

## A Homogenized Historical Temperature Extreme Dataset for the United States

ARTHUR T. DEGAETANO AND ROBERT J. ALLEN

*Northeast Regional Climate Center, Cornell University, Ithaca, New York*

KEVIN P. GALLO

*Office of Research and Applications, NOAA/NESDIS, Sioux Falls, South Dakota*

(Manuscript received 11 December 2001, in final form 29 March 2002)

### ABSTRACT

A subset of stations from the daily U.S. Historical Climatology Network (HCN) is used as a basis for a historical database of temperature extreme occurrence in the United States. The dataset focuses on daily temperature occurrences that exceed (fall below) the 90th (10th) percentiles of daily maximum and minimum temperature. Using a variety of techniques, the temperature extreme occurrence data are homogenized to account for nonclimatic shifts resulting from station relocations, changes in instrument type, and variations in the time of observations. Given the daily resolution of the extreme data, these potential sources of inhomogeneity require testing and adjustment using methods other than those conventionally used with mean temperature data. A data estimation technique, specific to extremes, is also used to produce serially complete exceedence records. Stations are also identified based on their current degree of urbanization using satellite observations. The dataset is intended to provide a research-quality source of temperature extreme data, analogous and complementary to the daily HCN dataset.

Two analyses are presented that illustrate the influence of adjustment. The change in temperature extreme occurrence with time reverses at between 15% and 20% of the HCN stations depending upon whether adjusted or unadjusted series is used. Changes in the distribution of extreme occurrences during drought and nondrought years are also shown to occur.

### 1. Introduction

Interest in the occurrence of temperature extremes has increased due, in part, to concerns of CO<sub>2</sub>-induced climate change. Based on statistical estimates, Mearns et al. (1984) expect a tripling in the likelihood of heat wave occurrences (5 consecutive days with maximum temperatures greater than 35°C) given a 1.7°C increase in the mean temperature of Des Moines, Iowa. Recent model estimates also reflect large changes in extreme events. Using an equilibrium solution for doubled CO<sub>2</sub>, Zwiers and Kharin (1998) show an increase of as much as 10°C in the 20-yr return period value of maximum temperature. The 20-yr recurrence interval minimum temperature increases by as much as 20°C across the central and eastern United States in these simulations. Such projections are critical to assessing the consequences of climate variations, since temperature extremes rather than averages are likely to produce the greatest societal impacts (Karl and Easterling 1999).

Despite such projections and impacts, relatively little

work has examined observed changes in extreme temperature events. Easterling et al. (2000) give a brief overview of the literature related to temporal variations in temperature extremes. In most cases, these studies have focussed on 0°C as an extreme temperature threshold.

DeGaetano (1996) examined trends in daily temperatures (both maximum and minimum) exceeding the 90th (or 10th) percentile of the distribution of all daily values across the northeastern United States. He found significant trends toward fewer cold minimum temperature threshold exceedences and more warm minimum temperature exceedences over the period from 1959 to 1993. A significant number of trends toward fewer warm maximum temperature threshold exceedences was also detected at the 22 stations that were analyzed.

David R. Easterling (2000, personal communication) examined U.S. extreme temperature trends based on the exceedence of fixed (0° and 32.2°C) thresholds and station-dependent percentiles. He shows a slight downward trend in the number of warm maximum temperature extremes, but notes large positive anomalies during the 1930s and 1950s drought years. Kalkstein and Davis (1989) also highlight the number of extreme temperatures that occurred in the 1930s. However, using a

---

*Corresponding author address:* Dr. Art DeGaetano, Northeast Regional Climate Center, Cornell University, 1119 Bradfield Hall, Ithaca, NY 14853.  
E-mail: atd2@cornell.edu

threshold equal to the 98.5th percentile, they find no exceedence trend over the 1931–96 period. Similarly, Kunkel et al. (1999) found no overall trend in 4-day heat waves with temperatures exceeding the 10-yr recurrence interval value.

Gaffen and Ross (1999) examined trends in extreme apparent temperatures and found increases in the number of days exceeding the 85th percentile over the period from 1948 to 1995. Although this agrees with the findings of Balling and Idso (1990), intuitively these results are in conflict with D. Easterling (2000, personal communication) and DeGaetano (1996). These differences can be attributed to the exclusion of the 1930s in the earlier studies, a potential urban influence at the stations used by Gaffen and Ross and water vapor trends (Ross and Elliot 1996).

It is likely that the absence of a high quality, long-term homogeneous set of daily maximum and minimum temperature data has led to this relatively small collection of studies and contributed to these contrasting results. Unlike monthly mean temperature data sets such as the U.S. Historical Climatology Network (HCN; Karl et al. 1990) and global HCN (Peterson and Vose 1997), a set of serially complete, homogenized daily temperatures has yet to be assembled. Although the daily HCN dataset (Easterling et al. 1999) provides a foundation for such a benchmark dataset, difficulties in adjusting daily series for changes in observation time, station relocations, and instrument changes have precluded the development of a homogenized daily dataset analogous to the monthly HCN. Recently, however, a method to homogenize daily temperature extreme series has been developed (Allen and DeGaetano 2000). This, in addition to the development of a temperature estimation procedure specific to extreme occurrences (Allen and DeGaetano 2001), has allowed us to create a long-term set of homogenized maximum and minimum temperature extremes (both cold and warm) for a subset of daily HCN stations. To facilitate the use of these data, not only in trend analysis, but potentially with regard to describing or forecasting interannual variations in temperature extreme occurrence, this paper documents the data and methods used to develop this set of daily temperature extremes, which is referred to as the Daily Historical Climatology Network for Extreme Temperature (HCN-XT). Examples of the influence of data homogenization are also presented.

