2.6 THE UNITED STATES HISTORICAL CLIMATOLOGY NETWORK MONTHLY TEMPERATURE DATA – VERSION 2

Matthew J. Menne*, Claude N. Williams, Jr. and Russell S. Vose
NOAA National Climatic Data Center, Asheville, North Carolina

ABSTRACT

In support of its climate monitoring and assessment activities, NOAA's National Climatic Data Center has developed an improved version of the U.S. Historical Climatology Network temperature dataset (i.e., HCN version 2). In this paper, the HCN version 2 temperature data are described in detail, with a focus on the quality-assured dataset sources and the systematic bias adjustments. The bias adjustments are discussed in the context of their impact on U.S. temperature trends from 1895-2006 and in terms of the differences between HCN version 2 and its widely used predecessor (now referred to as HCN version 1).

Because the collective impact of changes in observation practice at U.S. HCN stations is of the same order of magnitude as the background climate signal, bias adjustments are essential to reducing the uncertainty in U.S. climate trends. The largest biases in the HCN are shown to be associated with changes to the time of observation and with the widespread changeover from liquid in glass thermometers to the maximum minimum temperature sensor (MMTS). With respect to the version 1 annual and seasonal temperature trends, HCN version 2 trends are generally somewhat smaller because of an apparent over correction in version 1 for the MMTS instrument change and because of the inclusion of adjustments for undocumented shifts in version 2.

1. INTRODUCTION

Since 1987, NOAA's National Climatic Data Center (NCDC) has used the U.S. Historical Climatology Network (HCN) to quantify national- and regional-scale temperature change in the conterminous United States. To that end, HCN temperature records have been "corrected" to account for various historical changes in station location, instrumentation, and observing practice. The HCN is a designated subset of the NOAA Cooperative Observer (Coop) Network, with HCN sites having been selected according to their spatial coverage, record length, data completeness, and historical stability. The HCN therefore consists primarily of long-term Coop stations whose temperature records have been adjusted for systematic, non-climatic changes that bias temperature trends.

* *Corresponding Author Address*: Matthew J. Menne, National Climatic Data Center, 151 Patton Ave, Asheville, NC 28801; e-mail: Matthew.Menne@noaa.gov.

In support of its operational monitoring and climate assessment activities, NCDC has recently developed an improved HCN dataset (hereafter termed HCN version 2). In this paper we describe the HCN version 2 temperature data in detail, focusing on the quality-assured dataset sources as well as the bias adjustment techniques employed in version 2 to further reduce uncertainty in the U.S. instrumental temperature record. The HCN bias adjustments are discussed in the context of their impact on U.S. temperature trends and in terms of the differences between HCN version 2 and its widely used predecessor (now termed HCN version 1).

2. DATA

a. Network development

HCN is a reference station network (Collins et al. 1999) – i.e., a subset of long-term climate stations managed as part of a larger network, in this case the Coop network shown in Fig. 1. The original HCN stations were identified in the mid-1980s by examining temperature data (and metadata) from the Coop network with the goal of maximizing record length, data completeness, and stability in station location (Quinlan et al. 1987). To be designated as part of the HCN, a Coop station was ideally required to be active in 1987 and to have a period of record of at least 80 years. In practice, these criteria were sometimes relaxed in order to provide a more uniform distribution of stations across the country and to incorporate the recommendations of the nation's State Climatologists. The resulting network contained 1219 Coop stations, 84 of which were composites formed using consecutive records from two or more stations in order to achieve the minimum period of record requirement.

The actual subset of stations constituting the HCN has changed twice since 1987. By the mid-1990s, station closures and relocations had already forced a reevaluation of the composition of HCN as well as the creation of additional composite stations. The reevaluation led to 52 station deletions and 54 additions, for a total of 1221 stations (156 of which were composites). Since the 1996 release (Easterling et al. 1996), numerous station closures and relocations have again necessitated a revision of the network. As a result, HCN version 2.0 contains 1218 stations, 208 of which are composites; relative to the 1996 release, there have been 62 station deletions and 59 additions.
Fig. 1 depicts the locations of the 1218 stations in HCN version 2 (red triangles). Consistent with previous releases, the spatial distribution is relatively uniform across most of the conterminous United States.

However, some sampling variability is evident; for instance, parts of the intermountain West (e.g., southern Nevada) are more poorly sampled while portions of the East (e.g., northern New Jersey) are densely covered. In addition, as depicted by Fig. 2, the composition of the network is not uniform in time. For example, there is a rapid increase in the number of stations until about 1925, with the spatial coverage increasing most prominently in the West during these early years. The network remains relatively consistent in size until the late 20th century, after which it declines somewhat because of the closure of a number of stations. There is no spatial bias to the decline in recent years.

b. Source data

To maximize data completeness, HCN version 2 was derived from five complementary source datasets archived at NCDC:

- DSI-3200: U.S. Cooperative Summary of the Day;
- DSI-3206: U.S. Cooperative Summary of the Day – Pre-1948;
- DSI-3210: U.S. Summary of the Day First Order Data;
- DSI-3220: U.S. Summary of the Month; and
- U.S. HCN version 1 monthly data.

The first three datasets contain daily records while the last two consist of monthly means. Each dataset also contained "estimated" values and quality assurance (QA) flags; however, in order to standardize QA across data sources, neither the estimated values nor the quality flags were employed in building HCN version 2. Instead, each daily data source was first subjected to the suite of 12 QA reviews listed in Table 1. The QA checks were performed in the order in which they appear in the table, with each procedure operating on only those values that did not fail any of the preceding tests. The performance of each check was evaluated using the method of Durre et al. (in press). Collectively, the daily QA system had an estimated false-positive rate of 8% (i.e., the percent of flagged values that appear to be valid) and a miss rate of less than 5% (the percent of true errors that remain undetected). Monthly means were then derived from the quality assured daily data, with a requirement that no more than nine values be flagged or missing in any given month.

The five source datasets were subsequently merged by Coop number to form a comprehensive dataset of serial monthly temperature values. Duplicate records between data sources were eliminated based on a simple dataset priority scheme (i.e., DSI-3200 had the highest ranking, followed by DSI-3206, and so on). The resulting merged dataset was then subjected to the three additional QA reviews listed in Table 2; together, these checks had a false positive rate of 15% for maximum temperature and 10% for minimum temperature. Note that the two spatial checks were performed after the climatological check; furthermore, each was applied iteratively until no additional spatial inconsistencies were detected. The monthly QA reviews removed less than 0.2% of monthly temperature values.

3. SOURCES AND ASSESSMENT OF TEMPERATURE BIAS IN THE U.S. HCN

The process of removing systematic changes in the bias in a climate series is called homogenization, and the systematic artificial shifts in a series are frequently referred to as inhomogeneities. In the HCN, there are a number of causes behind inhomogeneities, including changes to the time of observation, station moves, instrument changes, and changes to conditions surrounding the instrument site. An assessment of each of these causes is discussed below.

a. Bias caused by changes to the time of observation

The majority of the Coop network observers (and also HCN) are volunteers who make observations at times that are more convenient than local midnight. However, the time at which daily maximum and minimum temperatures are observed has a systematic effect on the calculation of the monthly mean (Baker 1976; Karl et al. 1986). To illustrate, consider that a Coop observer who records daily maximum and minimum temperatures in the afternoon (e.g., 1700 LST) resets the temperature sensors at a time when the diurnal cycle is near its usual maximum. Suppose then that the observer resets the sensors at 1700 LST on April 1. If the maximum temperature on April 2 never exceeds the values at 1700 LST on April 1 (e.g., because of subsequent cold air advection), then the maximum temperature that gets recorded for April 2 will be the value at 1700 LST on April 1. The consequence of this type of "carry over" maximum is a positive (warm) bias in mean monthly maximums based on an afternoon reading schedule relative to those calculated from a midnight LST summary period (for which the observational day and calendar day coincide). Conversely, a morning Coop observer records daily minimum temperatures near the usual diurnal minimum, which leads to the issue of carry over minimums and a negative (cool) bias in mean monthly minimum temperatures relative to a midnight or afternoon daily summary period.

The systematic time of observation bias would be of little concern with regard to temperature trends provided that the observation time at a given station did not change during its operational history. As shown in Fig. 3, however, there has been a widespread conversion from afternoon to morning observation times in the HCN. Prior to the 1940s, for example, most observers recorded near sunset in accordance with U.S. Weather Bureau instructions. Consequently, the U.S. climate record as a whole contains a slight positive (warm) bias during the first half of the century. A switch to morning observation times has steadily occurred during the latter half of the century to support operational hydrological requirements. The result is a broad-scale reduction in mean temperatures that is simply caused by the conversion in the daily reading schedule of the Cooperative Observers. In other words, the gradual conversion to morning observation times in the United States during the past 50 years has artificially reduced the true temperature trend in the U.S. climate record (Karl et al. 1986; Vose et al. 2003; Hubbard and Lin 2006; Pielke et al. 2007a).

