed from all classes together, which tends to lie between the CRN 3 and CRN 4 trends because of the predominance of stations in those classes. Statistically significant differences relative to the CRN 1&2 trends are indicated by asterisks.

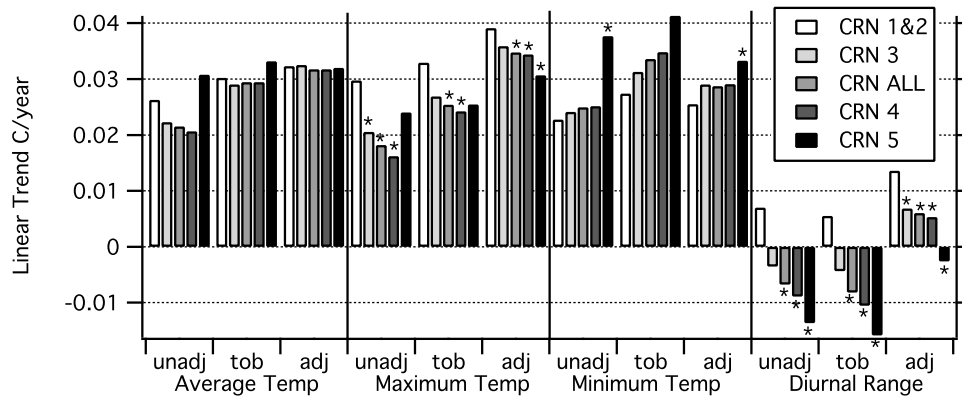
[28] The minimum temperature trend becomes progressively larger as the siting quality decreases, becoming statistically significant for the unadj and adj CRN 5 data. This is consistent with the analysis of *Runnalls and Oke* [2006] which indicated minimum temperatures were much more subject to change by microclimate influences. Consistent with a shift from late afternoon to morning observations, the tob adjustment increases the minimum temperature trends. The full adjustment (adj) is only able to reduce the trend differences between classes by about 50%.

[29] Maximum temperature trends are smaller with poorer siting quality, as found by *Menne et al.* [2010], but the decreasing trends are not monotonic with respect to siting except for the adj data. The unadj CRN 4 stations produce the smallest maximum temperature trend, and their difference with respect to CRN 1&2 is statistically significant at all levels of adjustment. Since CRN 4 stations make up the vast majority of all rated sites, the entire network also produces maximum temperature trends that are significantly different from the CRN 1&2 trends alone. The trend increase with the tob adjustment is again consistent with expectations. The full adjustment reduces the trend differences somewhat in most classes but does not eliminate them. Note that statistically significant differences are just as likely between classes of fully adjusted data even though the magnitude of the trend differences tends to be smaller.

[30] Relative to well-sited stations, the poorly sited stations have a more rapid minimum temperature increase and a less rapid maximum temperature increase. As a result, the difference in diurnal temperature range trends compared to CRN 1&2 is large and significant for most other classes of stations without adjustments and significant for all other classes after full adjustments. The linear trend of diurnal temperature range is almost unaffected by the tob adjustment, but the full adjustment increases the linear trend by about  $1^{\circ}\text{C}/\text{century}$  in most classes, making it positive for all but CRN 5. The magnitude of the linear trend in diurnal temperature range is over twice as large for CRN 1&2 ( $0.13^{\circ}\text{C}/\text{decade}$ ) as for any of the other CRN classes.

[31] Conversely, the differing trends in maximum and minimum temperature among classes cause the average temperature trends to be almost identical, especially for the fully adjusted data. In this case, no matter what CRN class is used, the estimated mean temperature trend for the period 1979–2008 is about  $0.32^{\circ}\text{C}/\text{decade}$ .

[32] The differences between CRN 3 and CRN 4 stations are small on average, and the question arises whether the differences between CRN 1&2 and CRN 5, and in turn the



**Figure 4.** Ordinary least squares linear trends of United States (contiguous 48 states) temperature, as estimated from data from USHCNv2 stations from the period 1979–2008 grouped by station siting classification. Trends are computed using raw or unadjusted data (unadj), data adjusted for time of observation (tob), and data with full inhomogeneity adjustments (adj). Diurnal range is the difference between maximum and minimum temperature, while average temperature is the average of the maximum and minimum temperatures. Asterisks indicate that trend differences relative to CRN 1&2 are significantly different at the 95% confidence level.

differences between these two classes and CRN 3&4, are an artifact of the uneven geographical distribution of the relatively small number of CRN 1&2 and CRN 5 stations. The number of stations in the two classes is not much smaller than what was shown to be adequate for decadal-scale climate monitoring in the United States [Vose and Menne, 2004], but that study assumed a near-uniform geographical distribution. Even though the Monte Carlo resampling randomizes the different geographical distributions, it is possible that the actual distributions are so peculiar that they artificially produce a statistically significant result. For example, Figure 2 shows that CRN 1&2 stations occupy a different portion of the NE climate region than do CRN 5 stations, and CRN 5 stations in the West Coast climate region tend to be concentrated along the coast.

[33] Figure 5 shows the 30 year trends of CRN 1&2, CRN 3&4, and CRN 5, as well as the 30 year trends of proxies to CRN 1&2 and CRN 5. The proxy networks (section 2.2) are composed entirely of CRN 3&4 stations but mimic the geographical distribution of the CRN 1&2 and CRN 5 stations. The unadj and tob trends estimated from the two proxy networks do not match the trends for CRN 3&4, so the combination of small sample size and irregular distribution almost certainly affects the trend estimates for CRN 1&2 and CRN 5 as well. In general, the difference between the CRN 1&2 proxy trends and the CRN 5 proxy trends has the same sign but smaller magnitude than the difference between the real CRN 1&2 trends and CRN 5 trends, implying that the geographical distribution of stations is contributing to the unadj and tob differences between CRN 1&2 and CRN 5. For the adj trends, though, the proxy trends are consistent with each other and with the CRN 3&4 trends, and differ from the CRN 1&2 and CRN 5 trends. This indicates that the adjusted trends are relatively insensitive to the geographical distribution and small number of stations, and that the trend differences between CRN 1&2 and CRN 5 arise from differences in the stations themselves.