## 2. Dataset development

### a. Station selection

The HCN-XT dataset is composed of a subset of 361 stations (Fig. 1) selected from the 1096-station Daily Historical Climatology Network (Easterling et al. 1999). The initial selection of stations was based on the completeness of the data record. Stations at which >10% of the daily observations were missing received no fur-

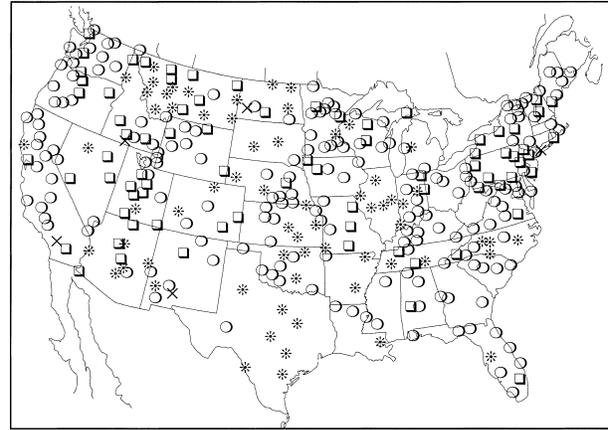


FIG. 1. Distribution of HCN-XT stations with beginning year of record indicated by  $\times$  ( $\geq 1950$ ),  $\circ$  (1930–49),  $\square$  (1910–29), or  $*$  (pre-1910).

ther consideration. Stations with nonstandard thermometry were also excluded, except in the case when this instrument was used only in the earliest portion of the record. In this case, the analyzed record was limited to that with either liquid-in-glass thermometers, a maximum–minimum temperature sensor (MMTS), or any of the hygrothermographs in use at first-order weather stations.

The records of the retained stations were divided into subseries, with each segment reflecting a change in location or thermometer type based on the HCN station history file. Based on this metadata, a relocation was assumed when the distance from the previous site was listed as  $\geq 0.1$  miles; the move was associated with a documented change in latitude, longitude, or elevation; or the instrumentation height changed. Although minimization of the number of station relocations was not an explicit aim, to facilitate subsequent homogenization, each subseries was required to be at least 10 yr in length.

Under several circumstances, stations not meeting the 10-yr constant location requirement were also selected. Less than 10-yr segments at the beginning of the station records were omitted and the remaining records retained. Stations at which the most recent subseries exceeded 49 yr and stations having the majority (>75%) of subseries with lengths  $\geq 20$  yr were also included. Finally, 24 stations were chosen from data-sparse areas to subjectively balance the spatial distribution of stations. Whenever possible these stations were selected to maximize record and subseries length while minimizing the number of subseries.

Of the 361 stations, the majority have records that begin between 1930 and 1950. At 76 stations, records begin prior to 1910. Another 87 stations have observations that start from 1910 to 1930. Five additional stations begin observations after 1950. The spatial distribution of these sites is shown in Fig. 1.

*b. Estimation of missing data*

Although data estimation techniques such as those described by Eischeid et al. (2000) and DeGaetano et al. (1995) provide rigorous methods for creating serially complete temperature series, Allen and DeGaetano (2001) show these methods to be biased when only the most extreme temperature observations are considered. Therefore, missing temperature data were estimated using a procedure developed by Allen and DeGaetano (2001). The method is a variation of a least squares regression approach that focuses on obtaining accurate estimates of annual exceedence counts (e.g., the number of days exceeding the 90th percentile,  $T_{90}$ ) and counts of consecutive extreme exceedences, while minimizing the estimation error associated with each individual extreme temperature observation.

Data estimation involves the selection of a set of the 15 closest neighboring predictor stations, each free of documented station moves during the period used to develop the estimation equations and with the same observation time category (Karl et al. 1986) as the data being estimated. Although a maximum separation distance of 800 km was imposed, most station pairs were separated by considerably less distance. This requirement of constant station location and observation time precluded the use of non-HCN cooperative network stations in the estimation procedure as a complete record of observation time is not available for cooperative network stations.

The 15 potential predictor stations were also required to share at least 100 extreme temperature days with the predictand station. Such days experienced nonmissing temperatures exceeding  $T_{90} - 1.1^{\circ}\text{C}$  at the predictand station, with the offset required to prevent the underprediction of temperatures near  $T_{90}$ . Furthermore, a temperature pair containing at least one observation for which the predictand's temperature is greater than its 97th percentile must be available at each predictor station.

Using binned least squares regression, one-predictor regression equations were developed using data from each of the 15 selected stations and the corresponding extreme temperature days from the missing data station. Binning prevents the unequal weighting (in terms of the number of observations) of points near the 90th percentile relative to those that are more extreme. The dependent sample of extreme temperature days was grouped such that each bin was associated with a unique temperature at the predictand (missing temperature) station. This set of unique predictand station temperatures serve as  $y$  values in the regression, with  $x$  values corresponding to the medians of all predictor station temperatures within each bin. Cases for which a negative regression slope resulted were not considered.

