THE U.S. HISTORICAL CLIMATOLOGY NETWORK MONTHLY TEMPERATURE DATA, VERSION 2

BY MATTHEW J. MENNE, CLAUDE N. WILLIAMS JR., AND RUSSELL S. VOSE

New bias adjustments reduce uncertainty in temperature trends for the United States.

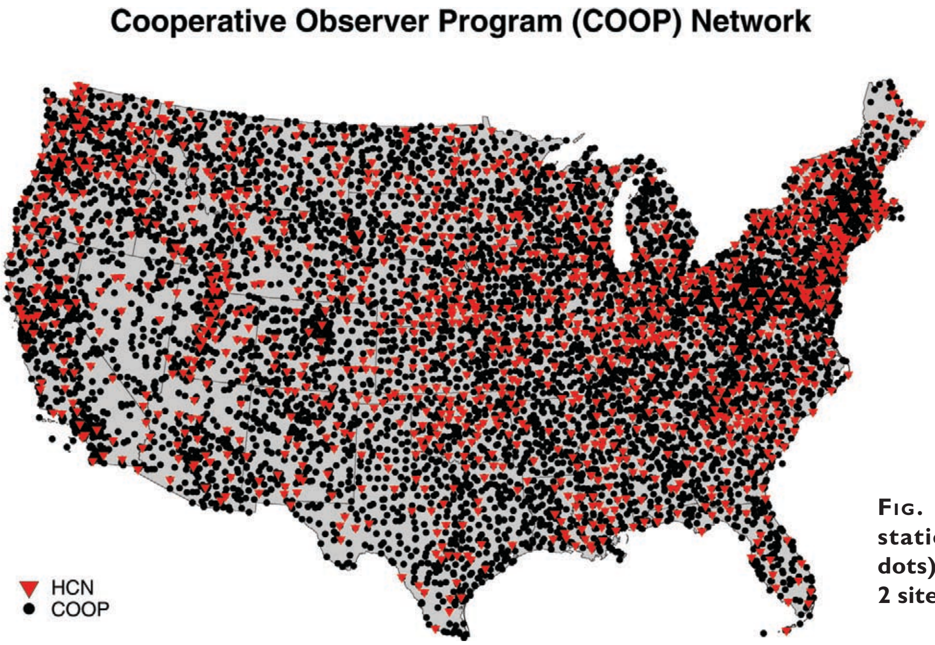

FIG. 1. Distribution of COOP stations in the CONUS (black dots) and the U.S. HCN version 2 sites (red triangles).

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

In support of its climate monitoring and assessment activities, NCDC has recently developed an improved U.S. HCN dataset (hereafter called 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 effect on U.S. temperature trends and in terms of the differences between version 2 and its widely used predecessor (now termed HCN version 1).

DATA. *Network development.* The U.S. HCN is a reference station network (Collins et al. 1999), that is, 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 station records (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 circa 1987 and to have a period of record of at least 80 years. In practice, these criteria were sometimes relaxed 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 1,219 COOP stations, 84 of which were composites formed using consecutive records from two or more stations to achieve the minimum period of record goal.

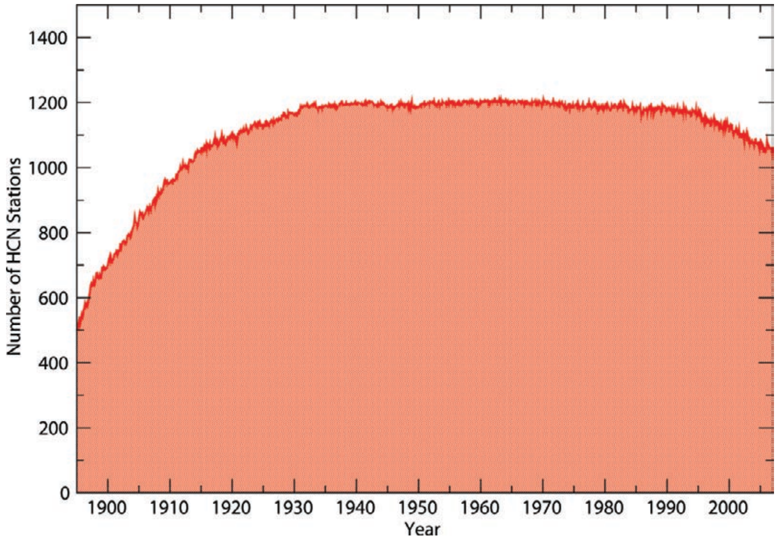

Fig. 2. Number of U.S. HCN stations with temperature records.

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 the U.S. HCN as well as the creation of additional composite stations. The reevaluation led to 52 station deletions and 54 additions, for a total of 1,221 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 contains 1,218 stations, 208 of which are composites; relative to the 1996 release, there have been 62 station deletions and 59 additions.

Figure 1 depicts the locations of the 1,218 stations in HCN version 2. Consistent with previous releases, the spatial distribution is reasonably uniform across the CONUS, although station density is higher across the eastern CONUS than in the intermountain west. Moreover, 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 reporting until about 1925, with spatial coverage increasing most prominently in the west during these early years. The number of stations reporting remained relatively consistent until the end of the twentieth century, after which it has declined because of station closures.