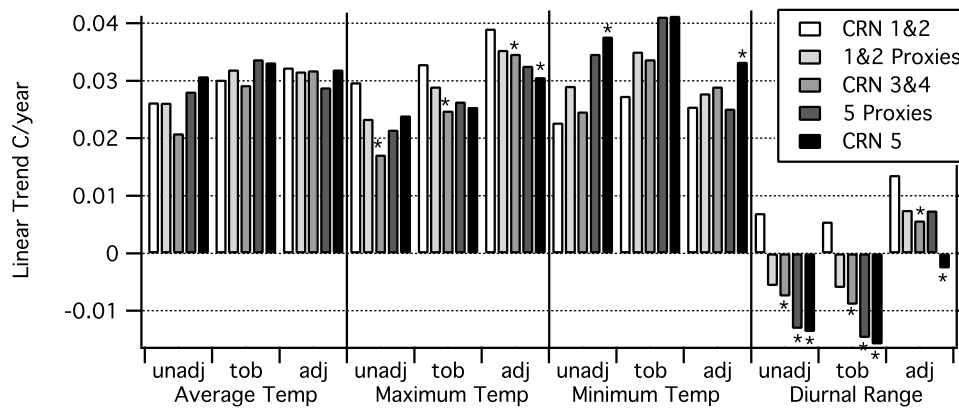
[34] The CRN 5 stations are particularly sparse and uneven across the central and eastern United States (see

Figure 2). Nonetheless, the fully adjusted trend differences between CRN 1&2 and CRN 5 computed using only stations in climate regions 1 through 6 are still large enough to be statistically significant, and analysis of CRN 5 proxies indicates that the trend differences are not attributable to the irregular spatial distribution of CRN 5 stations.

[35] Excluding nearly collocated stations, there are about 18 CRN 5 stations in the central and eastern United States, equivalent in density to a national network of about 25 stations. Vose and Menne [2004, Figure 9] found that a 25 station national network of COOP stations, even if unadjusted and unstratified by siting quality, is sufficient to estimate 30 year temperature trends to an accuracy of  $\pm 0.012^{\circ}\text{C}/\text{yr}$  compared to the full COOP network. The statistically significant trend differences found here in the central and eastern United States for CRN 5 stations compared to CRN 1&2 stations are as large ( $-0.013^{\circ}\text{C}/\text{yr}$  for maximum temperatures,  $+0.011^{\circ}\text{C}/\text{yr}$  for minimum temperatures) or larger ( $-0.023^{\circ}\text{C}/\text{yr}$  for diurnal temperature range).

[36] Figure 6 shows the 30 year trends computed from aggregations of different groups of stations. This grouping allows a direct comparison to Menne *et al.* [2010], and also tests possible divisions into well-sited and poorly sited stations. Because the United States trend estimates are computed from climate region trend estimates, the trends shown in Figure 6 are not simply linear combinations of the trends shown in Figure 4. Nonetheless, the picture is a similar one, and broadly confirms the more limited findings of Menne *et al.* [2010] that poorer-sited stations produce larger minimum temperature trends and smaller maximum temperature trends.

[37] The evolution of the maximum and minimum temperature differences is shown in Figure 7. In order to emphasize systematic differences among classes while eliminating noisy year-to-year climate variations, the interpolated monthly mean NARR temperatures are subtracted from the observed maximum and minimum temperatures prior to computing temperature anomalies. Since Figure 7 shows anomalies from individual time means, the instanta-



**Figure 5.** Linear trends of United States (contiguous 48 states) temperature, as in Figure 4, but for the station classifications shown. The proxy networks are nearest-neighbor networks to CRN 1&2 or CRN 5 stations and are composed of a mix of CRN 3 and CRN 4 stations. See text for details.

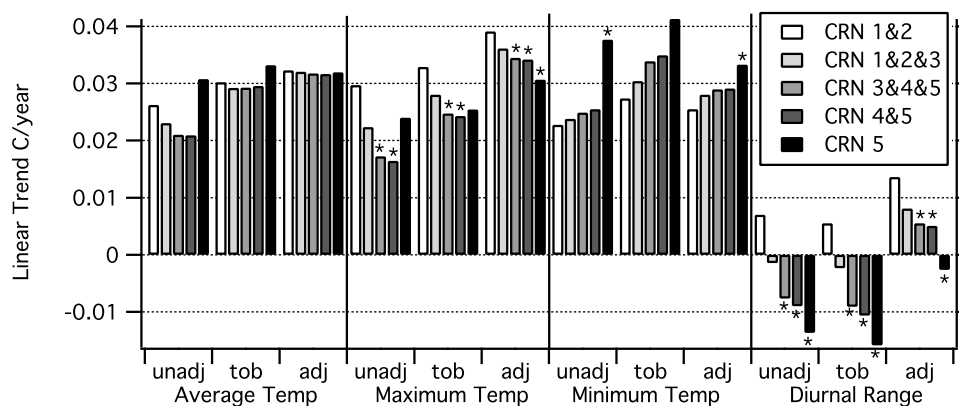
neous differences between classes are not as important as the changes in those differences over time. In particular, systematic differences among the different classes change dramatically around 1984–1987. This is coincident with a widespread transition to MMTS thermometers at most stations. In a field test, the MMTS recorded cooler maximum temperatures by about 0.4 C and warmer minimum temperatures by about 0.2 C than did a thermometer housed in a Cotton Region Shelter (CRS) [Wendland and Armstrong, 1993]. However, the instrumentation change at COOP stations was typically synchronous with a change in the siting characteristics [Menne et al., 2010]. The combined effect of the MMTS transition and simultaneous siting changes was an average maximum temperature decrease of 0.4 C and minimum temperature increase of 0.3 C relative to CRS stations [Quayle et al., 1991], similar to what might be expected from the instrumentation change alone, but the actual discontinuity varied widely from station to station [Hubbard and Lin, 2006], indicating that the microsite changes were similarly important on a station by station basis.

[38] Between-class temperature differences do not remain stable after the primary MMTS transition period. The extent to which this is due to siting classifications compared to imperfect estimations of United States temperature changes

is shown in Figure 8, which breaks down the difference between CRN 1&2 and CRN 5 temperatures as CRN 1&2 and their proxies, CRN 5 and their proxies, and the differences between the two sets of proxies. The latter difference should account for most of the representativeness error due to the small number of stations in each class and regionally dependent estimation errors associated with NARR, thereby isolating effects directly associated with siting. As with Figure 7, the important features are the changes in the differences over time.