The 15 binned regressions were then evaluated using cross validation and the equations associated with mean errors  $\geq 0.28^{\circ}$  or  $\leq -0.28^{\circ}\text{C}$  omitted. When rounded,

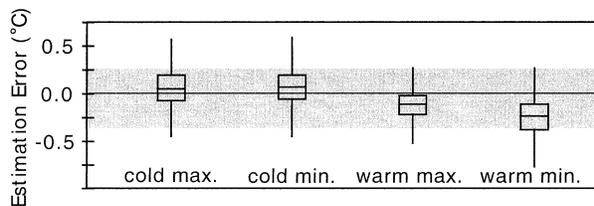


FIG. 2. Boxplots of average estimation errors, over all station subseries. Errors reflect temperatures within the range  $T_{92} - T_{97}$  ( $T_8 - T_3$  for cold max and min temperatures). The gray band indicates errors that, when rounded, equal zero.

larger errors gave errors  $\geq 1^{\circ}\text{F}$ . Although the remaining equations minimize the mean absolute error, they are still not optimal for categorizing a missing day as exceeding or not exceeding an extreme temperature threshold. In most cases, the number of days exceeding the 99th percentile is overestimated since the relatively large number of points near the 90th percentile still dominate the regression.

To illustrate this bias in exceedence counts, consider a set of warm temperature extremes and a binned regression equation that, on average, estimates temperatures in the range of  $33.9^{\circ}$ – $36.1^{\circ}\text{C}$  as  $35^{\circ}\text{C}$ . If  $35^{\circ}\text{C}$  is the 95th percentile, there are more temperatures (days) between  $33.9^{\circ}$  and  $35^{\circ}\text{C}$  than there are between  $35^{\circ}$  and  $36.1^{\circ}\text{C}$ . Thus, the chance of underestimating the extreme is greater. To compensate for this problem, the regressions were optimized by iteratively using percentiles other than median in the most extreme bins. Optimization continues until the number of estimated days exceeding the 99th and 95th percentiles is within 10% of the number of observed days, using cross validation.

After optimizing the 15 binned regression equations separately, the two that were previously associated with the lowest mean absolute error are combined. For each data pair, the combined estimate is simply the median of the estimates based on the individual equations. The combination of equations continues (in order of increasing mean absolute error) until the value  $100 - [(R_{90} + R_{95} + R_{99})/3]$  is minimized. Here,  $R_j$  is the ratio ( $\times 100$ ) of estimated to actual days exceeding the  $j$ th percentile based on cross validation.

Clearly, the final optimized regression is no longer the best (in terms of minimizing squared errors) for all temperatures  $\geq T_{90} - 1.1^{\circ}\text{C}$ , since the slopes and intercepts of the original binned regressions have been changed. The effect is minimal at temperatures near  $T_{90}$  but introduces a slight negative bias overall, since the more extreme temperatures are intentionally being underestimated. Nonetheless, the bias is similar (and generally less) than that associated with the more traditional data estimation techniques. This leads to individual estimates of sufficient accuracy and more importantly unbiased estimates of threshold exceedence days. Figure 2 shows the distribution of average cross-validation errors (over each homogeneous subseries) associated with the optimized regressions. Another desirable character-

istic of an extreme temperature estimation routine is the ability to replicate the occurrence of consecutive exceedences. Based on cross-validation, over 91% of the  $\geq 2$ ,  $\geq 3$ , and  $\geq 4$  day extreme temperature runs were simulated by the-estimation procedure.

In some cases, it was not possible to estimate missing temperature data. Generally, this occurred when either a set of predictor stations within 800 km of the missing data station could not be identified, or when there was an inadequate number predictor–predictand extreme temperature pairs to develop a regression. This latter condition occurred when there were fewer than 50 common extreme days or when a common extreme exceeding the 95th percentile could not be identified. If a regression could not be developed because of a deficient sample size, the missing data values were flagged according to their potential for exceeding the extreme threshold. This flag was based on the temperatures observed at the predictor stations. In cases where a temperature observation was available at one or more of these neighboring stations, the missing data value was flagged as unlikely to exceed the extreme threshold if the neighboring station's temperature was more than 5.5°C (10°F) below (or above for cold extremes) its corresponding percentile.

### *c. Threshold section*

Intuitively, the most recent period provides a logical base for computing station-specific extreme thresholds. However, the longest homogeneous data subseries also presents a viable alternative and was selected as a starting point for the temperature extreme database. Using the longest possible data record ensured a representative sample of daily temperature occurrences, particularly within the tails of the empirical distribution. It also avoided complications associated with documented data discontinuities within the base period and minimized the amount of adjusted data.

A primary consideration for choosing the recent record is to avoid adjusting subsequent (i.e., future) observations, at least until another discontinuity is introduced. This is of less concern in the analysis of the extreme series. Parameters such as the mean are operationally computed and archived, and therefore unprocessed values can be simply added to the series. This would not be the case if the addition of an adjustment factor was necessary. For temperature extreme data, adjustment involves altering the extreme threshold rather than a translation of the series mean (section 2e). Therefore, regardless of the current value of the threshold, extension of the extreme count series requires new tallies to be computed from the daily observations.

Although the choice of the base period affects the number of extreme temperature exceedences that a specific month or year experiences, in most cases it has no effect on the slope of the data series. An exception occurs if the base period corresponds to the warmest (or

coldest) portion of a nonstationary record. Since threshold exceedence count data are truncated at zero, a percentile threshold characteristic of the warmest (coldest) part of the record may result in no exceedences in the majority of the cooler (warmer) years. Clearly, this would skew any trend that may have been present. The potential for this problem was reduced by using the longest period, rather than the most recent, as the base for computing the extreme thresholds.