Source data. To maximize data completeness, HCN version 2 was derived from the following 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 source contains "estimated" values and quality assurance (QA) flags; however, 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 subjected to the suite of 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 thresholds were selected and the performance of each check was evaluated using the

AFFILIATIONS: Menne, Williams, and Vose—NOAA/NCDC, Asheville, North Carolina
CORRESPONDING AUTHOR: Matthew J. Menne, NOAA/NCDC, 151 Patton Ave., Asheville, NC 28801
E-mail: matthew.menne@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/2008BAMS2613.1

In final form 5 December 2008

method outlined in Durre et al. (2008). 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 sources were subsequently merged by COOP station 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 monthly QA checks 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 fewer than 0.2% of monthly maximum and minimum temperature values.

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 3-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 3-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

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

The process of removing systematic changes in the bias of 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.

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 1975; Karl et al. 1986). This "time of observation bias" would be of little concern with regard to tempera-

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 (i) its z score differs from the regional (target and neighbor) mean z score by at least 3.5 standard deviations and (ii) 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)

ture 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 since that time to satisfy operational hydrological requirements. The result has been 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 HCN version 2 monthly temperatures, the adjustment method described in Karl et al. (1986) was used. The robustness of this method, which was also used to produce version 1, 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 (DeGaetano 2000). In addition, the predictive skill of the Karl et al. (1986) approach to estimating the TOB was confirmed using hourly data from 500 stations during the period 1965–2001 (whereas the approach was originally developed using data from 79 stations during the period 1957–64). Given these verifications, the Karl et al. (1986) TOB adjustment procedure was used in HCN version 2 without modification.

To calculate the effect 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 1961–90 station mean. Anomalies were then interpolated to the nodes of a 0.25° × 0.25° latitude–longitude grid using the method described by Willmott et al. (1985). Finally, gridpoint values were area weighted into a mean anomaly for the CONUS 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 0.015°C decade^{-1} (±0.002) and in minimum temperatures by about 0.022°C decade^{-1} (±0.002) during the period

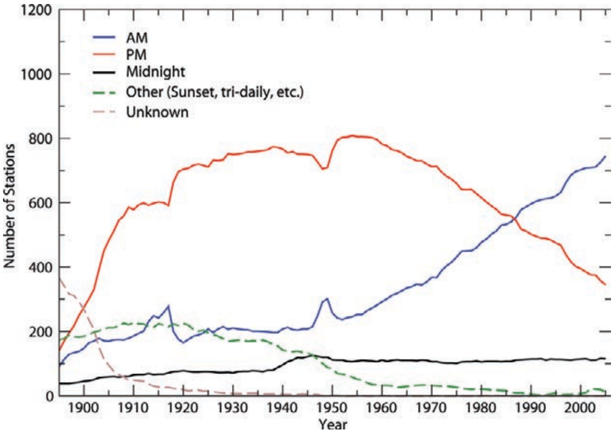

Fig. 3. Changes in the documented time of observation in the U.S. HCN.

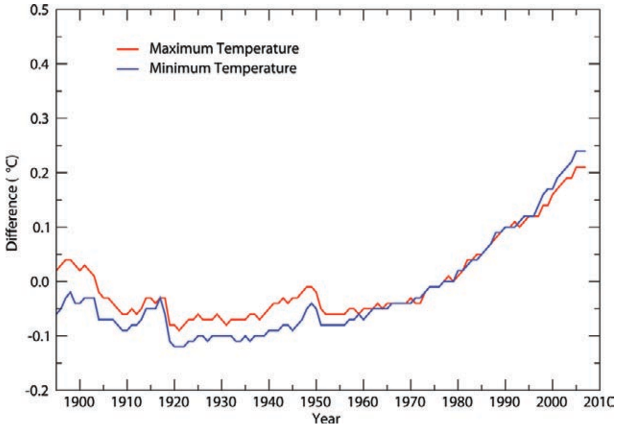

Fig. 4. Average annual differences over the CONUS between the TOB-adjusted data and the unadjusted (raw) data.

1895–2007. This net effect is about the same as that of the TOB adjustments in the HCN version 1 temperature data (Hansen et al. 2001), which is to be expected since the same TOB-adjustment method is used in both versions.

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 HCN version 1, the effects 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 HCN version 1 homogeneity algorithm was called 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 2009). This new algorithm addresses both documented and undocumented discontinuities via a pairwise comparison of temperature records, which avoids problems associated with the use of reference series in undocumented changepoint detection (Menne and Williams 2005). In the 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).

The pairwise approach works best when there are many neighboring series available for comparison with each target series. Thus, to maximize the number of potential neighbors for each HCN station, all COOP temperature series were used as input by the pairwise algorithm. In contrast, the SHAP used in HCN version 1 was restricted to intercomparing only HCN series, in large part because digital monthly COOP temperature data (and metadata) were more limited back in the 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 Fig. 5. As noted in the "Data" section, these historical temperature values were merged with other COOP Network data sources, which effectively increased the density of the observations (as well as the correlation between all series tested), thereby improving the ability of the pairwise algorithm to detect relative inhomogeneities.