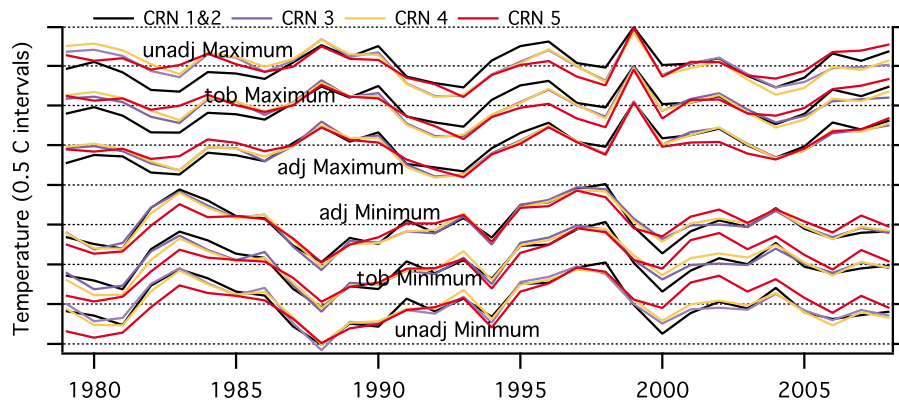
[39] Besides the transition in the mid-1980s discussed earlier, the difference in minimum tob temperatures estimated from the two CRN groups (Figure 8a) changes by about  $-0.2^{\circ}\text{C}$  around the year 2000. The mid-1980s change arises from both CRN 1&2/proxy differences and differences between the two proxy sets, while the 2000 change arises mainly from the CRN 5/proxy differences. The full adjustment (Figure 8b) reduces the change in differences, but the remaining change is primarily due to the CRN 5/proxy differences as the other two differences are stable over time. Most, but not all, of the CRN 5/proxy change remaining after full adjustments is in the mid-1980s.

[40] The difference in maximum tob temperatures (Figure 8c) has even more dramatic changes, climbing to a maximum in the mid-1990s and falling thereafter. Because the difference



**Figure 6.** Linear trends of United States (contiguous 48 states) temperature, as in Figure 4, but for the station classification groupings shown.

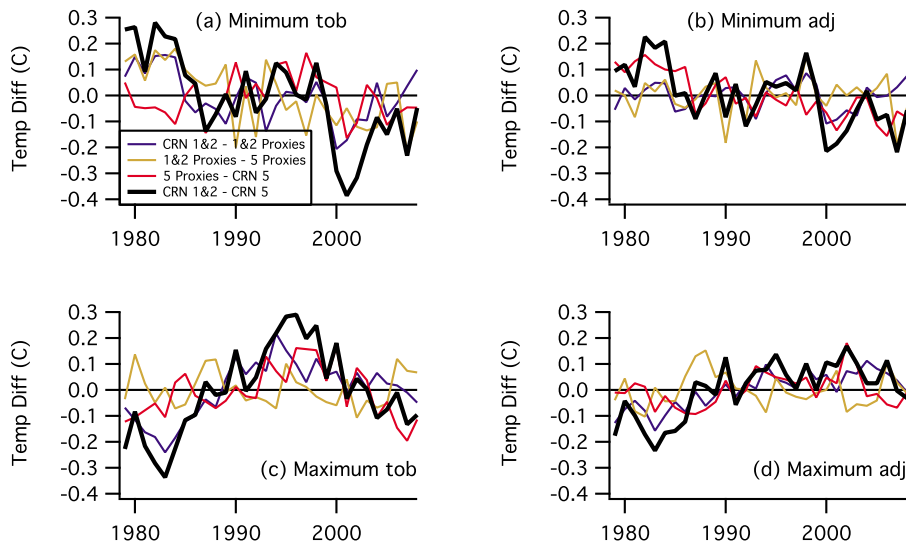




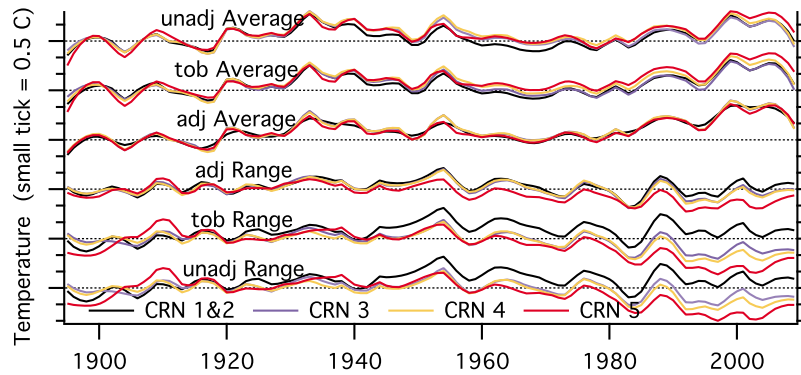
**Figure 7.** Maximum and minimum temperature anomaly differences for the United States (contiguous 48 states), as estimated using unadjusted and adjusted data from USHCNv2 stations in particular siting classes. The individual station anomalies are subtracted from the corresponding NARR annual mean temperature anomalies at the station locations in order to remove natural climate variability. The curves are graphed relative to their 30 year mean values (dotted lines), with the spacing between tick marks equal to 0.5°C.

between the two sets of proxies is stable over time, the changes cannot be attributed to estimation or representativeness errors. The rise receives its greatest contribution from the difference between CRN 1&2 and its proxies, while the fall receives its greatest contribution from the difference between CRN 5 and its proxies. The full adjustments (Figure 8d) virtually eliminate the fall since the mid-1990s, leaving as the primary change the one coincident with the MMTS transition in the 1980s, and the greatest contribution to the change comes from the difference between CRN 1&2 and its proxies.

[41] Collectively, Figures 7 and 8 show that most of the adj trend differences between CRN classes over the period 1979–2008 are due to uncorrected inhomogeneities around the time of the MMTS conversion. The diurnal temperature range differences can be inferred from the maximum and minimum temperature differences. According to Figure 7, the CRN 1&2 fully adjusted data show a large increase in diurnal temperature range, while the CRN 5 fully adjusted data show a small decrease in diurnal temperature range. According to Figure 8, most, but not all, of the difference in



**Figure 8.** Time series of the differences between United States temperature anomalies (relative to the 30 year means) estimated from CRN 1&2 and CRN 5 station networks using NARR temperatures as a first guess. Also shown is a breakdown of the differences into three additive components: the difference between estimates using CRN 1&2 and the proxy network for CRN 1&2, the difference between the estimates from the CRN 1&2 proxies and the CRN 5 proxies, and the difference between the estimates from the CRN 5 proxies and the CRN 5 stations. (a) Minimum temperatures, with time-of-observation corrections. (b) Fully adjusted minimum temperatures. (c) Maximum temperatures, with time-of-observation corrections. (d) Fully adjusted maximum temperatures.