To account for this Time of Observation Bias (TOB) in the version 2 monthly temperatures, the method described in Karl et al. (1986) was used. The robustness of this method, which was also used to produce version 1 temperatures, has been verified by Vose et al. (2003). In particular, because the TOB adjustment requires documentation of changes to the observation schedule, Vose et al. (2003) verified the accuracy of the U.S. HCN time of observation history using an independently generated source of metadata (i.e., DeGaetano 1999). In addition, the predictive skill of the Karl et al. (1986) approach to estimating the TOB was confirmed using hourly data from 500 stations over the period 1965-2001 (whereas the approach was originally developed using data from 79 stations over the period 1957-64). The Karl et al. (1986) TOB adjustment procedure was therefore used in version 2 without modification.

To calculate the impact of the TOB adjustments on the HCN version 2 temperature trends, the monthly TOB-adjusted temperatures at each HCN station were converted to an anomaly relative to the 1961-1990 station mean. Anomalies were then interpolated to the nodes of a $0.25° \times 0.25°$ latitude/longitude grid using the method described by Willmott et al. (1985). Finally, grid point values were area weighted into a mean anomaly for the conterminous United States for each month and year. The process was then repeated for the unadjusted temperature data, and a difference series was formed between the TOB-adjusted and unadjusted data as shown in Fig. 4.

Figure 4 indicates that removing the time of observation bias progressively elevates the mean U.S. temperature relative to the raw value during the period that coincides with the gradual shift to morning observation times in the network. The net effect of the TOB adjustments is to increase the overall trend in maximum temperatures by about 0.012°C dec^{-1} and in minimum temperatures by about 0.018°C dec^{-1} over the period 1985-2006. This net impact is about the same as that of the TOB adjustments in the version 1 temperature data (Hansen et al. 2001), which is to be expected since the same TOB-adjustment method is used in both versions.

b. Bias associated with other changes in observation practice

In addition to changes in the time of observation, most surface weather stations also experience changes in station location or instrumentation at various times throughout their histories. Such modifications generally entail alterations in sensor exposure and/or measurement bias that cause shifts in the temperature series that are unrelated to true climate variations. In version 1, the impacts of station moves and instrument changes were addressed using the procedure described by Karl and Williams (1987). Because this procedure addressed changes that are documented in the NOAA/NCDC station history archive, the version 1 homogeneity algorithm was termed the Station History Adjustment Program (SHAP).

Unfortunately, Coop station histories are incomplete. As a result, discontinuities may occur with no associated record in the metadata. Since undocumented discontinuities remain undetected by methods like SHAP, a new homogenization algorithm was developed for the HCN version 2 temperature data (Menne and Williams, submitted). This new algorithm addresses both documented and undocumented discontinuities via a pairwise comparison of temperature records. In this pairwise approach, comparisons are made between numerous combinations of temperature series in a region to identify and remove relative inhomogeneities (i.e., abrupt changes in one station series relative to many others).

As described in Menne and Williams (submitted), the pairwise approach works best when there are many neighboring series available for comparison with each target series. In order to maximize the number of potential neighbors for each HCN station, all Coop temperature series were used as input by the pairwise algorithm in version 2. In contrast, the SHAP in version 1 was restricted to inter-comparing HCN series only, in large part because of the more limited availability of digital monthly data (and metadata) during the late 1980s. Since that time, digitization efforts under the Climate Data Modernization Program (CDMP 2001) have markedly increased the volume of digital station data and histories available for the early years of the Cooperative Observer program, as shown in Figure 5. As noted in section 2, these historical temperature values were merged with other Coop data sources, which effectively increases the density of the network (and the correlation between all series tested), thereby improving the ability of the pairwise algorithm to detect relative inhomogeneities.

As in version 1, homogeneity testing in version 2 was conducted separately for mean monthly maximum and minimum temperature series. Fig. 6 depicts the frequency and magnitude of shifts detected by the pairwise algorithm for each variable. (A tabular summary of the shifts is also provided by variable in Table 3.) Overall, the pairwise algorithm identified 6487 (7854) statistically significant changepoints in maximum (minimum) temperature series. Since there are approximately 120,000 station-years of temperatures in the HCN version 2 dataset, this represents an average of about one significant artificial shift for every 15 years of station data. In terms of the adequacy of the HCN metadata, only about 44% (40%) of the significant inhomogeneities in maximum (minimum) temperatures are associated with documented changes.

Most of the documented changes in the HCN are associated with station relocations. In theory, minor station moves or other changes to the circumstances of sensor exposure would be expected to have a more pronounced effect on minimum temperatures than on maximum temperatures. The reason is that minimum temperatures generally occur near sunrise when calm and stable atmospheric boundary layer conditions are prevalent, at which time near surface temperature fields are strongly coupled to the local surface characteristics (Oke, 1987). On the other hand, during daylight hours, the boundary layer is more commonly well mixed, and microclimate differences between nearby locations should be less evident. The larger number of shifts detected in minimum temperature series relative to maximum temperature series is consistent with this reasoning.

While station changes can cause either an artificial rise or drop in temperature, the distribution of shifts identified in HCN version 2 is not necessarily symmetric about zero. For example, there are about 400 more negative shifts than positive shifts in maximum temperature series (Fig. 6a; Table 3). Most of this asymmetry appears to be associated with documented changes in the network (Fig. 6e) and, in particular, with shifts caused by the transition from liquid in glass (LiG) thermometers to the Maximum/Minimum Temperature System (MMTS; Fig. 6g). Quayle et al. (1991) concluded that this transition led to an average drop in maximum temperatures of about 0.4°C and to an average rise in minimum temperatures of 0.3°C for sites with no coincident station relocation. [These averages were subsequently used in version 1 to adjust the records from HCN stations that converted to the MMTS, primarily during the mid- and late 1980s (Easterling et al., 1996).] More recently, Hubbard and Lin (2006) estimated a somewhat larger MMTS impact on HCN temperatures and advocated for site specific adjustments in general, including those sites with no equipment move.

Notably, the pairwise algorithm in version 2 allows for such site specific adjustments to be calculated for all types of station changes. The subsets of changes associated with the conversion to the MMTS are shown in Figs. 6g and 6h. The pairwise results indicate that only about 33% (40%) of the maximum (minimum) temperature series experienced a statistically significant shift (out of ~850 total conversions to MMTS). As a result, the overall impact of the MMTS instrument change at all affected sites is substantially less than both the Quayle et al. (1991) and Hubbard and Lin (2006) estimates. However, the average impact of the statistically significant changes (-0.56°C for maximum temperatures and +0.39°C for minimum temperatures) is nearly identical to Hubbard and Lin (2006) results for sites with no coincident station move.

For the HCN version 2 as a whole, the combined impact of all adjustments for documented and undocumented changes is to increase the average U.S. trend in maximum temperatures by about 0.027°C dec^{-1} over the period of record relative to the values adjusted for TOB only (Fig. 7). In contrast, the impact of the pairwise homogenization algorithm on minimum temperature trends is effectively zero. As Fig. 7 indicates, the most significant impact of the adjustments on maximum temperatures begins after 1985, which coincides with the beginning of the changeover to the MMTS. The trend in the difference between the fully adjusted maximum temperature data and the TOB adjusted data reflects the cumulative impact of the individual instrument changes.

While the majority of MMTS changes occurred during the mid- and late 1980s, about 10% of HCN stations made the switch after 1994 (the last update to the version 1 digital metadata). In addition, a number of sites (about 5% of the network) converted to the Automated Surface Observation System (ASOS) after 1992. Like the MMTS, ASOS maximum temperature measurements have been shown to be lower relative to values from previous instruments (e.g., Guttman and Baker 1996). Such results are in agreement with the pairwise adjustments produced in version 2; that is, an average shift in maximum temperatures caused by the transition to ASOS in the HCN of about -0.44°C. The combined impact of the transition to MMTS and ASOS appears to be largely responsible for the continuing trend in differences between the fully and TOB-only adjusted maximum temperatures since 1985. On the other hand, while the impact of ASOS on minimum temperatures in the HCN is nearly identical to that on maximum temperatures (-0.45°C), the shifts associated with ASOS are opposite in sign to those caused by the transition to MMTS, which leads to a network-wide cancellation effect between the two instrument changes.

c. Bias associated with urbanization and nonstandard siting

In version 1, the regression-based approach of Karl et al. (1988) was employed to account for the impact of the urban heat island (UHI) bias on temperatures in the HCN (which they found to be important for minimum temperatures only). For a number of reasons, no specific urban correction is applied in HCN version 2. First, the adjustments for undocumented changepoints in version 2 actually account for the changes addressed by the Karl et al. (1988) UHI correction in version 1. In fact, as discussed in section 4, including adjustments for undocumented changepoints actually has a greater impact on minimum temperatures than the version 1 UHI correction. Second, adjusting for both documented and undocumented changepoints effectively removes most of the local, unrepresentative trends at individual HCN stations that may arise from gradual changes to the environment. The minimum temperature time series for Reno, Nevada (Fig. 8) illustrates this effect. Specifically, the TOB-adjusted data clearly indicate that the station developed a local trend beginning in the 1970s, possibly as a result of a growing urban heat island influence. In contrast, the fully adjusted version 2 data indicate that the relative trend changes have been largely removed. (Notably, the Reno series is also characterized by major step changes during the 1930s and 1990s caused by station relocations. Both abrupt changes were also removed by the version 2 adjustments.)