As in HCN version 1, homogeneity testing in HCN version 2 was conducted separately on monthly-mean maximum and minimum temperature series. Figure 6 depicts the frequency and magnitude of shifts detected by the pairwise algorithm for each variable. Overall, the pairwise algorithm identified around 6,000 statistically significant changepoints in maximum temperature series and roughly 7,000 shifts in minimum temperature series. Since there are approximately 120,000 station years of temperatures in the HCN version 2 dataset, this represents an aver-

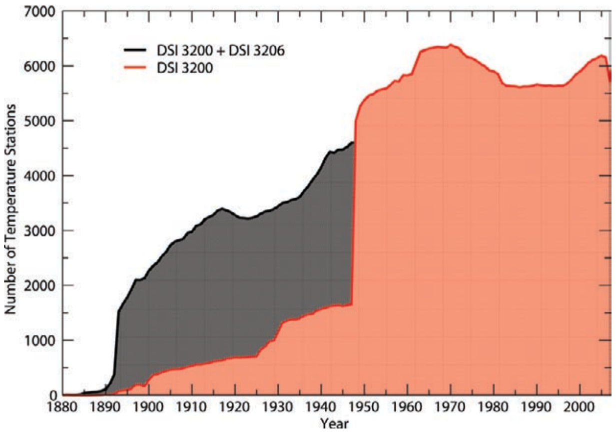

Fig. 5. Digital data availability for COOP stations before (DSI 3200) and after (DSI 3200 + 3206) the digitization efforts of the Climate Data Modernization Program.

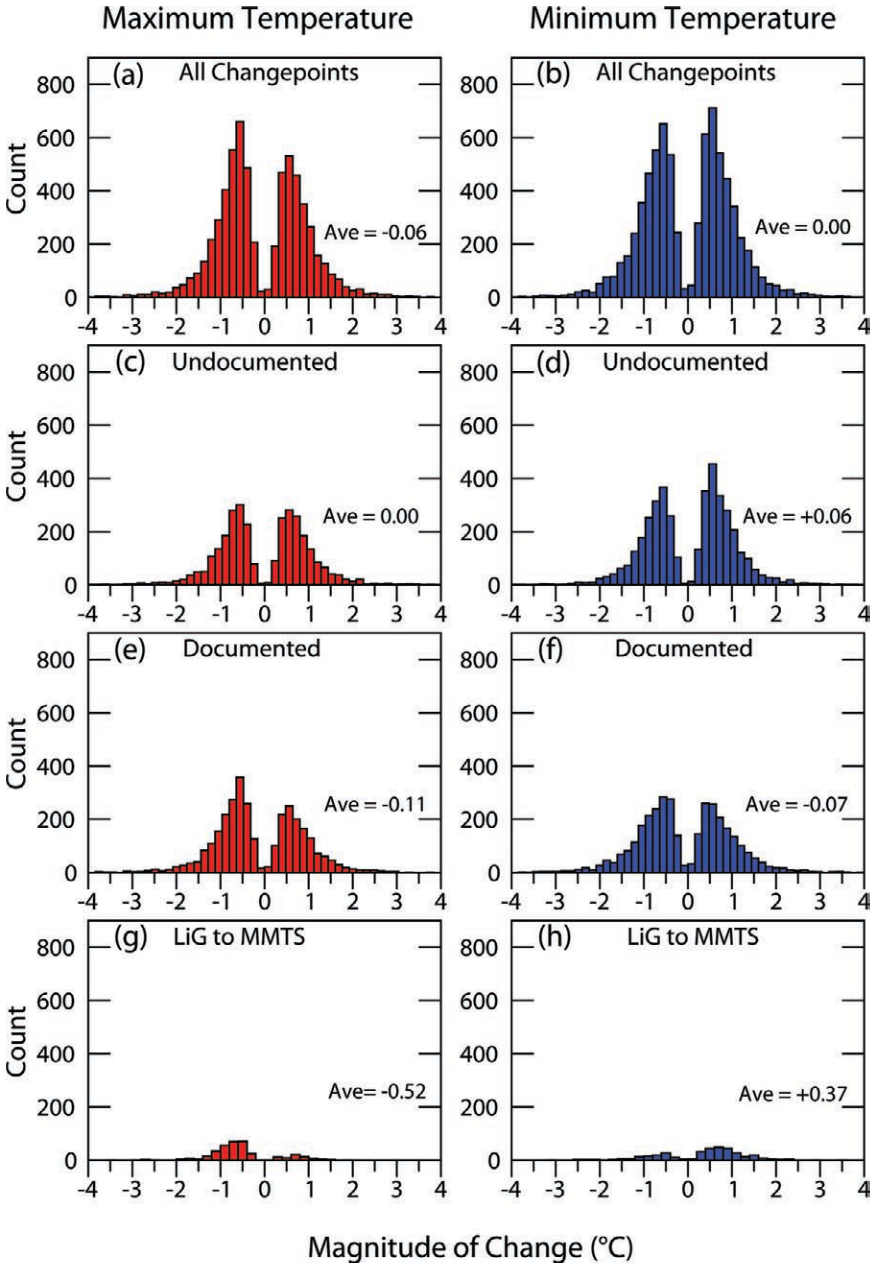

Fig. 6. Histograms of the magnitude of changepoints (shifts) in U.S. HCN mean monthly maximum and minimum temperature series: (a), (b) all changepoints; (c), (d) undocumented changepoints; (e), (f) changepoints associated with documented station changes; (g), (h) changepoints associated with the transition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

age of about one significant artificial shift for every 15–20 years of station data. In terms of the adequacy of the HCN metadata, about half of the identified inhomogeneities are undocumented.

Most of the documented changes in the HCN are associated with station relocations. In theory, minor station moves or other changes to 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.

Whereas 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). 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 effect on HCN temperatures and advocated for site specific adjustments in general, including those sites with no documented equipment move.

Notably, the pairwise algorithm in HCN 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 40% of the maximum and minimum temperature series experienced a statistically significant shift (out of ~850 total conversions to MMTS). As a result, the overall effect 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 effect of the statistically significant changes (−0.52°C for maximum temperatures and +0.37°C for minimum temperatures) is close to Hubbard and Lin's (2006) results for sites with no coincident station move.