**Figure 9.** Average temperature and diurnal temperature range anomalies for the United States relative to the 1895–1930 period, separated by siting classification and extent of adjustment. Curves have been smoothed using a loess filter to eliminate short-term climate variability and make differences between curves clearer. Each group of curves is offset from its neighbors by 1.5°C.

diurnal temperature range estimated from the two classes arises during the mid to late 1980s.

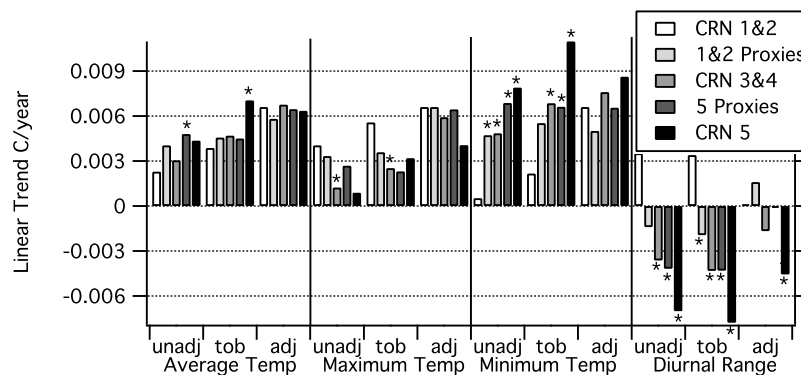
[42] The differences in temperature trends among USHCNv2 stations are not limited to the period 1979–2008. Figure 9 shows the average temperatures and average diurnal temperature ranges for the period of record, 1895–2009. Year-to-year variations have been removed with a loess filter, and curves are plotted relative to their 1895–1930 average, when the present-day siting classification should have the least relevance to siting quality.

[43] The average temperatures show the familiar long-term temperature variation pattern: general warming from 1895 to the mid-1930s, general cooling from the mid-1930s to the 1970s, and general warming since the 1970s. The full adjustments cause the smoothed average United States temperature anomalies estimated from the different classes to be within about 0.2°C of each other throughout the period of record, correcting a tendency for the more poorly sited stations to warm faster. The homogeneity corrections are not as successful in adjusting the diurnal temperature range. The large and systematic divergence of unadjusted diurnal temperature range among the various CRN classes is limited to CRN 5 by the full adjustments. There is one period of divergence around 1935–1950 and another, previously discussed, in the mid-1980s. The divergence around 1935–1950 is at least partially accounted for by errors in tem-

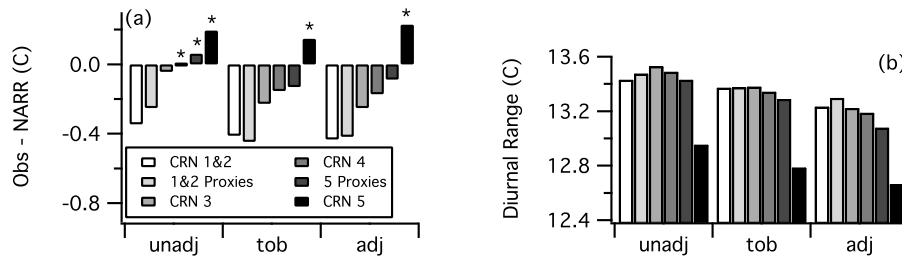
perature estimation using the limited number of CRN 1&2 and CRN 5 stations, according to analysis of the proxy stations (not shown).

[44] These long-term systematic variations among CRN classes of USHCNv2 stations lead to significant differences in the long term trends (Figure 10). The unadj and tob average temperature trends are about twice as large when estimated from CRN 5 stations as from CRN 1&2 stations; the CRN 5 tob trend difference is statistically significant and appears to be completely unrelated to differences in the distribution of stations. As with the 1979–2008 period, the adj trends are nearly identical, but the trend magnitude is much smaller for 1895–2009 than for 1979–2008. In contrast, the diurnal temperature range trend differences are statistically significant whether or not homogeneity corrections have been applied. The adj CRN 1&2 diurnal temperature range trend is almost exactly zero, while the adj CRN 5 diurnal temperature range trend is about  $-0.5^{\circ}\text{C}/\text{century}$ . The adjustments only reduce the trend difference between CRN 1&2 and CRN 5 by half, while the adjustments have a larger effect on the CRN 1&2 and CRN 3&4 trend differences.

[45] The adj trend differences in maximum and minimum temperature do not rise to statistical significance. The unadj and tob trends are significantly different for minimum temperature, but most of the difference is eliminated by the full adjustment. Maximum temperature trends are not sig-



**Figure 10.** Linear trends of United States temperature, as in Figure 4, but for the 1895–2009 period and for the station network groupings shown.



**Figure 11.** (a) United States (lower 48 states) mean temperature differences from NARR for the period 1979–2008, as estimated from USHCNv2 stations with the siting classifications shown. (b) United States mean diurnal temperature ranges for the period 1979–2008, as estimated from USHCNv2 stations with the siting classifications shown.

nificantly different except for tob trend differences between CRN 1&2 and CRN 3&4.

[46] Not only are the trends themselves smaller, but the 1895–2009 differences between CRN 1&2 and CRN 5 trends are also smaller than the corresponding 1979–2008 trend differences. This implies that, to the extent that the trend differences are caused by siting changes, a large portion of the siting changes has taken place since 1979. Nonetheless, Figure 9 shows that the diurnal temperature range trend was consistently lower for stations presently classified as more poorly sited from about 1920 onward.

[47] We computed the variance of the detrended aggregated monthly anomaly time series from stations of different classes and the correlation coefficients between the aggregated monthly anomalies and the NARR monthly anomalies aggregated from station sites to determine whether the siting differences influenced estimates of climate variability across the United States on an annual time scale. No statistically significant variance or correlation differences were found.