There are several alternative methods for computing percentile thresholds. Empirical percentiles, based on the sorted series of daily temperatures within the base period, offer one option. However, parametric percentiles based on some theoretical distribution (e.g., normal) could also be computed. As another option, separate percentiles can be computed for each year (or season) within the base period and the median (or average) of these annual values used as the relevant threshold. While the first and last approaches yield similar thresholds (Fig. 3), the parametric percentiles are consistently higher than their empirical counterparts. These differences are of little practical significance, provided a consistent computation method is adopted throughout the analysis. The first approach was adopted in this study, since it provided a more lenient definition of extreme allowing dataset users greater flexibility in choosing higher (lower) application-driven extreme thresholds.

For each station's base period, the 1st, 5th, 10th, 90th, 95th, and 99th percentile of daily maximum and minimum temperature was computed based on all complete years within the base period. Since the daily temperatures within the upper and lower deciles tend to be confined to specific seasons, the empirical distribution of temperatures could be limited to the relevant 3- or 6-month period. This would provide a much more stringent definition of extreme. Using all days, the 10th percentile approximates the 36th coldest day of a specific year. When confined to winter days, only the nine coldest days typically exceed the 10th percentile threshold. In creating the dataset, it was desirable to restrict the data as little as possible. Although some restriction was required for homogenization, the use of a lax definition of annual percentile extremes, allowed the creation of a dataset that includes the observations necessary to conduct analysis based on more stringent (such as seasonal) definitions of extreme. It should also be noted that the dataset is limited to the analysis of annual extreme exceedences. Investigations of extremes from a monthly perspective (e.g., extremely warm winter days) would likely require modification of the routines used for homogenization and data estimation.

### *d. Observation time adjustment*

The effect of observation time bias on monthly and annual temperature exceedence counts is analogous to its influence on average temperature. Figure 4 shows representative patterns of observation time bias for sea-

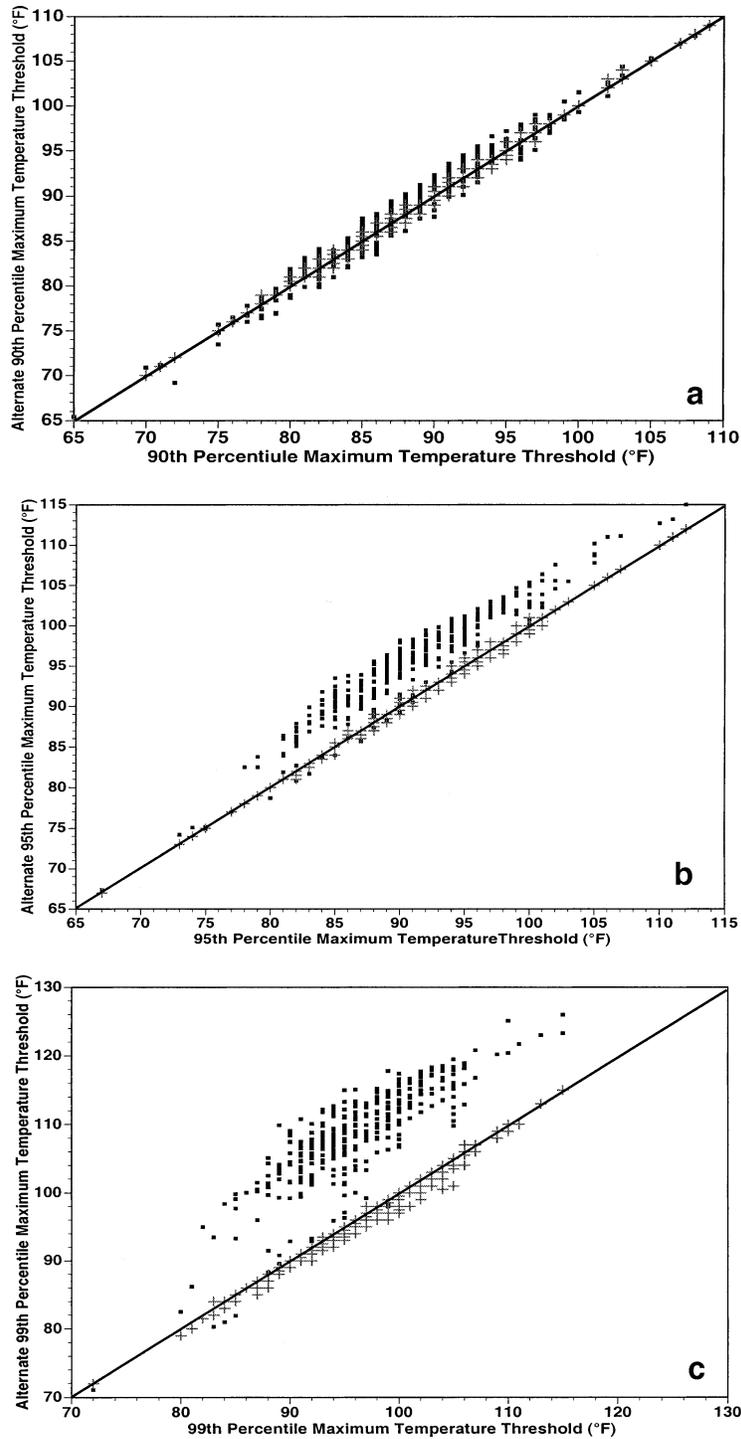


FIG. 3. Comparison of empirical (a) 90th, (b) 95th, and (c) 99th percentiles using all data with those based on a fit normal distribution (dots) or the median of a series of annual empirical percentiles (crosses).

sonal exceedence count data. Hourly data from 12 geographically diverse stations were used to simulate daily maximum and minimum temperatures based on 24 observations times. For each simulated observation sched-

ule, seasonal counts were made of exceedences of the 90th and 10th percentiles, with the percentiles being based on a midnight-to-midnight observation. Seasons were defined as June–August (JJA) and December–Feb-