For HCN version 2 as a whole, the combined effect of all adjustments for documented and undocumented temperature changes is to increase the average U.S. trend in maximum temperatures by about 0.031°C decade^{-1} (±0.007) over the period of record relative to the values adjusted only for the TOB (Fig. 7). In contrast, the effect of the pairwise homogenization algorithm on minimum temperature trends is effectively zero over the period of record. As Fig. 7 indicates, the most significant effect 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 effect of the individual instrument changes.

Although 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 HCN 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 HCN 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 effect 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 effect 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 partial cancellation effect between the two instrument changes. Undocumented changes, which are skewed in favor of positive shifts, further mitigate the effect of the MMTS on minimum temperatures.

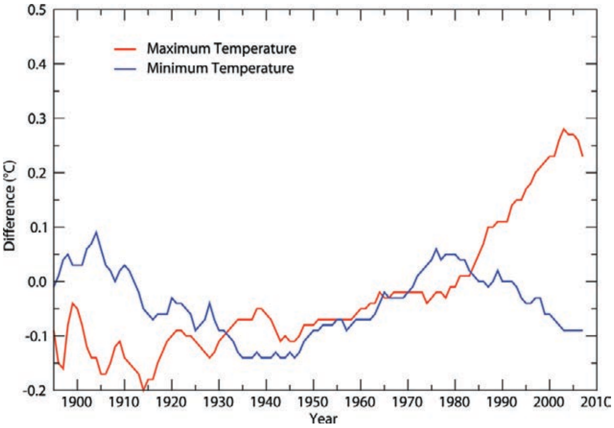

Fig. 7. Average annual differences over the CONUS between the fully adjusted (TOB + pairwise) HCN data and the TOB-only adjusted data.

Bias associated with urbanization and nonstandard siting. In HCN version 1, the regression-based approach of Karl et al. (1988) was employed to account for the effect of the urban heat island (UHI) bias on temperatures in the HCN (which they found to be important for minimum temperatures only). In contrast, no specific urban correction is applied in HCN version 2. The reason is that adjustments for undocumented changepoints in HCN version 2 appear to account for much of the changes addressed by the Karl et al. (1988) UHI correction used in HCN version 1. In fact, as discussed in the next section, including adjustments for undocumented changepoints actually has a greater impact on minimum temperatures than the HCN version 1 UHI correction. Moreover, 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 unadjusted data suggest that the station developed a local

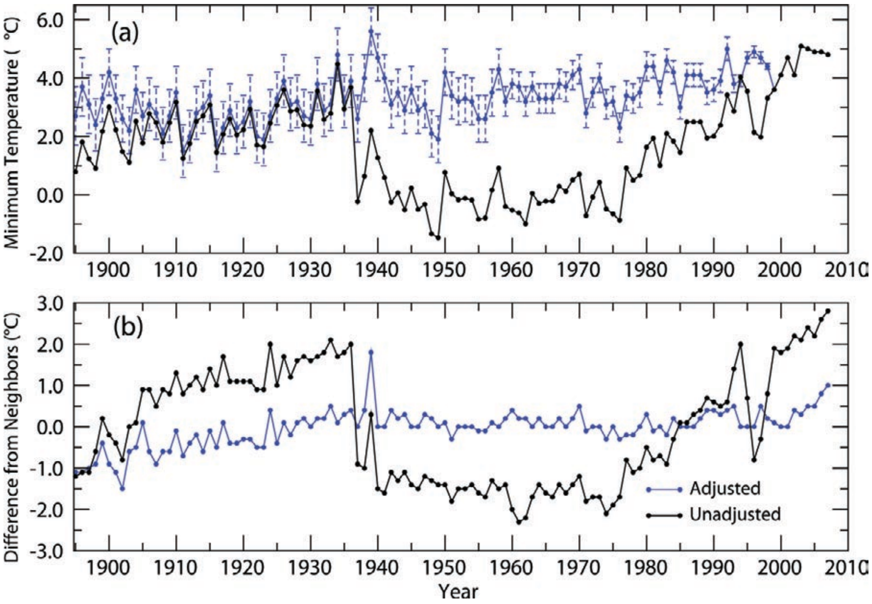

Fig. 8. (a) Mean annual unadjusted and fully adjusted minimum temperatures at Reno, Nevada. Error bars depict a measure of the cumulative uncertainty (95% confidence limits) in the pairwise algorithm's bias adjustments. The estimated uncertainty was determined using 100 Monte Carlo simulations in which a value within the range of pairwise estimates for the magnitude of each shift was randomly selected and used to adjust the series accordingly. (b) Difference between minimum temperatures at Reno and the mean from its 10 nearest neighbors.

trend beginning in the 1970s, possibly as a result of a growing urban heat island influence. In contrast, the fully adjusted HCN 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 HCN version 2 adjustments.) For these reasons, the average CONUS minimum temperature trend calculated from the 30% most urban HCN stations (based on population metadata) are about the same as that calculated from the remaining more rural locations (i.e., 0.071° and 0.077°C decade^{-1}, respectively) during the period 1895–2007.

It is important to note, however, that although 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 difficult to robustly identify the true interval of a trend inhomogeneity (Menne and Williams 2009).