#### 4. Temperature Bias Analysis

[48] The evidence presented in the preceding section supports the hypothesis that station characteristics associated with station siting quality affect temperature trend estimates. If this is so, the temperature values themselves must necessarily be affected as well. This enables an independent test of the hypothesis: if the observed temperatures vary systematically according to station siting, and if that variation is consistent with the trend differences identified earlier, the hypothesis is confirmed. An analysis of the temperature bias may also provide useful information regarding the consistency of observed temperatures across classes and the size of the temperature errors associated with poor siting. On the other hand, because mean temperatures are much more sensitive to local climatic influences than are long-term temperature trends, an absolute temperature signal may be difficult to detect. Comparison to a reference temperature data set (in this case, NARR) is essential, and any reference data set comes with its own errors.

[49] Figure 11a shows the observed mean temperature differences from NARR mean temperatures, computed as described in section 2.3. Poorer siting is associated with temperatures that are warmer than CRN 1&2 stations compared to NARR. The CRN 5 biases are statistically significant. However, the differences between NARR and the CRN 1&2 proxy temperatures are similar to the differ-

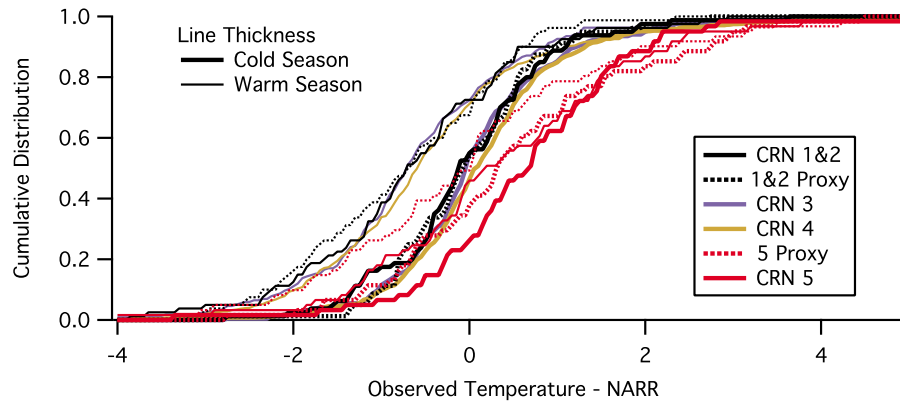
ences between NARR and the CRN 1&2 stations themselves. This means that the differences between CRN 1&2, CRN 3, and CRN 4 average temperatures compared to NARR are mostly or entirely attributable to the geographic distribution of CRN 1&2 stations rather than the siting characteristics. On the other hand, the CRN 5 stations are warmer  $0.3^{\circ}\text{C}$  than even the CRN 5 proxies when compared to NARR.

[50] Diurnal temperature range is not directly available from NARR so we calculate it directly. Figure 11b shows that diurnal temperature ranges are essentially undistinguishable among CRN 1&2, CRN 3, and CRN 4 stations, but are considerably smaller (by  $0.5\text{--}0.6^{\circ}\text{C}$ ) for the CRN 5 stations. The difference between CRN 1&2 and CRN 5 fully adjusted diurnal temperature ranges fails the significance test ( $p = 0.08$ ).

[51] The smaller diurnal temperature range for CRN 5 stations compared to other CRN classes is consistent with the trend differences identified earlier. The lack of a substantial average temperature difference across classes, once the geographical distribution of stations is taken into account, is also consistent with the lack of significant trend differences in average temperatures.

[52] The mean biases mask substantial variability from station to station. Figure 12 shows the cumulative distribution of average station temperatures relative to the NARR means. No attempt is made in Figure 12 to adjust for differences in the geographical distribution of stations, but the signs of the differences in Figure 12 among CRN classes are consistent with the mean biases analyzed spatially and depicted in Figure 11a. Figure 12 also separates the temperature differences into cool-season and warm-season differences in order to illustrate seasonal dependencies.

[53] For all quantities and seasons, the distribution of CRN 5 differences from NARR is broader and flatter than other CRN class distributions. However, the CRN 5 proxies have a similar cumulative distribution shape, indicating that the greater variability of CRN 5 measured surface temperatures as a group compared to NARR analyses is due to the station distribution. The offset between the CRN 5 cumulative distribution curve and the CRN 5 proxy cumulative distribution curve is a few tenths of a degree Celsius and is somewhat larger in the warm season than the cool season. Two-tailed paired sample  $t$  tests show that the differences between the CRN 5 and CRN 5 proxy temperature departures from NARR reanalysis fail to rise to statistical



**Figure 12.** Cumulative distribution of 1979–2008 seasonal mean differences between temperatures measured at USHCNv2 stations in various siting classifications and NARR mean temperatures. Warm season (summer) includes May through October and is illustrated with thin lines, and cool season (winter) includes November through April and is illustrated with thick lines.

significance in either the warm season ( $p = 0.13$ ) or the cool season ( $p = 0.43$ ).

[54] Overall, the calculations of bias relative to NARR reanalysis are broadly consistent with the difference in trends found in section 3. However, the greater difficulty of identifying a statistically significant signal in temperatures than in temperature trends means that further work is necessary to quantify the specific temperature effects associated with USHCN siting deficiencies.

[55] Also notable in Figure 12 is a seasonal dependence on the difference between observed average temperatures and NARR mean temperatures. Across all CRN classes, observed average surface temperatures are several tenths of a degree warmer relative to NARR means in the cool season than in the warm season. Because this difference is present across all CRN classes, its origin seems more likely to lie in the NARR reanalysis or differences in average temperature computation than in the USHCN observations.

## 5. Summary and Discussion

[56] The classification of 82.5% of USHCNv2 stations based on CRN criteria provides a unique opportunity for investigating the impacts of different types of station exposure on temperature trends, allowing us to extend the work initiated by *Watts* [2009] and *Menne et al.* [2010].