It is important to note, however, that while the pairwise algorithm uses a trend identification process to discriminate between gradual and sudden changes, trend inhomogenieties in the HCN are not actually removed with a trend adjustment. Rather, the pairwise approach uses a simple difference in means in the target minus neighbor series (before and after a step change) to estimate the magnitude of the shift even when there was a relative trend between the two series (as in the case of Reno). Ideally, trend inhomogeneities would be removed with gradual adjustments and step changes with abrupt adjustments. Unfortunately, unlike relative step changes, which occur simultaneously in all difference series formed between an HCN temperature series and those of its neighbors, a trend inhomogeneity may begin and end at different times with respect to its various neighbors. This makes it nearly impossible to robustly identify the true interval of a trend inhomogeneity (Menne and Williams, submitted).

In contrast, using a simple difference in means test addresses both gradual and sudden changes, producing what arguably approximates the "best objective hypothetical climate record available for the corrected station" (Pielke et al. 2007b). More generally, accounting for both sudden and gradual changes is critical because spurious results may occur if only the sudden changes are corrected (Menne and Williams submitted). The reason is that, in some cases, gradual and sudden changes may not reflect station moves and the impact of urbanization, but rather some kind of microclimate peculiarity such as the growth and removal of a single tree. In such an instance, correcting for the sudden change, but not for the gradual change, would likely produce unrealistic adjusted temperature values. Even in a case like Reno, preserving the local trend (i.e., not adjusting for the gradual change) would probably result in a bogus "double counting" of the UHI signal because the station would have experienced urbanization when it was located in the city and then again after its relocation in the mid-1990s to the airport (whose surroundings became urbanized somewhat later).

One implication of using a difference in means test to adjust for all change points is that local trends are "aliased" onto the estimates of step changes (DeGaetano 2006). To quantify the impact of this aliasing effect, the pairwise approach was temporarily modified such that only abrupt shifts were removed, creating a "non-production" version of HCN in which local trends were retained. In the case of minimum temperature, the resulting distribution of documented shifts became somewhat less skewed in favor of negative changes while the distribution of undocumented shifts became more skewed in favor of positive changes (relative to the results presented in Table 3). The reason for these distributional changes is that there is a sizable preference for relative trends between HCN stations and their neighbors to be negative. In fact, considering all the pairwise difference series segments formed to calculate the magnitude of documented and undocumented shifts, there are 25% more negative trends than positive trends. In other words, there is a general tendency for HCN minimum temperature trends to be smaller relative to surrounding Coop stations. This means that the local trend aliasing effect, on the whole, is removing more negative than positive trend inhomogeneities at HCN stations, in spite of cases like Reno. Thus, while there may be residual trend inhomogeneities left in some HCN series, they are more likely to be negative than positive and, collectively, there is no evidence of a positive bias in HCN trends caused by the UHI or other local changes.

Finally, a number of recent articles have raised concerns about the siting characteristics of HCN stations by way of photographic documentation (e.g., Davey and Pielke 2005; Pielke et al. 2007a,b). In at least one case (i.e., Mahmood et al. 2006) photographic documentation and other sources of information regarding the exposure characteristics of Coop and HCN sites were used to link poor siting with measurement bias. Such evidence raises legitimate questions about the representativeness of temperature measurements from a number of HCN sites. Nevertheless, from a climate change perspective, the primary concern is not so much the absolute measurement bias of a particular site, but rather the changes in that bias over time. The TOB and pairwise adjustments effectively address those changes.

The goal of the version 2 adjustments, therefore, was not to ensure that observations conform to an absolute standard, but rather to remove the impact of relative bias changes that occur during a station's history of observation. In this regard, photographic documentation is most valuable when it is used to document the timing and causes of changes in bias through time. Moreover, relative changes in the bias of observations, whatever the source, cannot be inferred from the metadata. Instead, the impact of station changes and non standard instrument exposure on temperature trends must be determined via a systematic evaluation of the observations themselves (Peterson 2006).

d. Bias assessment of estimates for missing monthly temperature values

As in version 1, HCN version 2 provides estimates for missing monthly maximum and minimum temperatures. Estimates are generated using an optimal interpolation technique known as FILNET (for *FIL*L in the

*NET*WORK) that makes use of the fully adjusted temperature values at neighboring Coop stations. In essence, the FILNET procedure iterates to find an optimal set of neighboring series that minimizes the confidence limits for the difference between the target series and the average of neighboring series (optimized separately for each calendar month). The difference between the target and neighbor average is used as an offset in the interpolation to account for climatological differences between the target and neighbors. The FILNET technique is also used to estimate data in a series where changepoints occur too close together in time (i.e., less than 24 months apart) to reliably estimate the magnitude of shift identified by the pairwise algorithm.

To assess the performance of FILNET, estimates were generated for all mean monthly maximum and minimum temperatures in the HCN and compared with the observed values. Specifically, both the mean difference and the mean absolute difference between the estimated and observed values were calculated separately for each decade in the HCN period of record. As shown in Fig. 9, the mean difference between the FILNET estimates and the observed values is less than 0.1°C in all decades. In addition, the mean absolute difference between the FILNET estimates and the observed values decreases with time as the density of stations in the Coop network increases. For the period of record as a whole, the mean difference between the FILNET estimates and the observed monthly values in the HCN is 0.01°C while the mean absolute difference is just under 0.5°C. Not surprisingly, the FILNET procedure has virtually no systematic impact on HCN temperature trends as shown in Fig. 10.

4. COMPARISON OF HCN VERSION 1 AND 2 TEMPERATURES

To assess the basic temperature differences between HCN version 1 and 2 at the national scale, the annual U.S. averages from the two datasets were differenced using the same gridding procedure described in section 3. Because the HCN version 1 release provides an optional UHI correction, two difference series were formed for each variable: (a) version 2 minus version 1 (with TOB and SHAP adjustments), and (b) version 2 minus version 1 (with TOB, SHAP, and UHI adjustments).

Fig. 11 indicates that there is a decreasing trend in the difference series for minimum temperatures before 1970. The trend is especially evident when the UHI adjustment is excluded from version 1. The existence of this trend can be traced to the impact that the SHAP adjustments on minimum temperatures in version 1. Specifically, the SHAP adjustments are limited to documented changes which have a preference for downward shifts (Table 3). When these shifts are removed, a mean warming is introduced into the SHAP-adjusted temperature record relative to the raw and TOB-only adjusted data (see also Hansen et al. 2001). Notably, the version 1 UHI adjustment "cools" HCN temperature series as a function of population growth, thereby indirectly compensating for much (but not all) of the SHAP-induced warming. In contrast, the undocumented changepoints in minimum temperatures identified in version 2 are skewed in favor of positive shifts, which collectively compensate for the negatively skewed documented shifts (the only changes known to the SHAP). For this reason, the version 2 pairwise adjustments do not increase the minimum temperature trend relative to the TOB-adjusted data (Fig. 7).

Fig. 11 also suggests a divergence between version 1 and 2 temperatures after 1985, a difference associated with the adjustments for the MMTS instrument change in version 1. As discussed in section 3b, the version 1 MMTS correction is too large when the impact of the full subset of HCN sites is considered (i.e., when stations with changes coincident to MMTS installation are included). However, as Fig. 11 indicates, maximum temperatures recover from the apparent overcorrection in version 1 after the mid-1990s. Unfortunately, this recovery is accidental; in fact, it appears to be a consequence of two factors: (a) the version 1 metadata were last updated with the Easterling et al. (1996) release; and (b) the continued conversion to MMTS (and later Nimbus), as well as the introduction of ASOS (HO 88), which have artificially (but unknown to SHAP) cooled maximum temperatures to a level that currently compensates for the version 1 overcorrection.

5. TEMPERATURE TRENDS IN THE U.S. HCN

Figure 12 depicts the U.S. annual time series for maximum, minimum, and mean ([maximum+minimum]/2) temperature during the period 1895-2006. In general, all variables exhibit a slight increase until the early 1930s, followed by a slight decrease until the early 1970s, and finally a more prominent increase into the early 21st Century. Interannual variability is markedly lower from the mid-1950s to the mid-1970s, the so-called "benign climate" period (Baker et al. 1993). For maximum temperature, the two highest ranking years are 2006 and 1934; for minimum temperature, the two highest values occurred in 1998 and 2006.

Table 4 summarizes U.S. annual and seasonal (linear) trends in maximum and minimum temperature for the raw, TOB, and fully adjusted (TOB+pairwise) version 2 data as well as the fully adjusted version 1 data (TOB+SHAP+UHI). On an annual basis, the version 2 trend in maximum temperature is 0.056 ºC dec^{-1}, and the trend in minimum temperature is 0.070 ºC dec^{-1} (both of which are comparable to the global mean trend of ~0.060 ºC dec^{-1} for the same period). Trends in both variables are largest in winter and lowest in fall, and increases in the minimum exceed those in the maximum in all seasons except spring. For reasons described in sections 3a and 3b, trends in the adjusted data always exceed those in the raw data. However, as discussed in section 3d, the version 2 trends are somewhat smaller than the fully adjusted version 1 trends (with the exception of maximum temperature in fall).