Use of a simple difference in means test does, however, address 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 (e.g., Fig. 10 in Menne and Williams 2009). The reason is that, in some cases, gradual and sudden changes may not reflect station moves and the effect 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 such as the Reno observations, preserving the local trend (i.e., not adjusting for the gradual change) would result in a "double counting" of the UHI signal, because the station likely experienced urbanization effects when it was located in the city and then again after its relocation in the mid-1930s to an airport site (whose surroundings became urbanized much later).

One implication of using a difference in means test to adjust for all changepoints is that local trends are "aliased" onto the estimates of step changes (DeGaetano 2006). To quantify the influence of this aliasing effect, the pairwise approach was modified such that only abrupt shifts were removed, thereby creating a "nonproduction" version of HCN in which local trends were retained (see Menne and Williams 2009 for details). 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 Fig. 6). The reason for these distributional changes is that there is an apparent and sizable preference for relative trends between HCN stations and their neighbors to be negative. 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, despite cases like Reno. Thus, whereas there are apparent residual trend inhomogeneities that remain in some HCN series, they are more likely to be negative than positive and, collectively, there appears to be little evidence of a positive bias in HCN trends caused by the UHI or other local changes. It should be noted, however, that if there is a regional signal that affects a number of stations, its effect will be largely preserved by the homogenization procedure.

A number of recent articles have also raised concerns about the site characteristics of U.S. HCN stations by way of photographic documentation (e.g., Davey ahnd Pielke 2005; Pielke et al. 2007a,b). Moreover, there is evidence that a large fraction of HCN sites have poor ratings with respect to the site classification criteria used by the U.S. Climate Reference Network (A. Watts 2008 personal communication; refer also to www.surfacestations.org[1]). In at least one study (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 U.S. HCN sites. However, 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, which the TOB and pairwise adjustments effectively address (Vose et al. 2003; Menne and Williams 2009).

The goal of the HCN version 2 adjustments (and homogenization in general) is not to ensure that observations conform to an absolute standard but rather to remove the effect of relative bias changes that occur during a station's history of observation. In this regard, photographic documentation, though valuable, is most valuable when it is used to document the timing and causes of such shifts in bias through time. Ultimately, the magnitude of relative changes in the bias of observations, whatever the source, cannot be inferred from the metadata. Instead, the effect of station changes and nonstandard instrument exposure on temperature trends must be determined via a systematic evaluation of the observations themselves (Peterson 2006), generally through relative comparisons. Such an analysis suggests that the effect of undocumented changes appears to be at least as significant as documented changes in the HCN and that homogeneity testing for both types of shifts is critical.

Bias assessment of estimates for missing monthly temperature values. As in HCN version 1, HCN version 2 provides estimates for missing monthly maximum and minimum temperatures. Estimates are generated using an optimal interpolation technique known informally as FILNET (short for "fill in the network"), which 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 FILNET estimates and the observed

---

[1] Site classifications are based on a modification of Leroy (1999), as described in the U.S. Climate Reference Network (2002) Site Information Handbook.

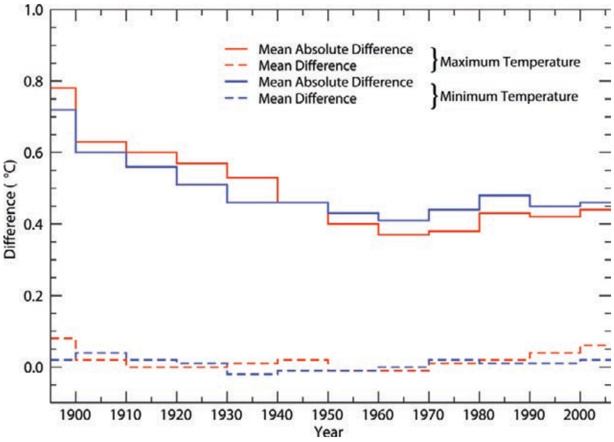

Fig. 9. Difference (by decade) between FILNET estimates and observed monthly values at all U.S. HCN stations.

monthly values in the HCN is 0.01°C, while the mean absolute difference is slightly less than 0.5°C. As shown in Fig. 10, the FILNET procedure has virtually no systematic effect on HCN temperature trends.

COMPARISON OF U.S. HCN VERSIONS 1 AND 2 MONTHLY TEMPERATURES. To assess the basic temperature differences between HCN versions 1 and 2 at the national scale, the annual CONUS averages from the two datasets were compared using the same gridding procedure described in the "Sources and assessment of temperature bias in the U.S. HCN" section. Because the HCN version 1 release provides an optional UHI correction, two difference series were formed for each variable: (i) HCN version 2 minus HCN version 1 (with TOB and SHAP adjustments), and (ii) HCN version 2 minus HCN version 1 (with TOB, SHAP, and UHI adjustments).

Figure 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 HCN version 1. The existence of this trend can be traced to the effect the SHAP adjustments had on minimum temperatures in HCN version 1. Specifically, the SHAP adjustments are limited to documented changes that have a preference for downward shifts (Fig. 6). 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 HCN version 1 UHI adjustment depresses 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 HCN 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 HCN version 2 pairwise adjustments do not increase the minimum temperature trend relative to the TOB-adjusted data (Fig. 7).

Figure 11 also suggests a divergence between HCN versions 1 and 2 temperatures after 1985, a difference associated with the adjustments for the MMTS instrument change in HCN version 1. As discussed in the "Bias associated with other changes in observation practice" section, the HCN version 1 MMTS correction appears to be too large when the effect on the full subset of HCN sites is considered (i.e., when stations with documented moves 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: first, the HCN version 1 metadata were last updated with the Easterling et al. (1996) release; second, the continued conversion to MMTS (and later Nimbus)—as well as the introduction of ASOS—have artificially (but unknown to SHAP) cooled maximum temperatures to a level that currently compensates for the HCN version 1 overcorrection.