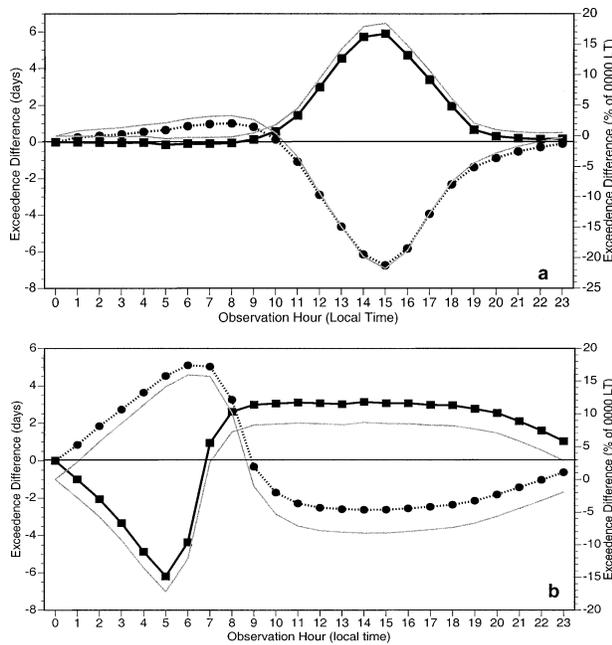


FIG. 4. Representative patterns of temperature extreme occurrence observation time bias relative to a midnight-to-midnight observation for (a) max and (b) min temperature. Squares correspond to the 90th percentile and circles to the 10th percentile. The gray curves show the observation time bias as a percentage of the midnight value.

bruary (DJF) for the 90th and 10th percentile thresholds, respectively.

For warm ( $\geq 90$ th percentile) maximum temperatures, there is almost a 20% (6 days on average) increase in exceedences when observations are based on an afternoon instead of a midnight (or morning) observation time (Fig. 4a). This is a reflection of the maximum temperature from a relatively warm afternoon being the highest reading in successive observation intervals. Cold ( $\leq 10$ th percentile) maximum temperatures display the opposite pattern of bias, with over 20% fewer exceedences associated with counts from afternoon observations. In this case, a particularly cold maximum temperature during one afternoon can be exceeded by a warmer reading near the end of the 24-h afternoon observation interval. The colder reading may not be superseded in the 0000 local time (LT) observation, since each set of afternoon hours is contained within separate midnight-to-midnight intervals. An analogous situation is responsible for the slight increase in cold maximum temperature exceedences based on morning observations, since, on average, the temperature at midnight is warmer than at an early morning hour.

As opposed to maximum temperatures, few observation hours are free of observation time bias when minimum temperature exceedences are considered. Seasonal warm minimum temperature exceedences decrease by almost 20% (about 6 days) for morning observation hours, but increase by 10% when based on afternoon observation schedules. As with cold maxi-

TABLE 1. Hourly stations used to develop and validate the observation time adjustment equations.

Development		Validation	
Station	Abbreviation	Station	Abbreviation
Brownsville, TX	BRO	Asheville, NC	AVL
Cape Hatteras, NC	HSE	Billings, MT	BIL
Caribou, ME	CAR	Corpus Christi, TX	CRP
Fort Myers, FL	FMY	Flagstaff, AZ	FLG
New Orleans, LA	MSY	Fresno, CA	FAT
Pocatello, ID	PIH	Los Angeles, CA	LAX
Quillayute, WA	UIL	Miami, FL	MIA
Sacramento, CA	SAC	Mobile, AL	MOB
San Diego, CA	SAN	Omaha, NE	OMA
St. Louis, MO	STL	Phoenix, AZ	PHX
Salt Lake City, UT	SLC	Portland, ME	PWM
Tucson AZ	TUS	San Francisco, CA	SFO

um temperatures, this pattern reflects the possibility that the minimum occurring near the end of the morning-to-morning interval (or midnight-to-midnight) will be cooler than an extremely warm minimum occurring near the beginning of the period. The pattern of bias for cold minimum temperature exceedences mirrors that of the warm minimum temperature exceedences, since extremely cold temperatures occurring during early morning (late night) hours often reflect the minimum temperature of two morning-to-morning (midnight-to-midnight) observation periods.

To compensate for these biases in exceedence count series that experience changes in observation time, regression-based adjustments were developed. As opposed to previous methods of compensating for observation time bias in average temperature series (e.g., Karl et al. 1986), it was not feasible to apply the same monthly adjustment to data from each year. Rather, since the adjustments rely heavily on the number of exceedences experienced and there is considerable year-to-year variability in monthly exceedences, it was necessary to compute separate adjustments for each individual month. Simulated daily observations representing each observation hour were computed from hourly data at the 12 stations listed in Table 1 for the period 1985–95 and are used as developmental data for obtaining regression-based observation time adjustments.

Initially, the simulated data were grouped into observation time categories (i.e., morning, afternoon, and midnight) based on Karl et al. (1986) and a set of six regression equations developed for each of 12 temperature extreme categories (i.e., 1st, 5th, 10th, 90th, 95th, and 99th percentile exceedences for maximum and minimum temperature). In each case, the original monthly exceedence count, the number of monthly runs, a measure of the observation time bias associated with average temperature, and the current observation hour were considered as potential predictors of the monthly exceedence count for an hour in a different observation time category. Predictors were eliminated from the final re-

gression equations using a backward elimination procedure based on  $\alpha = 0.10$ .