In Fig. 13, the geographic distribution of linear trends in maximum and minimum temperatures for the period 1895-2006 are shown both for the adjusted version 2 data and for the raw data. Geographically, maximum temperature (Fig. 13a) has increased in most areas except parts of the East Central and Southern regions. Minimum temperature (Fig. 13c) exhibits the same pattern of change, though the pockets of decreasing temperature are displaced slightly to the south and west relative to maximum temperature. Figures 13b and 13d suggest that the raw data exhibit more extreme trends as well as larger spatial variability; in other words, the bias adjustments tend to have a spatial smoothing effect on rates of change. The reduction in the extent of negative trends across eastern parts of the country is a function of removing the time of observation bias and of the adjustments associated with the MMTS instrument change.

In spite of the improved pattern, Pielke et al. (2007b) suggest that homogenized data are inappropriate for calculating regional trends because the homogenized series lack independence. This is surprising considering that the unadjusted temperature series are themselves highly correlated, with monthly first difference (Peterson et al. 1998) correlation coefficients generally exceeding 0.9 at station densities found in the United States. While the homogenization process does enhance the correlation between series, the increase is negligible (i.e., <0.01 on average) over that between raw series. It is likely for this reason that Vose and Menne (2004) found that the same basic relationship exists between station density and the error in calculating the mean U.S. temperature trend whether unadjusted or adjusted data are used. Moreover, the Vose and Menne (2004) assessment of the network density required to capture the overall U.S. trend is about an order of magnitude less than the current configuration of the HCN. This suggests that the HCN should be sufficient to calculate regional trends in most areas. In any case, all Coop temperature series are homogenized by the version 2 pairwise algorithm, which greatly expands the pool of adjusted series. Consequently, if there is a concern over the characteristics of a particular HCN site or inadequate station density in some areas, adjusted Coop temperature series may be helpful. This is only one of the benefits of this unique climate network, made possible by the efforts of dedicated Cooperative Observer volunteers for well over a century.

6. SUMMARY AND CONCLUSION

Overall, the collective impact of changes in observation practice at U.S. HCN stations is of the same order of magnitude as the background climate signal (e.g., artificial bias in maximum temperatures is about -0.039°C dec^{-1} compared to the background trend of 0.056°C dec^{-1}). Consequently, bias adjustments are essential in reducing the uncertainty in U.S. climate trends. The bias changes that have had the biggest impact on the climate network include changes to the time of observation (which impacts both maximum and minimum temperature trends) and the widespread conversion to the MMTS (which impacts primarily maximum temperatures). Adjustments for undocumented changes are especially important in removing bias in minimum temperature records.

Trends in the adjusted series are more spatially uniform than in unadjusted data. This indicates that the homogenization procedures remove changes in relative bias and that the background climate signal is more accurately represented by the homogenized data. It is important to point out, however, that while homogenization generally ensures that climate *trends* can be confidently inter-compared between sites, the impact of relative biases will still be reflected in the *mean* temperatures of homogenized series. The reason is that, by convention, temperatures are adjusted to conform to the latest (i.e., current) observing status at all stations. This detail helps to explain why Peterson and Owen (2005) found evidence of a systematic difference in mean temperatures at rural versus urban HCN stations but little evidence of a comparable difference in their homogenized trends.

Finally, while changes in observation practice have clearly had a systematic impact on average U.S. temperature trends, homogeneity matters most at the station level where even one change in bias can have a drastic impact on the series trend. Consequently, the goal behind the version 2 adjustments was to make each of them site-specific, which is especially valuable in the development of widely used products such as the U.S. Climate Normals. Nevertheless, there is always room for improvement. For example, while the monthly adjustments used in version 2 are constant for all months, there is evidence that bias changes often have impacts that vary seasonally and/or synoptically (Trewin and Trivitt 1995; Guttman and Baker 1996). As shown by Della-Marta and Wagner (2006), it is possible to estimate the differential impacts indirectly by evaluating the magnitude of change as a function of the frequency distribution of daily temperatures. Daily adjustments are thus a promising area for future HCN development.

7. ACKNOWLEDGEMENTS

The authors wish to thank Tom Peterson and Tamara Houston for helpful comments on this manuscript. Partial support for this work was provided by the Office of Biological and Environmental Research, U.S. Department of Energy (Grant number DE-AI02-96ER62276).

8. REFERENCES

Baker, D. G., 1975: Effect of observation time on mean temperature estimation. *J. Appl. Meteor.*, **14**, 471–476.

Baker, D.G., D.L. Rushy, and R.H. Skaggs, 1993: Agriculture and the recent benign climate. *Bull. Amer. Meteor. Soc.*, **74**, 1035-1040.

CDMP, 2001: Annual report. 8 pp. [Available online at http://www.ncdc.noaa.gov/oa/climate/cdmp/files/annualreport2001.pdf. c

Collins, DA, S. Johnson, N. Plummer, A.K. Brewster, Y. Kuleshov, 1999: Re-visiting Tasmania's reference climate stations with a semi-objective network selection scheme. *Australian Meteorological Magazine*, **48**, 111-122.

Davey, C.A., and R.A. Pielke Sr., 2005: Microclimate exposure of surface-based weather stations. *Bull. Amer. Meteor. Soc.*, **86**, 497-504.

DeGaetano, A. T., 1999: A serially complete simulated observation time metadata file for U.S. Daily Historical Climatology Network stations. *Bull. Amer. Meteor. Soc.*, **81**, 49– 67.

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

Della-Marta, P.M. and H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179-4197.

Durre, I., M.J. Menne, and R.S. Vose, 2008: Strategies for evaluating quality assurance procedures. *J. Appl. Meteor. Climatol.*, **47**, 1785-1791.

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.

Guttman, N. B., and C. B. Baker, 1996: Exploratory analysis of the difference between temperature observations recorded by ASOS and conventional methods. *Bull. Amer. Meteor. Soc.,* **77,** 2865–2873.

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

Hubbard, K. G., X. Lin, 2006: Reexamination of instrument change effects in the US Historical Climatology Network. *Geophys. Res. Lett.,* **33,** L15710, doi:10.1029/2006GL027069.

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

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

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

Mahmood, R., 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.

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

Menne, M.J., and C.N. Williams, Jr., submitted: Homogenization of temperature series via pairwise comparisons. *J. Climate*, in review.

Peterson, T.C., T.R. Karl, P.F. Jamason, R. Knight, and D.R. Easterling, 1998b: First difference method: Maximizing station density for the calculation of the long-term global temperature change. *J. Geophys. Res.*, **74**, 22967-22974.

Peterson, T.C. and T.W. Owen, 2005: Urban heat island assessment: Metadata are important. *J. Climate*, **18**, 2637-2646.

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

Pielke, R.A., Sr., 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. *Bull. Amer. Meteor. Soc.*, **88**, 913–928.

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. Amer. Meteor. Soc.*, **72**, 1718-1723.

Quinlan, F. T., T. R. Karl, and C. N. Williams, Jr. 1987. United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data. NDP-019, Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.

Trewin, B.C., and A.C.F. Trevitt, 1996: The development of composite temperature records. *Int. J. Climatol*., **16**, 1227-1242.

Vose, R.S., C.N. Williams, T.C. Peterson, T.R. Karl, and D.R. Easterling, 2003: An evaluation of the time of observation bias adjustment in the US Historical Climatology Network. *Geophys. Res. Lett.*, **30**, doi:10.1029/2003GL018111.

Vose, R.S., and M.J. Menne, 2004: A method to determine station density requirements for climate observing networks. *J. Climate*, **17**, 2961-2971.

Willmott, C. J., C. M. Rowe, and W. D. Philpot, 1985: Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring. *Am. Cartographer*, **12**, 5–16.

Table 1. Quality assurance checks applied to daily data.

Data Problem	Description of Check
Simultaneous zeros	Identifies days on which both maximum and minimum temperature are -17.8°C (0°F)
Duplication of data	Identifies duplication of data between entire years, different years in the same month, different months within the same year, and maximum and minimum temperature within the same month
Impossible value	Determines whether a temperature exceeds known world records
Streak	Identifies runs of the same value on > 15 consecutive days
Gap	Identifies temperatures that are at least 10°C warmer or colder than all other values for a given station and month
Climatological outlier	Identifies daily temperatures that exceed the respective 15-day climatological means by at least six standard deviations
Internal inconsistency	Identifies days on which the maximum temperature is less than the minimum temperature
Interday inconsistency	Identifies daily maximum temperatures that are less than the minimum temperatures on the preceding, current, and following days as well as for minimum temperatures that are greater than the maximum temperatures during the relevant three-day window
Lag range inconsistency	Identifies maximum temperatures that are at least 40°C warmer than the minimum temperatures on the preceding, current, and following days as well as minimum temperatures that are at least 40°C colder than the maximum temperatures within the three-day window
Temporal inconsistency	Determines whether a daily temperature exceeds that on the preceding and following days by more than 25°C
Spatial inconsistency	Identifies temperatures whose anomalies differ by more than 10°C from the anomalies at neighboring stations on the preceding, current, and following days
"Mega" inconsistency	Looks for daily maximum temperatures that are less than the lowest minimum temperature and for daily minimum temperatures that are greater than the highest maximum temperature for a given station and calendar month

Table 2. Quality assurance checks applied to monthly data.