TEMPERATURE TRENDS FROM 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 1895–2007. In general, all variables exhibit a slight increase

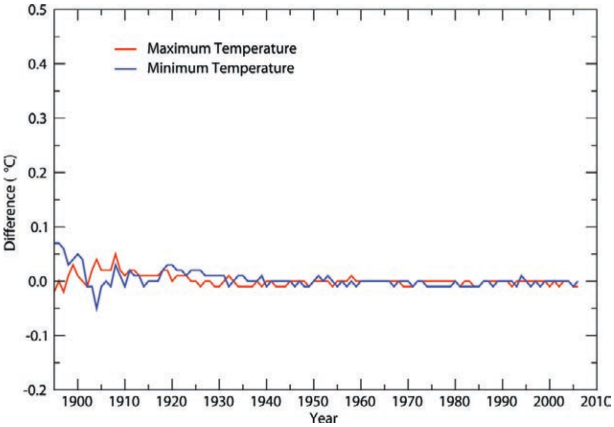

Fig. 10. Average annual differences over the CONUS between the fully adjusted HCN data with estimates for missing values (TOB + pairwise + FILNET) and the fully adjusted data without missing data estimates (TOB + pairwise).

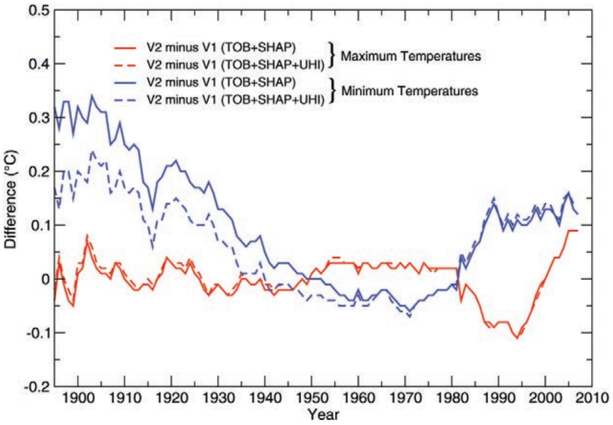

Fig. 11. Average annual differences over the CONUS between HCN version 2 and HCN version 1 (Revision 3; Easterling et al. 1996)

until the early 1930s, followed by a slight decrease until the early 1970s, and finally a more prominent increase into the early twenty-first 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 1934 and 2006; for minimum temperature, the two highest values occurred in 1998 and 2006.

Table 3 summarizes U.S. annual and seasonal (linear) trends in maximum and minimum temperature for the raw, TOB, and fully adjusted (TOB + pairwise) HCN version 2 data as well as the fully adjusted HCN version 1 data (TOB + SHAP + UHI). On an annual basis, the HCN version 2 trend in maximum temperature is 0.064°C decade^{-1}, and the trend in minimum temperature is 0.075°C decade^{-1} (both of which are comparable to the global mean trend of ~0.060°C decade^{-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 the "Bias caused by changes to the time of observation" section and "Bias associated with other changes in observation practice" section, trends in the adjusted data always exceed those in the raw data. However, as discussed in last section, the HCN version 2 trends in minimum temperature are somewhat smaller than the fully adjusted HCN version 1 trends.

In Fig. 13, the geographic distribution of linear trends in maximum and minimum temperatures for the period 1895–2007 are shown both for the adjusted HCN version 2 data and for the raw data. Geographically, maximum temperature (Fig. 13a) has increased in most areas except in 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 is a function of removing the time of observation bias and of the adjustments associated with the MMTS instrument change.

Despite the more coherent pattern, Pielke et al. (2007a,b) argue that homogenized data are not useful for calculating regional trends because the homogenized series lack independence, noting, in

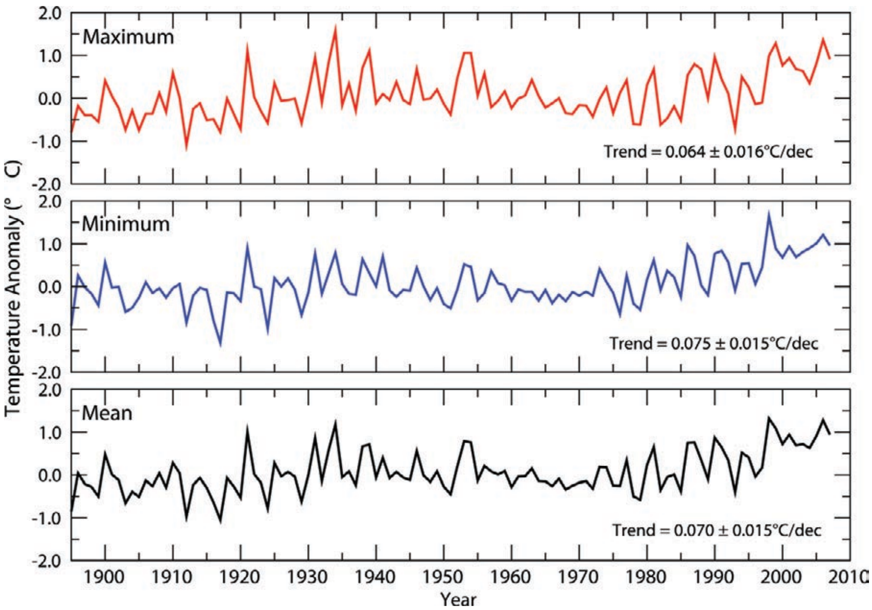

Fig. 12. Time series of annual temperature anomalies from HCN version 2 averaged over the CONUS. Base period is 1961–90. The trends include 95% confidence limits (± one standard error) that were calculated by adding the error in the least squares regression coefficient for the series trend and a factor quantifying the uncertainty in the adjusted temperature values (as described in Fig. 8).