[57] The comparison of time series of annual temperature records from good and poor exposure sites shows that differences do exist between temperatures and trends calculated from USHCNv2 stations with different exposure characteristics. Unlike *Menne et al.* [2010], who grouped all USHCNv2 stations into two classes and found that “the unadjusted CONUS minimum temperature trend from good and poor exposure sites ... show only slight differences in the unadjusted data,” we found the raw (unadjusted) minimum temperature trend to be significantly larger when estimated from the sites with the poorest exposure sites relative to the sites with the best exposure. These trend differences were present over both the recent NARR overlap period (1979–2008) and the period of record (1895–2009). We find that the partial cancellation *Menne et al.* [2010] reported between the effects of time of observation bias

adjustment and other adjustments on minimum temperature trends is present in CRN 3 and CRN 4 stations but not CRN 5 stations. Conversely, and in agreement with *Menne et al.* [2010], maximum temperature trends were lower with poor exposure sites than with good exposure sites, and the differences in trends compared to CRN 1&2 stations were statistically significant for all groups of poorly sited stations except for the CRN 5 stations alone. The magnitudes of the significant trend differences exceeded  $0.1^{\circ}\text{C}/\text{decade}$  for the period 1979–2008 and, for minimum temperatures,  $0.7^{\circ}\text{C}$  per century for the period 1895–2009.

[58] Additional assessment of the relationship between station siting and surface temperature is done using linear trends of time series that have been corrected for time of observation changes and other apparent inhomogeneities. The full adjustments tended to reduce but not eliminate the trend differences, which remained significant over the 1979–2008 period compared to CRN 1&2 stations for CRN 5 minimum temperatures and CRN 4 maximum temperatures and became significant for CRN 5 maximum temperatures.

[59] The opposite-signed differences in maximum and minimum temperature trends at poorly sited stations compared to well-sited stations were of similar magnitude, so that average temperature trends were statistically indistinguishable across classes. For 30 year trends based on time-of-observation corrections, differences across classes were less than  $0.05^{\circ}\text{C}/\text{decade}$ , and the difference between the trend estimated using the full network and the trend estimated using the best-sited stations was less than  $0.01^{\circ}\text{C}/\text{decade}$ .

[60] While the opposite-signed differences cancel for average temperature, they magnify differences in trends of the diurnal temperature range. Such trends were significantly different from CRN 1&2 trends for all siting classes and both short term and century scale, with the short-term (1979–2008) trend for diurnal temperature range being negative for the most poorly sited stations and positive for the best-sited stations. The best-sited stations show essentially no long-term trend in diurnal temperature range, while the most poorly sited stations have a diurnal temperature range trend of  $-0.4^{\circ}\text{C}/\text{century}$ .

[61] The absence of a long-term trend in diurnal temperature range across the lower 48 states, as measured by well-

sited surface stations, has not previously been noted. Past studies of large-scale diurnal temperature range trends, such as by Karl *et al.* [1984], Easterling *et al.* [1997], and Vose *et al.* [2005b], identified a downward trend from the 1940s or 1950s to at least the 1980s, with little or no trend since. Karl *et al.* [1993] note that there was no downward trend in the United States prior to the mid-1950s. The present analysis confirms the multidecade downward trend beginning in the mid-1950s, but finds that upward trends during other periods resulted in zero diurnal temperature range trend for the period of record, 1895–2009.

[62] Assessments comparing observed and analyzed (NARR) monthly mean temperature anomalies illustrate how changes in temperature differences through time contribute to the trend differences. Using CRN 1&2 sites as a baseline, time-of-observation corrected minimum temperature measurements at CRN 5 stations have grown increasingly warm, while corresponding maximum temperatures cooled steadily until the mid-1990s and warmed thereafter. The full inhomogeneity adjustments reduce the rate of change of temperature differences but do not eliminate them except for removal of the post-1990s maximum temperature warming. The remaining trend differences imply that the adjustments did not fully correct for changes in instrumentation and microclimate with the transition to MMTS temperature sensors.

[63] An initial attempt at estimating the magnitudes of the temperature biases themselves is made by analyzing the differences between NARR temperatures and observed temperatures. The CRN 5 stations are on average warmer than the CRN 1&2 stations compared to interpolated NARR temperatures by about 0.7°C. However, when the differing geographical distribution of stations is taken into account, the difference attributable to siting characteristics alone is about 0.3°C. The diurnal temperature range is smaller for CRN 5 stations than for all other station classes by about 0.5°C, but this difference is not significant at the 5% level ( $p = 0.08$ ).

[64] In cases where no statistical significance was found, the absence of statistical significance does not necessarily imply a lack of influence of station siting characteristics, only that other variables, such as instrumentation differences or local climatic differences, are important and may be responsible for the calculated differences among classes. Conversely, statistically significant differences may in some cases be due to factors that covary with siting characteristics rather than the specific siting characteristics themselves. A follow-up study is underway to distinguish and quantify the separate effects of siting, instrumentation, urbanization, and other factors.

[65] Overall, this study demonstrates that station exposure does impact USHCNv2 temperatures. The temperatures themselves are warmest compared to independent analyses at the stations with the worst siting characteristics. Temperature trend estimates vary according to site classification, with poor siting leading to an overestimate of minimum temperature trends and an underestimate of maximum temperature trends, resulting in particular in a substantial difference in estimates of the diurnal temperature range trends. Homogeneity adjustments are necessary and tend to reduce the trend differences, but statistically significant differences remain for all but average temperature trends.

[66] Trend differences tend to become progressively larger (and more likely to be statistically significant) as siting quality degrades, except for average temperature trends which are relatively insensitive to CRN classification. It seems that the accuracy of maximum and minimum trend estimates can be improved by using only better-sited stations, but the appropriate quality criterion probably varies from situation to situation. There is a necessary tradeoff between the number of stations (more stations improve the signal-to-noise ratio) and the siting quality criterion (more lenient standards increase the observation biases). For the long-term trends considered here, the optimal network may consist exclusively of the CRN 1&2 stations. However, even the fully adjusted data from the highest-quality stations may be affected by trend biases in lower-quality stations in the interval surrounding the change point [Pielke *et al.*, 2007a]. It may be beneficial to exclude the most poorly sited stations from the adjustment procedure at better-sited stations.

[67] We recommend that this type of comprehensive siting study be extended to the global historical climate network [GHCN] temperature data (<http://www.ncdc.noaa.gov/oa/climate/ghcn-monthly/index.php>), as part of the improvement in metadata and benchmarking of data adjustment algorithms proposed in the meeting organized by Stott and Thorne [2010].