Data Problem	Description of Check
Climatological outlier	Identifies temperatures that exceed their respective climatological means for the corresponding station and calendar month by at least five standard deviations
Spatial inconsistency	Compares z-scores (relative to their respective climatological means) to concurrent z-scores at the nearest 20 neighbors located within 500 km of the target; a temperature fails if (1) its z-score differs from the regional (target and neighbor) mean z-score by at least 3.5 standard deviations and (2) the target's temperature anomaly differs by at least 2.5°C from all concurrent temperature anomalies at the neighbors
Spatial inconsistency	Identifies temperatures whose anomalies differ by more than 4°C from concurrent anomalies at the five nearest neighboring stations whose temperature anomalies are well-correlated with the target (correlation > 0.7 for the corresponding calendar month)

Table 3. Summary of shifts in HCN mean monthly maximum and minimum temperature series. A negative shift indicates that the inhomogeneity led to a decrease in the mean level of the temperature series relative to preceding values.

	Maximum Temperature				**Minimum Temperature**			
	Positive Shifts		Negative Shifts		Positive Shifts		Negative Shifts	
	Count	Average	Count	Average	Count	Average	Count	Average
All Changepoints	3057	1.50	3430	-1.54	3957	1.51	3897	-1.58
Undocumented Changepoints	1786	1.50	1836	-1.53	2509	1.48	2210	-1.51
Documented Changepoints	1271	1.51	1594	-1.55	1448	1.57	1687	-1.66
LiG to MMTS Changepoints	52	0.73	231	-0.85	241	0.93	104	-0.85

Table 4. U.S. annual and seasonal temperature trends (°C dec^{-1}) 1895 to 2006 for adjusted and unadjusted temperature series.

Season	Maximum Temperature	Minimum Temperature
	Fully Adjusted – Version 2 (TOB+Pairwise)	
Annual	0.056	0.070
D-J-F	0.094	0.101
M-A-M	0.069	0.059
J-J-A	0.036	0.064
S-O-N	0.014	0.046
	Unadjusted (Raw) – Version 2	
Annual	0.017	0.052
D-J-F	0.047	0.066
M-A-M	0.030	0.046
J-J-A	0.005	0.062
S-O-N	-0.024	0.028
	Adjusted for TOB only – Version 2	
Annual	0.029	0.070
D-J-F	0.067	0.102
M-A-M	0.042	0.060
J-J-A	0.008	0.063
S-O-N	-0.014	0.046
	Fully Adjusted – Version 1 (TOB+SHAP+UHI)	
Annual	0.061	0.089
D-J-F	0.101	0.109
M-A-M	0.090	0.085
J-J-A	0.038	0.085
S-O-N	0.008	0.069

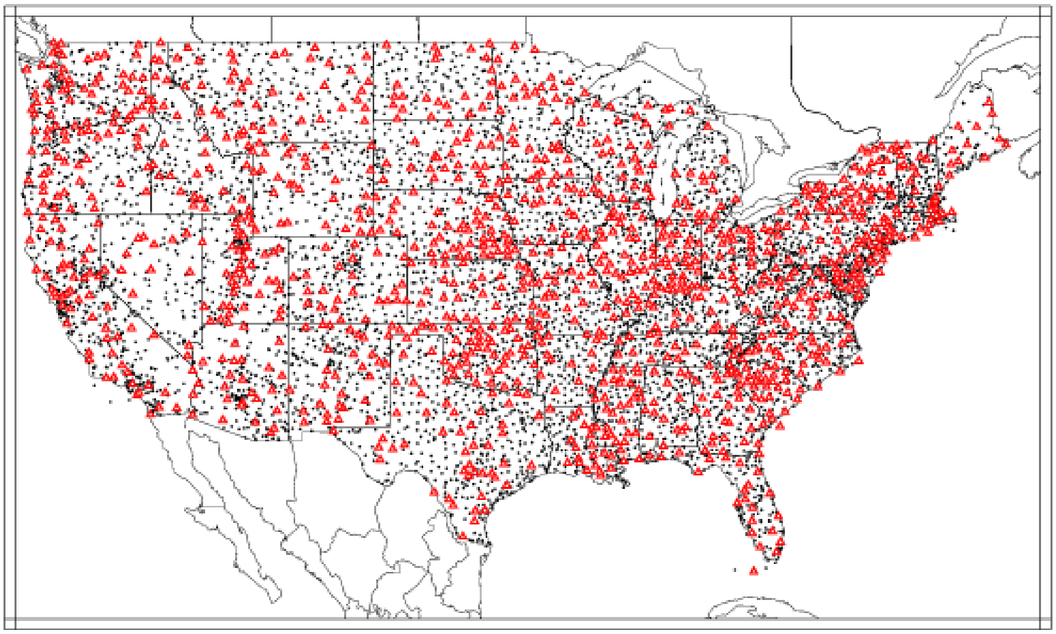

Figure 1. Distribution of Coop Observer Network stations (black dots) and the U.S. HCN version 2 sites (red triangles).

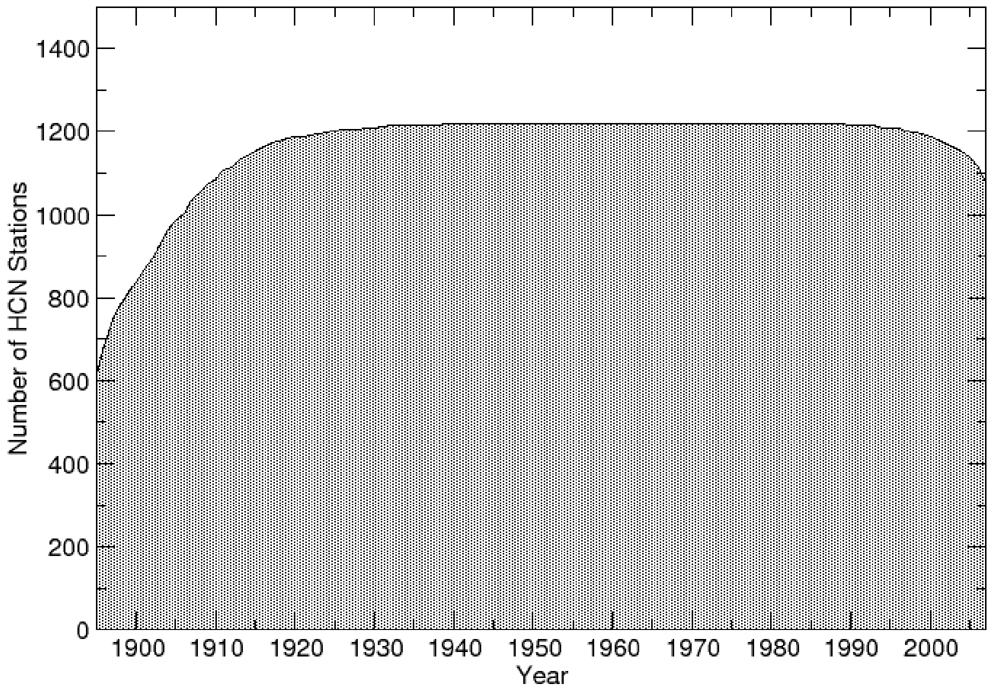

Figure 2. Number of U.S. HCN stations by year.

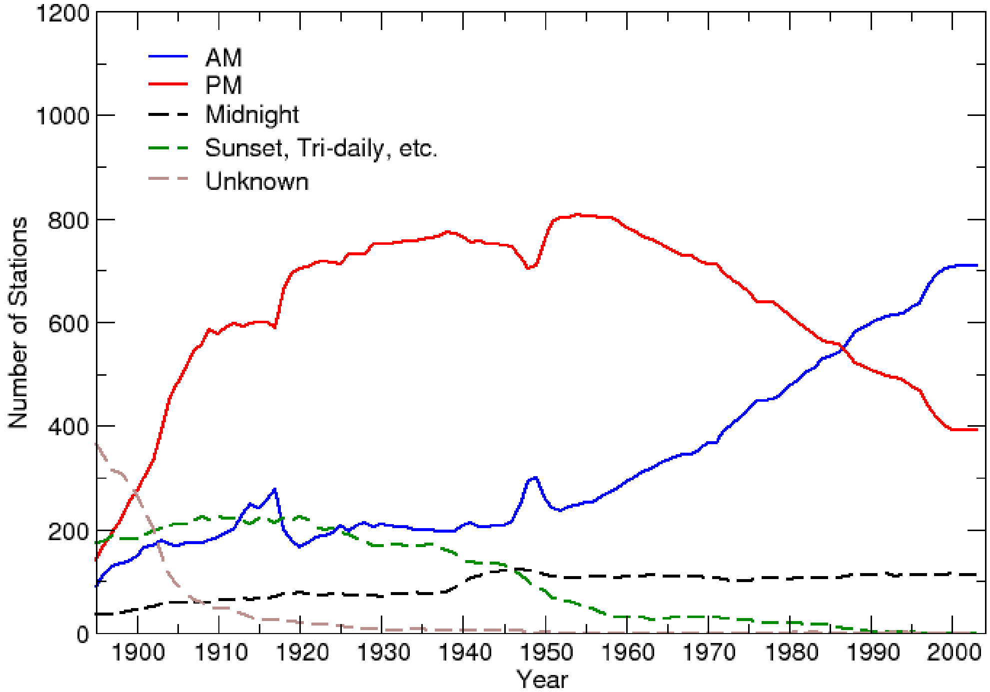

Figure 3. Changes in the documented time of observation in the U.S. HCN.