Monthly runs were defined as single or consecutive-day occurrences of a temperature exceeding the extreme threshold, separated by at least one day on which the temperature fell below the threshold. The measure of average temperature bias was computed based on the empirical model given by Karl et al. (1986). This model estimates the observation time bias associated with monthly maximum or minimum temperatures at a station as a function of some base bias, which is a function of local solar month and hour (i.e., latitude, longitude, and time zone), the mean monthly interdiurnal temperature difference (absolute value of day-to-day differences in mean temperature), and a measure of the end-of-month effect. For each individual month, the model was run using the average interdiurnal temperature difference within the specific month, as opposed to the static monthly means used by Karl et al. (1986). From the output, the difference between the bias associated with the new observation hour and that of the original hour was computed for use as a potential predictor.

Examination of the residual plots (not shown) associated with these regression equations indicated that the assumption of constant variance was not valid. As this feature could not be remedied through transformation of the predictands, the separate regression equations were reformulated based on specific predictand hours as opposed to categories. Thus, adjustment of a monthly temperature exceedence count based on any morning (0600–0900 LT) observation hour to that of a 1700 observation schedule required a different equation than conversion to a different (say 1800 LT) observation hour. Initially this led to a substantial increase in the number of regression equations required. However, basing the regressions on specific predictand hours allowed the separate equations for the 90th, 95th, and 99th (or 1st, 5th, and 10th) percentile exceedence counts to be combined. Thus, conversion between all but the most uncommon observation hours could be accomplished with four sets (warm and cold maximum and minimum temperature) of 20 equations.

Besides the formulation of these regression equations, it was also necessary to develop a criterion to determine if adjustment was required in months with a zero exceedence count. For instance, when based on an afternoon observation, it is possible for a month to have an exceedence of a warm minimum temperature threshold, even though no exceedences are reported based on a morning observation threshold. This stems from the possibility that the minimum occurring near the end of the morning-to-morning interval will be cooler than an extremely warm minimum occurring near the beginning of the period. Thus, while the single warm minimum associated with an afternoon observation interval may stand as the lowest temperature observed in that particular 24-h period, the second, cooler temperature is likely to represent the minimum of the morning-to-

morning period. Based on the patterns of observation time bias shown in Fig. 4, it is not possible to increase a zero monthly warm maximum or cold minimum temperature exceedence count through a change in observation time. Thus for these variables, the regression equations apply only to those months with nonzero exceedence counts. Conversely, a zero exceedence count may increase when going from a morning to afternoon observation time for warm minimum temperatures or from an afternoon to morning observation hour for cold maximum temperatures. For these variables, the regressions were fitted and applied to all months in which the highest (lowest) reported temperature was at most  $1.1^{\circ}\text{C}$  less (greater) than the extreme threshold. Although arbitrary, this  $1.1^{\circ}\text{C}$  ( $2^{\circ}\text{F}$ ) interval provided a means of identifying months in which it was unlikely that an observation time change would influence a zero exceedence count.

The regression equations were evaluated using an independent set of simulated daily observations representing the relevant observation hours. Hourly data from 12 sites (Table 1) for the period 1985–95 were used for these simulations. Figure 5 shows the results of these evaluations for the 90th (Figs. 5a,c) and 10th (Figs. 5b,d) percentiles using boxplots. The distributions of estimation errors are similar for both thresholds and maximum and minimum temperatures. In all cases, estimates are relatively unbiased, as the median errors are less than  $\pm 0.5$  days. When rounded, counts within this range are associated with no adjustment error. However, based on the validation data there is a slight tendency to overadjust morning observations (when converting to afternoon based values) and likewise underadjust afternoon observations (Figs. 5a,b,c). In all but a few cases, 75% of the monthly estimates are within  $\pm 1$  day of the observed values, and 95% of the estimates fall within 3 days of the observations. For comparison, Fig. 6 shows boxplots of the estimation errors associated with 95th and 99th percentile maximum temperature exceedences. In these cases the distribution of residuals is similar to that for the 90th percentile exceedences.

In Figs. 5 and 6, each error distribution represents 12 stations, 11 yr, 12 months and, in some cases, a range of original hours. To identify any biases related to these parameters, Fig. 7 shows the distribution of residuals for specific stations, years, months, and original hours. The conversion of 90th percentile minimum temperature exceedences from an afternoon hour to that of an 0700 LT observation time is used as a representative example. The choice of original observation hour and year have little effect on the distribution of adjustment errors. Despite the apparently large biases during winter, the distributions of adjustment errors are similar for each month. The boxplots in Figs. 5–7 omit cases in which adjustment was not required. In the case of warm minimum temperatures, this includes months in which the highest reported temperature was more than  $1.1^{\circ}\text{C}$  ( $2^{\circ}\text{F}$ ) below the extreme threshold. Thus, the monthly error