[68] **Acknowledgments.** The authors wish to acknowledge the many cooperative observers who unselfishly carry out COOP observations, which are the backbone of climate monitoring. We also acknowledge the many volunteers who 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, Don and Liz Healy, Eric Gamberg, John Goetz, Don Kostuch, John Slayton, Ted Semon, Russell and Ellen Steele, and Barry Wise. Acknowledgment is given to former California State Climatologist James Goodridge, who was inspirational with surveys he made of California COOP stations during his tenure. Station photographs are courtesy of Evan Jones, Warren Meyer, Michael Denegri, Rex E. Kirksey (via SurfaceStations.org), and Christopher A. Davey [Davey and Pielke, 2005]. We would also like to thank Richard McNider for his assistance and thoughtful comments. We acknowledge Dallas Staley for her standard outstanding editorial support.

## References

- Baker, D. G. (1975), Effect of observation time on mean temperature estimation, *J. Appl. Meteorol.*, *14*, 471–476, doi:10.1175/1520-0450(1975)014<0471:EOOTOM>2.0.CO;2.
- Balling, R. C., Jr., and C. D. Idso (2002), Analysis of adjustments to the United States Historical Climatology Network (USHCN) temperature database, *Geophys. Res. Lett.*, *29*(10), 1387, doi:10.1029/2002GL014825.
- 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 data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.
- Cao, H. X., J. F. B. Mitchell, and J. R. Lavery (1992), Simulated diurnal range and variability of surface temperature in a global climate model for present and doubled CO<sub>2</sub> climates, *J. Clim.*, *5*, 920–943, doi:10.1175/1520-0442(1992)005<0920:SDRAVO>2.0.CO;2.
- Christy, J. R. (2002), When was the hottest summer? A state climatologist struggles for an answer, *Bull. Am. Meteorol. Soc.*, *83*, 723–734, doi:10.1175/1520-0477(2002)083<0723:WWTHS>2.3.CO;2.
- 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, *J. Clim.*, *19*, 548–563, doi:10.1175/JCLI3627.1.

- Christy, J. R., W. B. Norris, and R. T. McNider (2009), Surface temperature variations in East Africa and possible causes, *J. Clim.*, *22*, 3342–3356, doi:10.1175/2008JCLI2726.1.
- 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, *Bull. Am. Meteorol. Soc.*, *86*, 497–504, doi:10.1175/BAMS-86-4-497.
- DeGaetano, A. T. (2006), Attributes of several methods for detecting discontinuities in mean temperature series, *J. Clim.*, *19*, 838–853, doi:10.1175/JCLI3662.1.
- 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*, Carbon Dioxide Inf. and Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Easterling, D. R., et al. (1997), Maximum and minimum temperature trends for the globe, *Science*, *277*, 364–367, doi:10.1126/science.277.5324.364.
- Fall, S., D. Niyogi, R. A. Pielke Sr., A. Gluhovsky, E. Kalnay, and G. Rochon (2010), Impacts of land use land cover on temperature trends over the continental United States: Assessment using the North American Regional Reanalysis, *Int. J. Climatol.*, *30*, 1980–1993, doi:10.1002/joc.1996.
- Hansen, J. E., R. Ruedy, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl (2001), A closer look at United States and global surface temperature change, *J. Geophys. Res.*, *106*, 23,947–23,963, doi:10.1029/2001JD000354.
- Hubbard, K. G., and X. Lin (2006), Reexamination of instrument change effects in the U.S. Historical Climatology Network, *Geophys. Res. Lett.*, *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 Sens. Environ.*, *61*, 361–370, doi:10.1016/S0034-4257(97)00046-1.
- Kalnay, E., and M. Cai (2003), Impact of urbanization and land-use change on climate, *Nature*, *423*, 528–531, doi:10.1038/nature01675.
- 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, *J. Geophys. Res.*, *111*, D06106, doi:10.1029/2005JD006555.
- Kalnay, E., M. Cai, M. Nunez, and Y. Lim (2008), Impacts of urbanization and land surface changes on climate trends, *Int. Assoc. Urban Clim. Newsl.*, *27*, 5–9.
- Karl, T. R., and C. N. Williams Jr. (1987), An approach to adjusting climatological time series for discontinuous inhomogeneities, *J. Clim. Appl. Meteorol.*, *26*, 1744–1763, doi:10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.
- Karl, T. R., G. Kukla, and J. Gavin (1984), Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980, *J. Clim. Appl. Meteorol.*, *23*, 1489–1504, doi:10.1175/1520-0450(1984)023<1489:DDTRIT>2.0.CO;2.
- Karl, T. R., C. N. Williams Jr., P. J. Young, and W. M. Wendland (1986), A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperature for the United States, *J. Clim. Appl. Meteorol.*, *25*, 145–160, doi:10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2.
- Karl, T. R., H. F. Diaz, and G. Kukla (1988), Urbanization: Its detection and effect in the United States climate record, *J. Clim.*, *1*, 1099–1123, doi:10.1175/1520-0442(1988)001<1099:UIDAEI>2.0.CO;2.
- 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, *Environ. Sci. Div. Publ. 3404*, 389 pp., Carbon Dioxide Inf. and Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Karl, T. R., P. D. Jones, R. W. Knight, D. Kukla, N. Plummer, V. Razuvayev, K. P. Gallo, J. Lindsey, R. J. Charlson, and T. C. Peterson (1993), A new perspective in recent global warming: Asymmetric trends of daily maximum and minimum temperature, *Bull. Am. Meteorol. Soc.*, *74*, 1007–1023, doi:10.1175/1520-0477(1993)074<1007:ANPORG>2.0.CO;2.
- Leroy, M. (1999), Classification d’un site, *Note Tech. 35*, 12 pp., Dir. des Syst. d’Obs., Météo-France, Trappes, France.
- 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, *Geophys. Res. Lett.*, *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, *Int. J. Climatol.*, *26*, 1091–1124, doi:10.1002/joc.1298.
- Mahmood, R., et al. (2010), Impacts of land use land cover change on climate and future research priorities, *Bull. Am. Meteorol. Soc.*, *91*, 37–46, doi:10.1175/2009BAMS2769.1.
- Menne, M. J., and C. N. Williams Jr. (2005), Detection of undocumented change-points using multiple test statistics and composite reference series, *J. Clim.*, *18*, 4271–4286, doi:10.1175/JCLI3524.1.
- Menne, M. J., and C. N. Williams Jr. (2009), Homogenization of temperature series via pairwise comparisons, *J. Clim.*, *22*, 1700–1717, doi:10.1175/2008JCLI2263.1.
- Menne, M. J., C. N. Williams Jr., and R. S. Vose (2009), The United States Historical Climatology Network monthly temperature data, version 2, *Bull. Am. Meteorol. Soc.*, *90*, 993–1007, doi:10.1175/2008BAMS2613.1.
- 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.
- Mesinger, F., et al. (2006), North American Regional Reanalysis, *Bull. Am. Meteorol. Soc.*, *87*, 343–360, doi:10.1175/BAMS-87-3-343.
- Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*, 693–712, doi:10.1002/joc.1181.
- National Oceanic and Atmospheric Administration and National Environmental Satellite, Data, and Information Service (NOAA and NESDIS) (2002), Climate Reference Network site information handbook, *NOAA-CRN/OSD-2002-0002ROUD0*, 19 pp., U.S. Dep. of Commer., Natl. Clim. Data Cent., Asheville, N. C.
- National Weather Service (NWS) (1989), *National Weather Service Observing Handbook*, vol. 2, *Cooperative Station Observations*, 1st ed., Off. of Syst. Oper., Silver Spring, Md. [Available at <http://www.nws.noaa.gov/om/coop/Publications/coophandbook2.pdf>].
- Office of the Federal Coordinator for Meteorological Services and Supporting Research (1994), Temperature and dew point sensors, in *Federal Standard for Siting Meteorological Sensors at Airports, FCM-S4-1994*, pp. 2–4, U.S. Dep. of Commer., Washington, D. C. [Available at <http://www.ofcm.gov/siting/pdf/fcm-s4-1994%28Siting%29.pdf>].
- Peterson, T. C. (2003), Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found, *J. Clim.*, *16*, 2941–2959, doi:10.1175/1520-0442(2003)016<2941:AOUVRI>2.0.CO;2.
- Peterson, T. C. (2006), Examination of potential biases in air temperature caused by poor station locations, *Bull. Am. Meteorol. Soc.*, *87*, 1073–1089, doi:10.1175/BAMS-87-8-1073.
- Peterson, T. C., and D. R. Easterling (1994), Creation of homogeneous composite climatological reference series, *Int. J. Climatol.*, *14*, 671–679, doi:10.1002/joc.3370140606.
- Peterson, T. C., et al. (1998), Homogeneity adjustments of in situ atmospheric climate data: A review, *Int. J. Climatol.*, *18*, 1493–1517, doi:10.1002/(SICI)1097-0088(199811)18:13<1493::AID-JOC329>3.0.CO;2-T.
- 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, *Int. J. Climatol.*, *22*, 421–434, doi:10.1002/joc.706.
- Pielke, R. A., Sr., et al. (2007a), Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment, *Bull. Am. Meteorol. Soc.*, *88*, 913–928, doi:10.1175/BAMS-88-6-913.
- Pielke, R. A., Sr., et al. (2007b), Unresolved issues with the assessment of multidecadal global land surface temperature trends, *J. Geophys. Res.*, *112*, D24S08, doi:10.1029/2006JD008229.
- Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes (1991), Effects of recent thermometer changes in the Cooperative Station Network, *Bull. Am. Meteorol. Soc.*, *72*, 1718–1723, doi:10.1175/1520-0477(1991)072<1718:EORTCI>2.0.CO;2.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu (2007), A review and comparison of change-point detection techniques for climate data, *J. Appl. Meteorol. Climatol.*, *46*, 900–915, doi:10.1175/JAM2493.1.
- Runnalls, K. E., and T. R. Oke (2006), A technique to detect microclimatic inhomogeneities in historical records of screen-level air temperature, *J. Clim.*, *19*, 959–978, doi:10.1175/JCLI3663.1.
- 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 CCCma coupled model, *Clim. Dyn.*, *20*, 435–445.
- Stott, P., and P. Thorne (2010), Proposals for surface-temperature databank now open for scrutiny, *Nature*, *466*, 1040, doi:10.1038/4661040d.
- Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears (2005), Uncertainties in climate trends: Lessons from upper-air temperature records, *Bull. Am. Meteorol. Soc.*, *86*, 1437–1442, doi:10.1175/BAMS-86-10-1437.