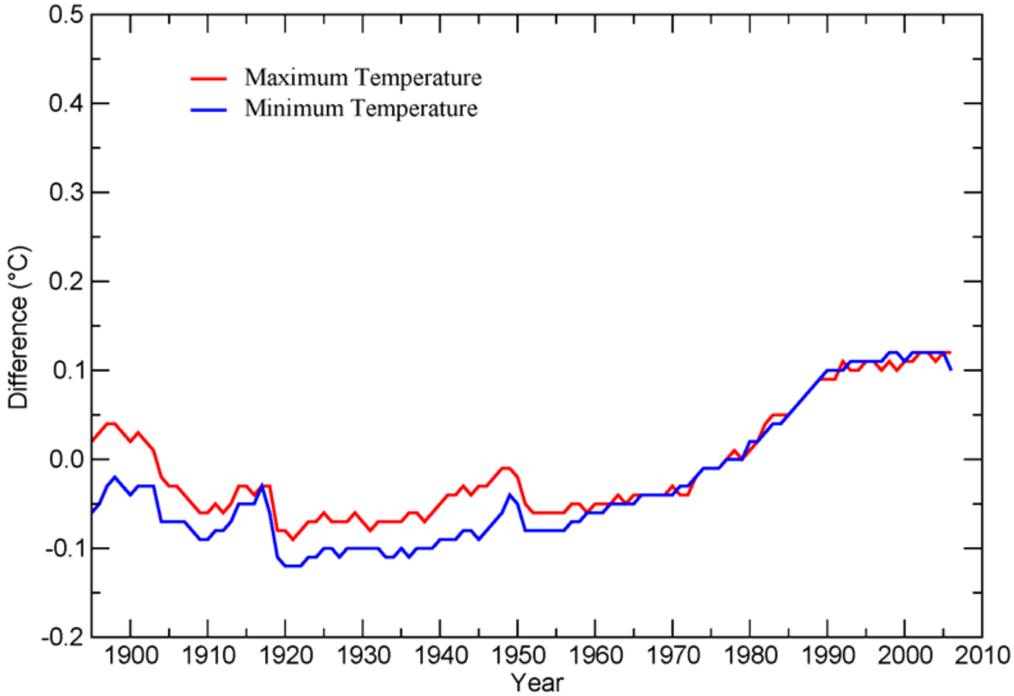

Figure 4. Average year by year difference over the conterminous United States between the TOB-adjusted data and the unadjusted (raw) data.

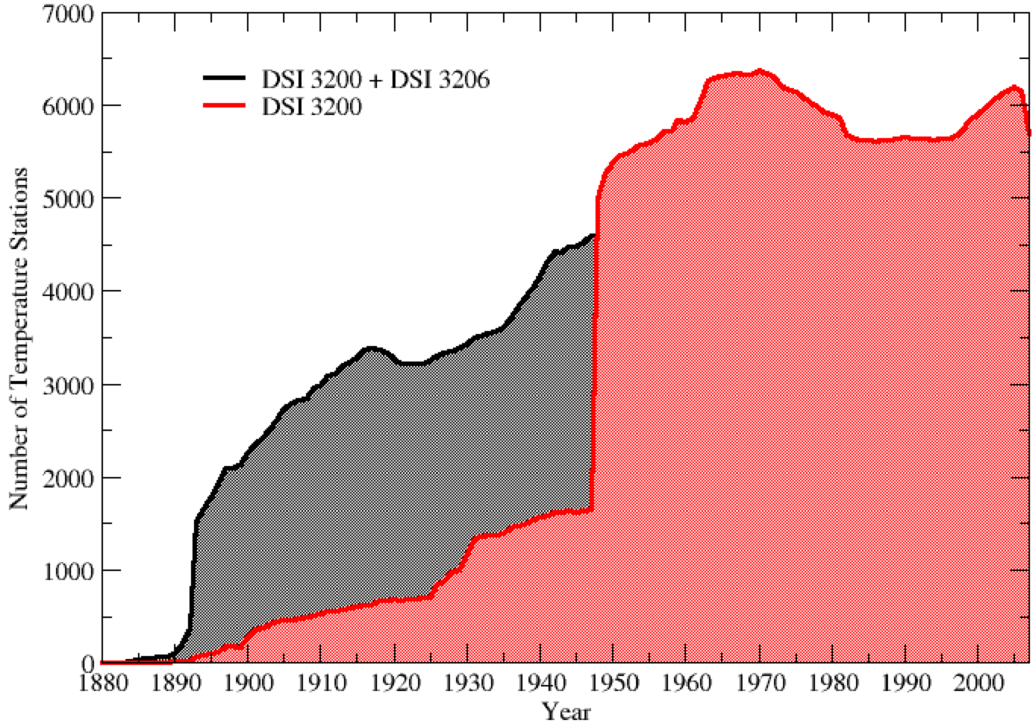

Figure 5. Digital monthly data availability for U.S. Cooperative Observer stations before (DSI 3200) and after (DSI 3200 + 3206) the digitization efforts of the Climate Data Modernization Program.

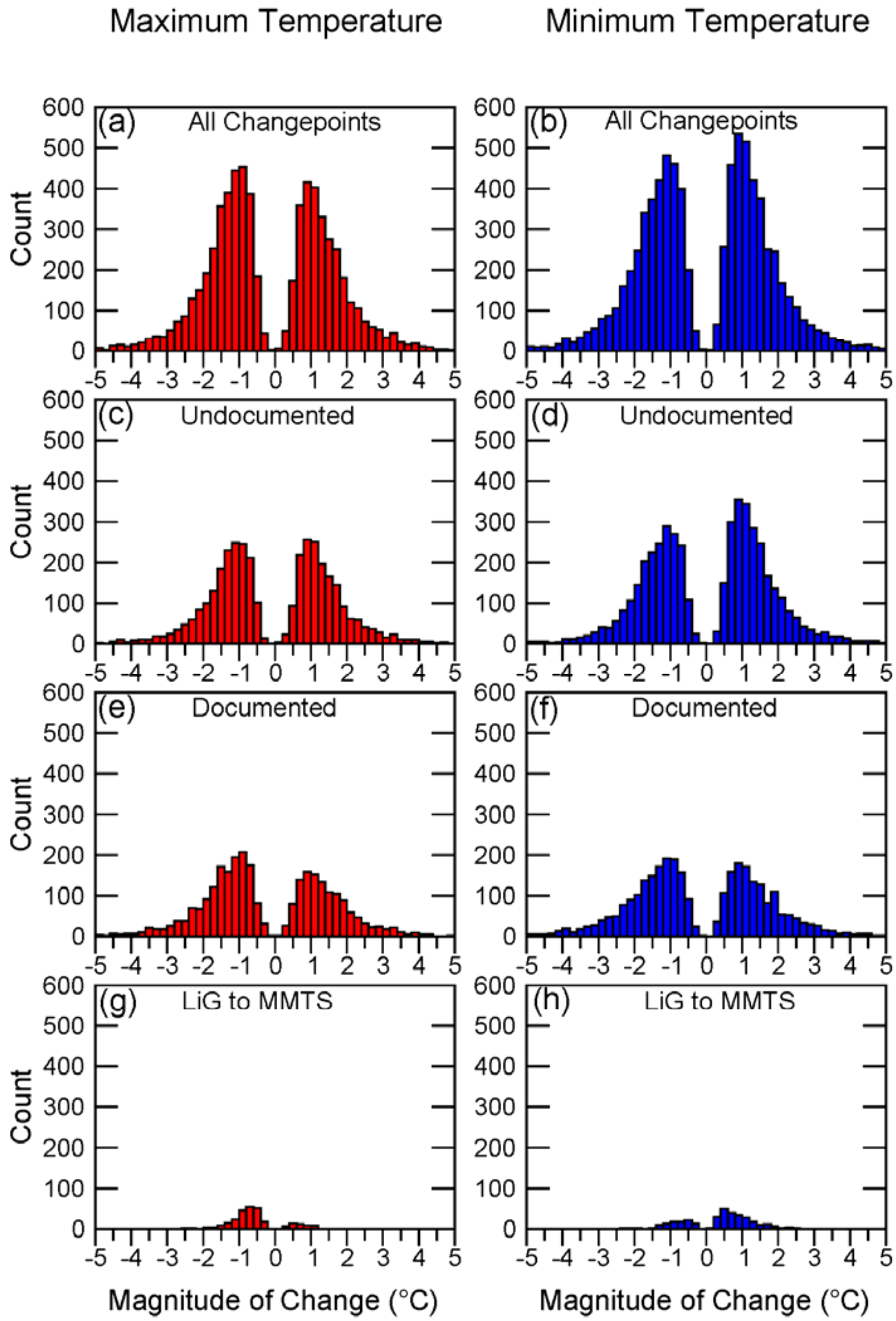

Figure 6. Histograms of the magnitude of changepoints (shifts) in U.S. HCN mean monthly maximum and minimum temperature series. (a) and (b) – all changepoints; (c) and (d) – undocumented changepoints; (e) and (f) – changepoints associated with documented station changes; (g) and (h) – changepoints associated with the transition from Liquid in Glass (LiG) thermometers to the Maximum Minimum Temperature System (MMTS). A negative shift indicates that the inhomogeneity led to a decrease in the mean level of the temperature series relative to preceding values.