particular, that the site-specific information that would have been obtained from a well-sited, stable station cannot be derived retrospectively. Nonetheless, Pielke et al. (2007b) state that the adjusted temperature series "may well be the best objective hypothetical climate record available." We believe that it follows that the adjusted series can be used to infer patterns of climate variability and change at the surface (which is one of the principal motivations behind climate data homogenization). Moreover, the increase in interstation correlation in the adjusted data relative to the unadjusted data is negli-

Table 3. U.S. annual and seasonal temperature trends (°C decade^{-1}) 1895–2007 for adjusted and unadjusted temperature series.

Season	Maximum temperature	Minimum temperature
	Fully adjusted—Version 2 (TOB + Pairwise)	
Annual	0.064	0.075
Dec–Feb	0.101	0.107
Mar–May	0.082	0.066
Jun–Aug	0.044	0.067
Sep–Nov	0.025	0.054
	Unadjusted (Raw)—Version 2	
Annual	0.018	0.054
	Adjusted for TOB only—Version 2	
Annual	0.033	0.076
	Fully adjusted—Version 1 (TOB + SHAP + UHI)	
Annual	0.063	0.090

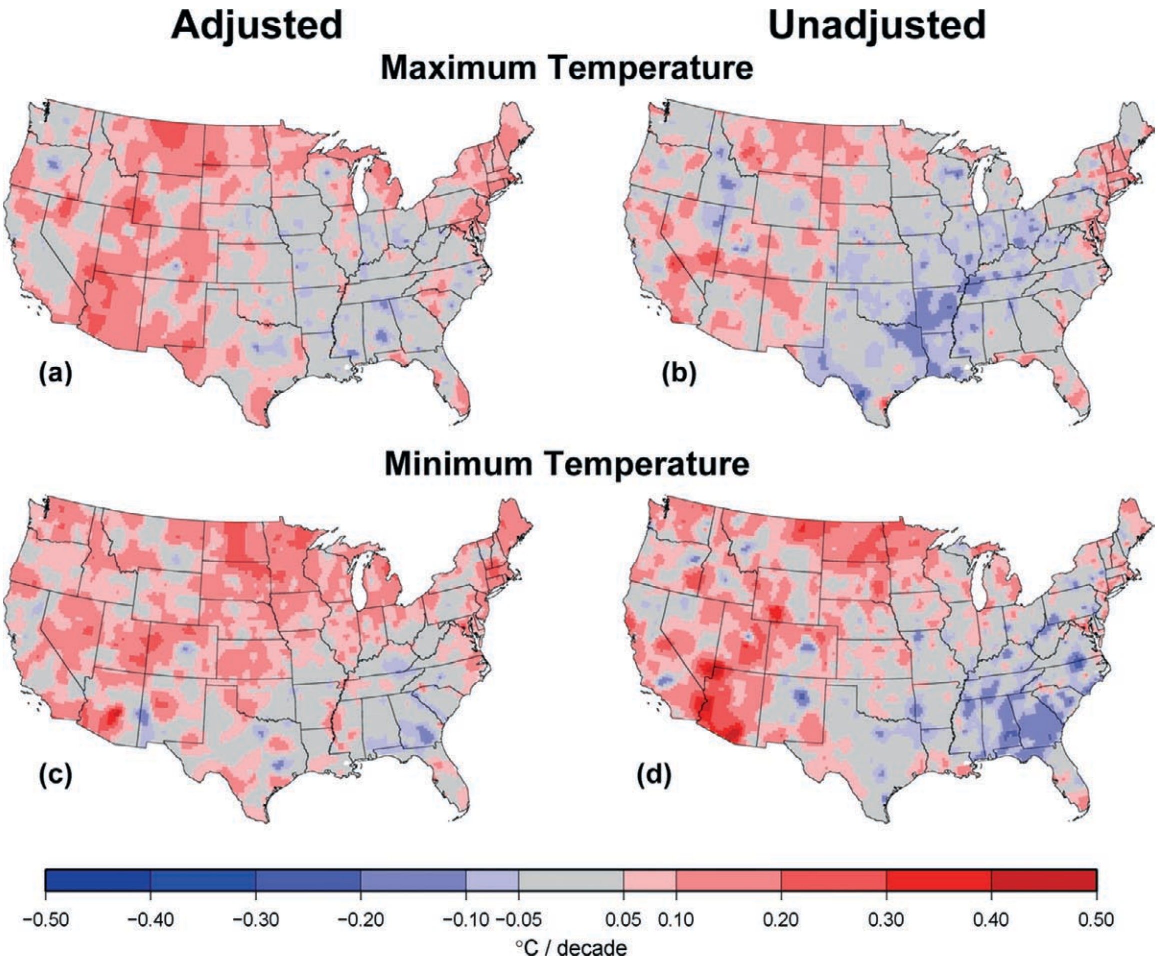

Fig. 13. Geographic distribution of linear trends in HCN version 2 temperatures for the period 1895–2007. (a) adjusted maximum temperatures; (b) unadjusted maximum temperatures; (c) adjusted minimum temperatures; (d) unadjusted minimum temperatures.

gible (accounting for the effect of shifts). 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. In addition, 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 temperature observations from the HCN should be sufficient to calculate regional trends in most areas. In any case, all COOP temperature series are homogenized by the HCN version 2 pairwise algorithm, which expands the pool of adjusted series beyond the HCN subset. Consequently, if there is a concern about the characteristics of a particular HCN site or inadequate station density in some areas, adjusted COOP temperature series can supplement the HCN. This is only one of the benefits of this unique climate network, made possible by the efforts of dedicated volunteers for more than a century.