- Vose, R. S., and M. J. Menne (2004), A method to determine station density requirements for climate observing networks, *J. Clim.*, *17*, 2961–2971, doi:10.1175/1520-0442(2004)017<2961:AMTDSD>2.0.CO;2.
- 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, *Geophys. Res. Lett.*, *30*(20), 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,” *Bull. Am. Meteorol. Soc.*, *86*, 504–506, doi:10.1175/BAMS-86-4-504.
- 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, *Geophys. Res. Lett.*, *34*, L12709, doi:10.1029/2007GL029505.
- Watts, A. (2009), *Is the U.S. Surface Temperature Record Reliable?*, 28 pp., Heartland Inst., Chicago, Ill.
- Wendland, W. M., and W. Armstrong (1993), Comparison of maximum–minimum resistance and liquid-in-glass thermometer records, *J. Atmos. Oceanic Technol.*, *10*, 233–237, doi:10.1175/1520-0426(1993)010<0233:COMRAL>2.0.CO;2.
- Willmott, C. J., S. M. Robeson, and J. J. Feddema (1991), Influence of spatially variable instrument networks on climatic averages, *Geophys. Res. Lett.*, *18*, 2249–2251, doi:10.1029/91GL02844.
- Yilmaz, H., S. Toy, M. A. Irmak, S. Yilmaz, and Y. Bulit (2008), Determination of temperature differences between asphalt concrete, soil, and grass surfaces of the city of Ezurum, Turkey, *Atmosfera*, *21*, 135–146.
- 
- J. R. Christy, Department of Atmospheric Science, University of Alabama in Huntsville, Huntsville, AL 35899, USA.
- S. Fall, College of Agricultural, Environmental and Natural Sciences, Tuskegee University, 1200 W. Montgomery Rd., Tuskegee, AL 36088, USA.
- E. Jones and A. Watts, IntelliWeather, 3008 Cohasset Rd., Chico, CA 95973, USA.
- J. Nielsen-Gammon, Department of Atmospheric Sciences, Texas A&M University, 3150 TAMUS, College Station, TX 77843, USA.
- D. Niyogi, Indiana State Climate Office, Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA.
- R. A. Pielke Sr., CIRES/ATOC, University of Colorado at Boulder, Boulder, CO 80309, USA. (pielkesr@cires.colorado.edu)