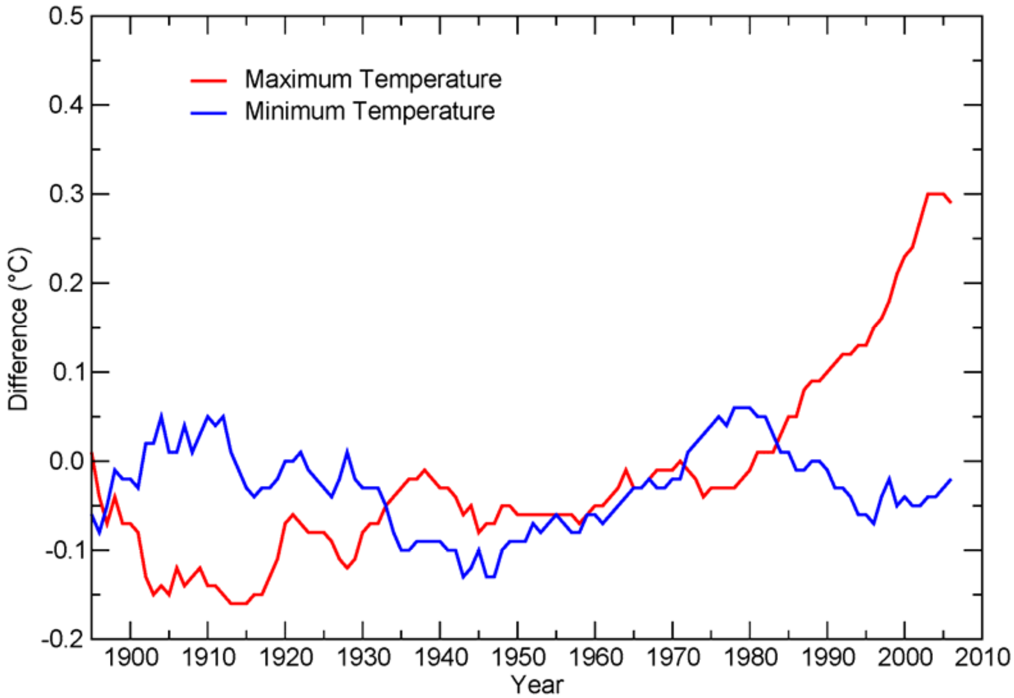

Figure 7. Year by year differences averaged over the conterminous United States between the fully adjusted (TOB+pairwise) HCN data and the TOB-only adjusted data.

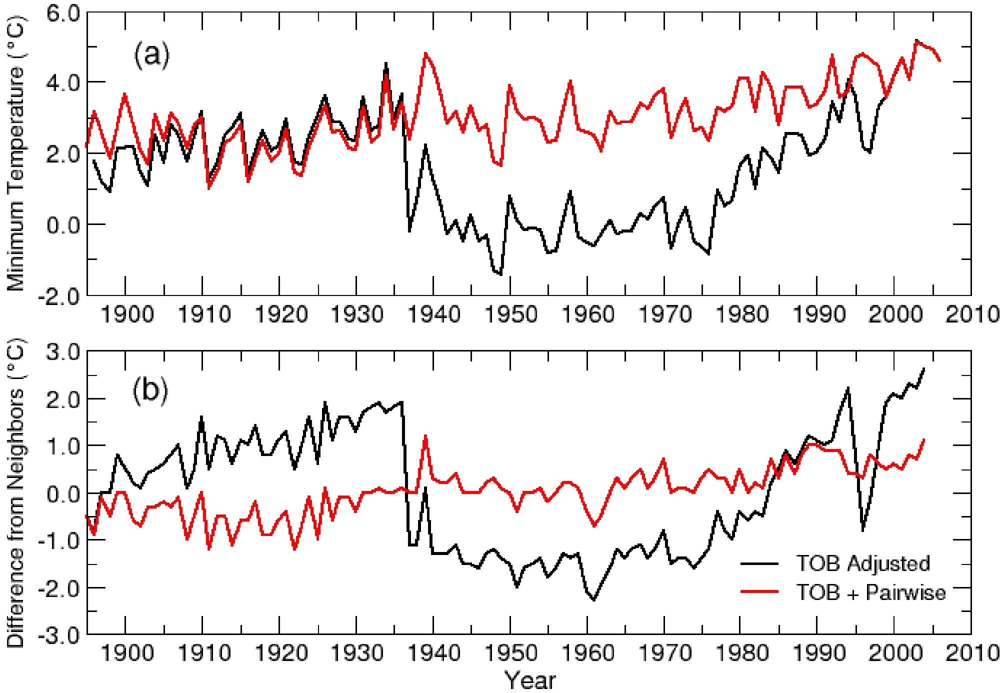

Figure 8. (a) Mean annual TOB and fully adjusted (TOB+Pairwise) minimum temperatures at Reno, Nevada; (b) difference between minimum temperatures at Reno and the mean from its 10 nearest neighbors.

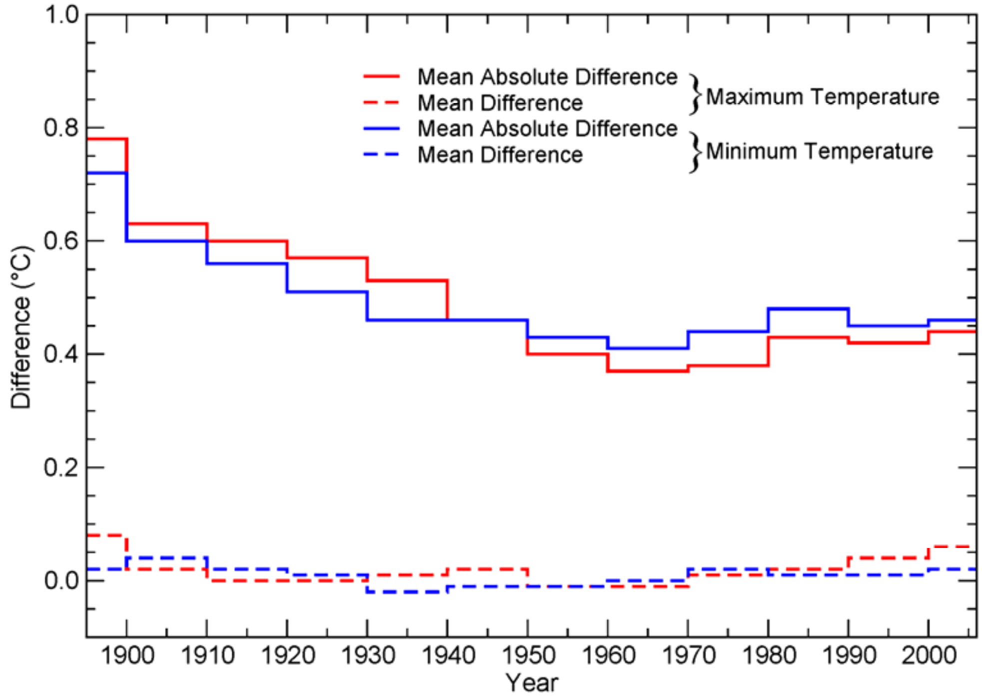

Figure 9. Difference (by decade) between FILNET estimates and observed monthly values at all U.S. HCN stations.