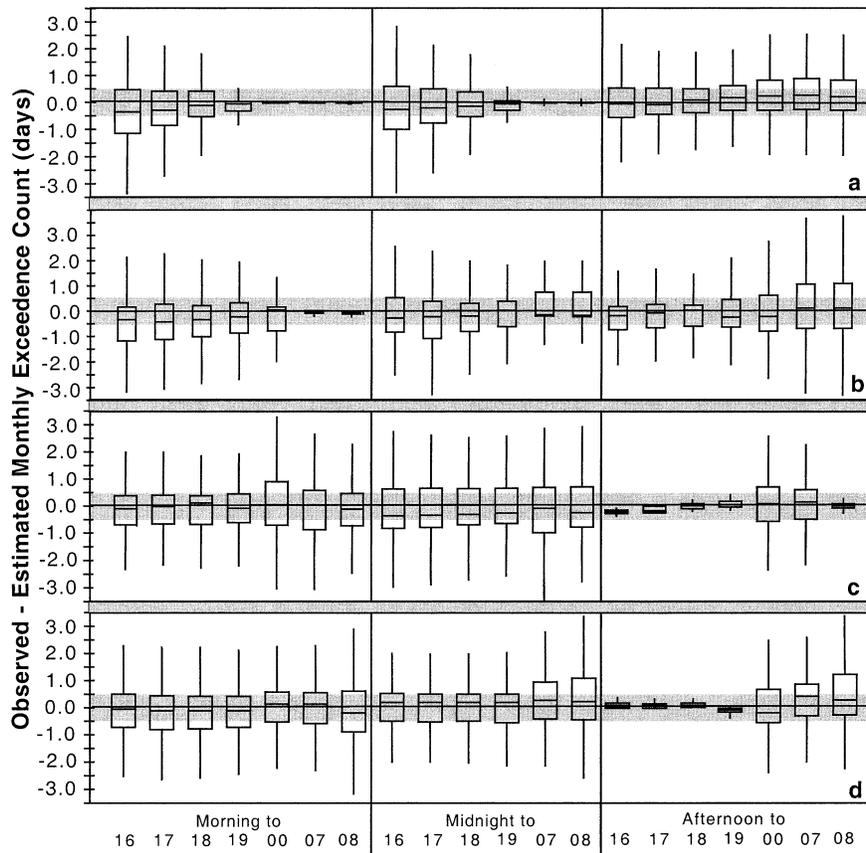


FIG. 5. Boxplots of observation time bias adjustment errors for 12 independent stations for (a) 90th percentile max temperature, (b) 10th percentile max temperature, (c) 90th percentile min temperature, and (d) 10th percentile min temperature. Whiskers denote the 95th percentile.

distributions for winter months are based on a very small sample (Fig. 7).

The distributions of adjustment errors are, however, influenced by station location. The regression equations overadjust warm minimum temperature exceedance counts at Phoenix, Asheville, and Omaha, while un-

deradjustment is most pronounced at Miami, Fresno, and Mobile. With the exception of Omaha and Asheville, these verification stations are generally located in regions where the influence of observation time bias is minimal (DeGaetano 2000) and therefore provide a stringent evaluation of the regressions. The error char-

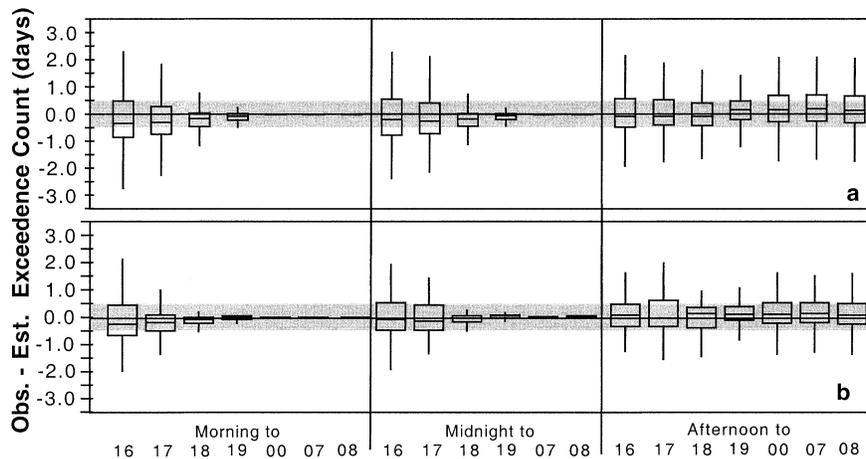


FIG. 6. As in Fig. 4, but for the (a) 95th and (b) 99th max temperature percentiles.