SUMMARY AND CONCLUSIONS. Overall, the collective effect 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 −0.04°C decade^{-1} compared to the background trend of about 0.06°C decade^{-1}). Consequently, bias adjustments are essential in reducing the uncertainty in U.S. climate trends. The bias changes that have had the biggest effect on the climate network as a whole include changes to the time of observation (which affects both maximum and minimum temperature trends) and the widespread conversion to the MMTS (which affects primarily maximum temperatures). Adjustments for undocumented changes are especially important in removing bias in minimum temperature records. Tests for undocumented shifts, however, are inherently less sensitive than in cases where the timing of changes is known through metadata. Thus, metadata are exceedingly valuable when it comes to adjusting and evaluating climate trends.

Trends in the HCN version 2 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 although homogenization generally ensures that climate *trends* can be more confidently intercompared between sites, the effect 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. Moreover, while changes in observation practice have clearly had a systematic effect on average U.S. temperature trends, homogeneity matters most at the station level where even one change in bias can have a drastic effect on the series trend (which can occasionally be missed by changepoint tests). Therefore, the goal behind the HCN version 2 dataset (and future improvements) is to make the adjustments as site specific and comprehensive as possible, which is especially valuable in the development of widely used products, such as the U.S. Climate Normals.

Finally, the U.S. HCN data will be updated monthly and fully reprocessed periodically to detect and adjust for shifts from the recent past (see www.ncdc.noaa.gov/oa/climate/research/uschcn/ for further information, including access to the data and uncertainty calculations). Plans are also in place to ensure that U.S. HCN monthly means are internally consistent with NCDC's global daily dataset (the Global Historical Climatology Network—Daily dataset). Still, there is always room for improvement in the field of climate data homogenization. For example, although the monthly adjustments used in HCN version 2 are constant for all months, there is evidence that bias changes often have effects that vary seasonally and/or synoptically (Trewin and Trivitt 1996; Guttman and Baker 1996). As shown by Della-Marta and Wanner (2006), it is possible to estimate the differential effects 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.

ACKNOWLEDGMENTS. The authors wish to thank Anthony Watts for his considerable efforts in documenting the current site characteristics of U.S. HCN stations. The authors also thank Tom Peterson, Tami Houston, and three anonymous reviewers whose helpful comments greatly improved this manuscript. Partial support for this work was provided by the Office of Biological and Environmental Research, U.S. Department of Energy (Grant DE-AI02-96ER62276).

REFERENCES

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

——, D. L. Ruschy, and R. H. Skaggs, 1993: Agriculture and the recent benign climate. *Bull. Amer. Meteor. Soc.,* **74,** 1035–1040.

CDMP, 2001: Annual report. Climate Database Modernization Program Rep., 8 pp. [Available online at www.ncdc.noaa.gov/oa/climate/cdmp/files/annual-report2001.pdf.]

Climate Reference Network, 2002: Site information handbook. NOAA/NESDIS CRN Series X030, CRN Rep. NOAA-CRN/OSD-2002-0002R0UD0. [Available online at ftp://ftp.ncdc.noaa.gov/pub/data/uscrn/documentation/program/X030FullDocumentD0.pdf.]

Collins, D. A., S. Johnson, N. Plummer, A. K. Brewster, and Y. Kuleshov, 1999: Re-visiting Tasmania's reference climate stations with a semi-objective network selection scheme. *Aust. Meteor. Mag.,* **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., 2000: A serially complete simulated observation time metadata file for U.S. daily Historical Climatology Network stations. *Bull. Amer. Meteor. Soc.,* **81,** 49–67.

——, 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, cited 1996: United States Historical Climatology Network (U.S. HCN) monthly temperature and precipitation data. ORNL/CDIAC-87, NDP-019/R3. [Available online at http://cdiac.ornl.gov/epubs/ndp019/ndp019.html.]

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,** 23 947–23 963.

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.

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.

——, 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.

——, H. F. Diaz, and G. Kukla, 1988: Urbanization: Its detection and effect in the United States climate record. *J. Climate,* **1,** 1099–1123.

Leroy, M., 1999: Classification d'un site. Météo-France, Direction des Systèmes d'Observation. Tech. Note 35, 12 pp.

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.

——, and ——, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate,* **22,** 1700–1717.

Oke, T. R., 1987: *Boundary Layer Climates.* 2nd ed. Routledge, 435 pp.

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

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

Pielke, R. A., Sr., and Coauthors, 2007a: Documentation of uncertainties and biases associated with surface temperature measurement sites for climate change assessment. *Bull. Amer. Meteor. Soc.,* **88,** 913–928.

——, and Coauthors, 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, TN.

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

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

——, 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,** 2046, doi:10.1029/2003GL018111.

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. *Amer. Cartogr.,* **12,** 5–16.