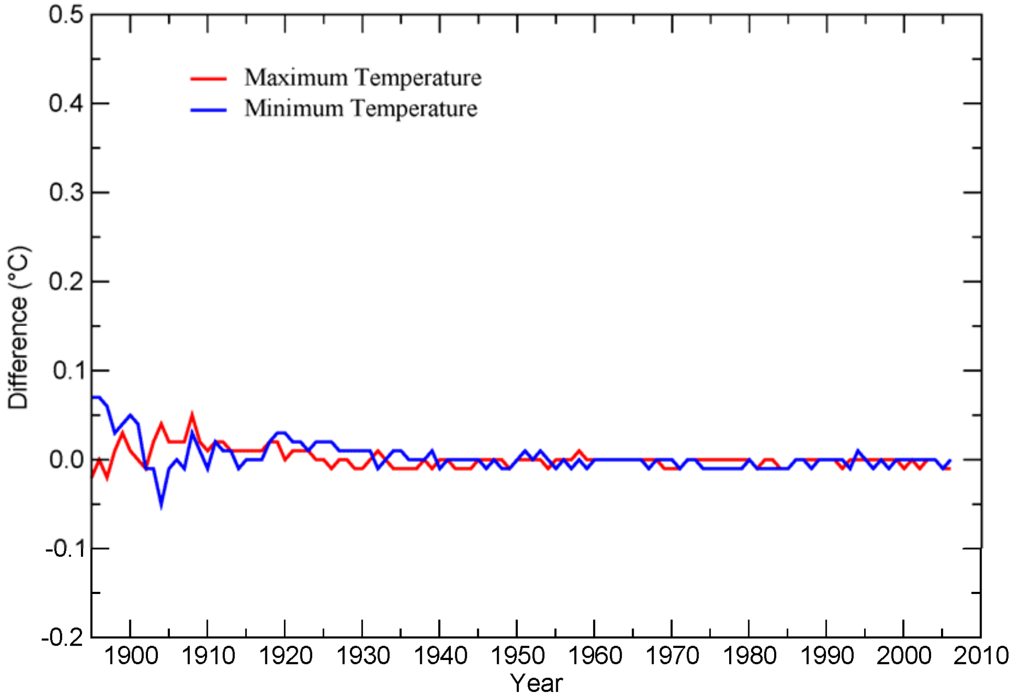

Figure 10. Year by year differences between the fully adjusted HCN data with estimates for missing values (TOB+pairwise+FILNET) and the fully adjusted data without missing data estimates (TOB+pairwise).

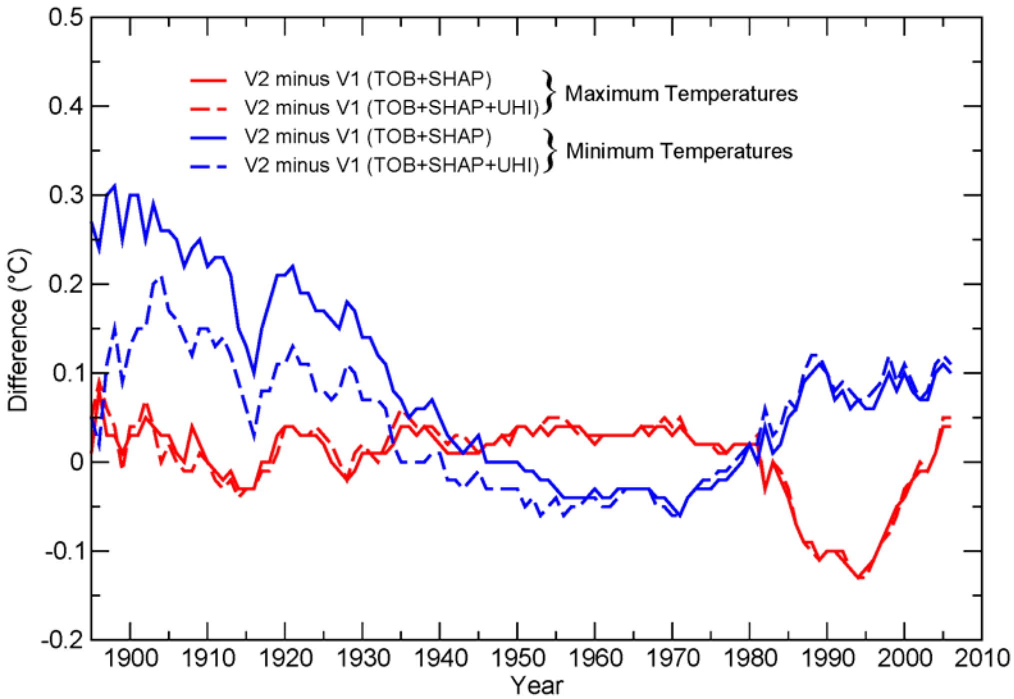

Figure 11. Year by year differences between U.S. average temperatures in HCN version 2 and HCN version 1 (Revision 3; Easterling et al. 1996).

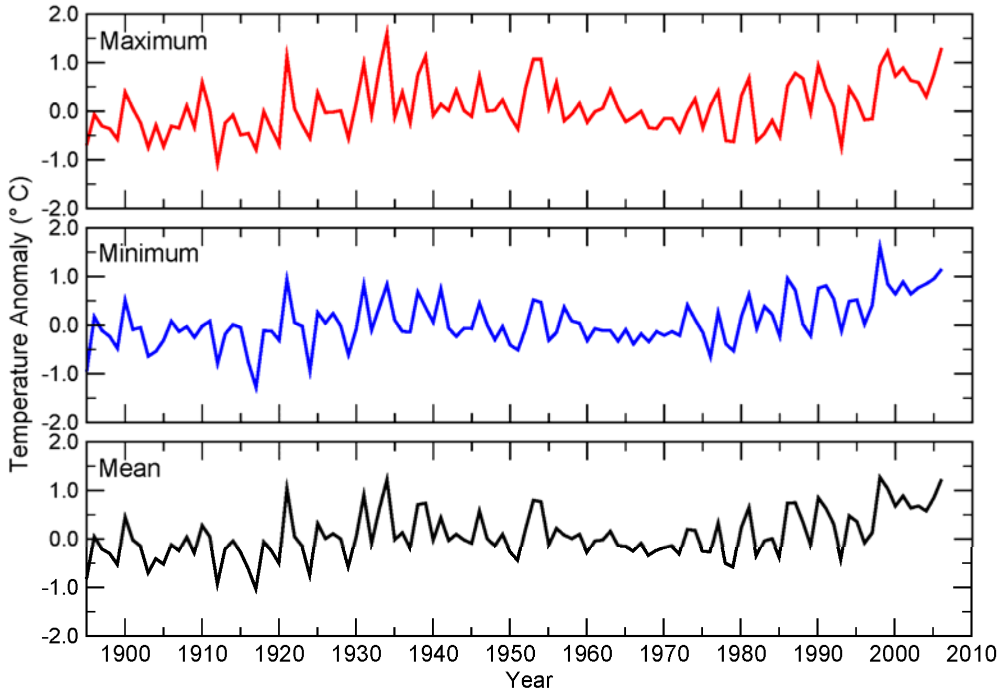

Figure 12. Time series of annual temperature anomalies from HCN version 2 averaged over the conterminous United States. Base period is 1961 to 1990.

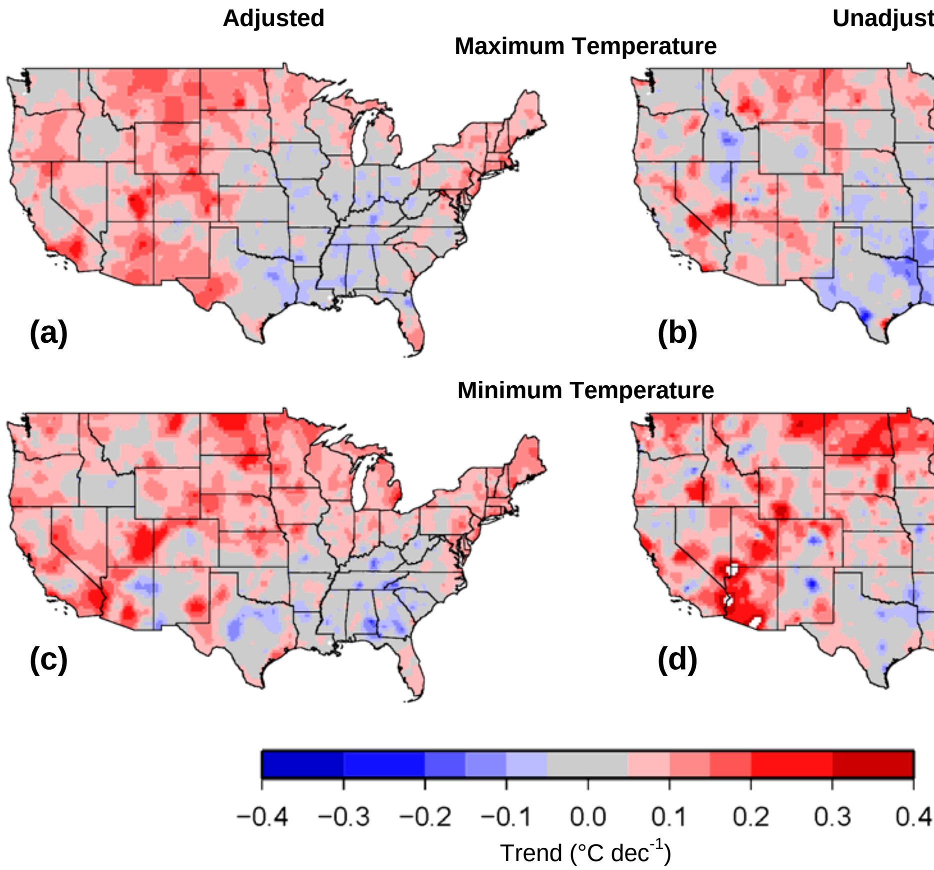

Figure 13. Geographic distribution of linear trends in HCN version 2 temperatures for the period 1895 to 2006. Areas in white indicate trends in excess of 0.4°C dec^{-1}: (a) adjusted maximum temperatures; (b) unadjusted maximum temperatures; (c) adjusted minimum temperatures; (d) unadjusted minimum temperatures.