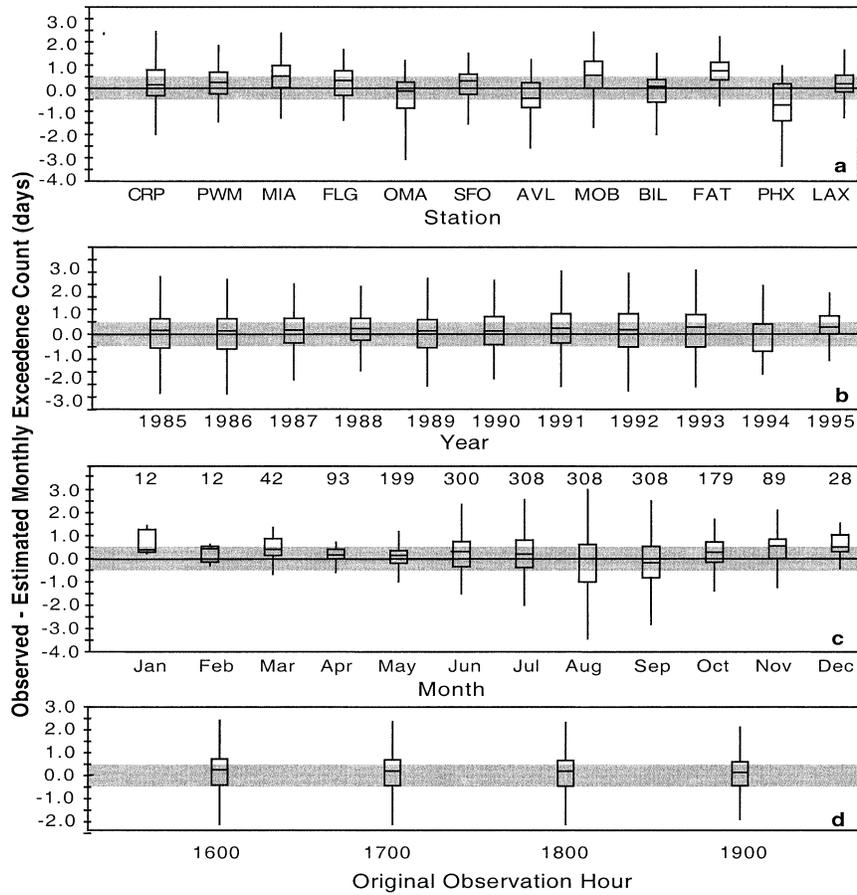


FIG. 7. As in Fig. 4, but for 90th percentile min temperature exceedences for specific (a) stations, (b) years, (c) months, and (d) hours. (d) Observations are adjusted to that of an 0700 observation hour. The numbers in (c) are the adjustments used in forming the boxplots.

acteristics of Portland, Billings, and Flagstaff are more characteristic of stations located away from the immediate Gulf and Pacific coasts where the effects of observation time are more pronounced (DeGaetano 1999).

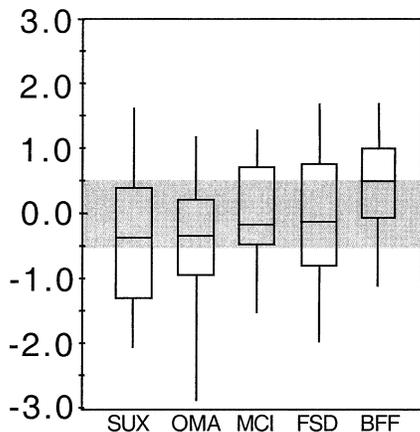


FIG. 8. As in Fig. 6, but for specific stations in the vicinity of Omaha, Nebraska (OMA).

The skewed exceedence count errors for Omaha and Asheville likely result from station-specific characteristics. DeGaetano (2000) showed suboptimal observation time classification success at Asheville, despite favorable results for surrounding stations. Likewise, Omaha lies within an area characterized by low interdiurnal minimum temperature range, relative to that for maximum temperature (DeGaetano 1999). Over the majority of the country the magnitude of these two ranges are essentially equal. To further investigate whether a regional bias was present in the vicinity of Omaha, adjustment errors at a set of four neighboring stations—Sioux Falls, South Dakota (FSD); Sioux City, Iowa (SUX); Kansas City, Missouri (MCI); Scottsbluff, Nebraska (BFF)—were analyzed (Fig. 8). Sioux City displays a comparable tendency toward overadjustment. However, the adjustments at the other stations are either unbiased (FSD) or exhibit a slight (MCI) or modest (BFF) underadjustment. Although such station-specific biases can be identified based on simulated observation times, they could not be inferred from daily HCN observations. Therefore, the development of a set of sta-

tion-specific (or even regional) observation time adjustment functions would not have been practical.

#### e. Inhomogeneity adjustments

Time series of temperature exceedence counts can also be influenced by changes in instrument type and station location. The methods of Allen and DeGaetano (2000) were used to test and potentially adjust each documented station relocation or instrument change for a nonclimatic discontinuity. Prior to this screening, the data series were standardized to a common observation time reflecting the predominant historic observation schedule. This facilitated the inhomogeneity tests by maximizing the number and length of homogeneous periods at each station.

Potential inhomogeneities were identified using the HCN station history file. Based on these metadata, it was possible to identify changes in station location, instrument type, and instrument height (e.g., roof top versus ground). A change in any one of these three attributes identified a potential discontinuity that required testing. Changes associated with site characteristics (e.g., nearby paving or construction) and routine weather station maintenance (e.g., shelter repainting or the replacement of a broken thermometer) are not documented electronically and therefore it was difficult to consider such changes as potential inhomogeneities.

For each potential discontinuity, a set of neighboring reference stations was assembled from the pool of 1096 daily HCN stations in a fashion similar to Karl and Williams (1987). However, due to differences between the testing and adjustment procedures, the selection of stations was based on minimizing the pooled standard deviation of the difference series rather than the confidence interval width associated with the  $t$  test. These differences arise from Karl and Williams' use of the Student's  $t$ -test as opposed to the adoption of a nonstandard test statistic by Allen and DeGaetano (2000).

Once a set of reference stations was selected, a combined reference series was formed by weighting corresponding values from each series by their respective pooled standard deviation and then summing each set of annual exceedences. This reference exceedence count series was used to compute a difference (reference – potentially inhomogeneous series), which was divided into two periods according to the documented inhomogeneity. The 75th and 25th percentiles of the longer of the two periods ( $T75_1$  and  $T25_1$ ) were then calculated. Using these values, the proportion of years in the shorter period that exceeded  $T75_1$  ( $P75_2$ ) was calculated, as was the proportion of years that fell below  $T25_1$  ( $P25_2$ ) and used to compute the test statistic

$$t_s = (0.25 - P75_2) - (0.25 - P25_2). \quad (1)$$

When the two periods are similar (i.e., the metadata change does not produce a significant discontinuity),

$P75_2 \approx 0.25$ . Similarly,  $P25_2 \approx 0.25$  and thus,  $t_s \approx 0$ . If, however, the discontinuity introduces a significant warming or cooling during the second period, then the quartiles of the two periods will be different with  $t_s < 0$  or  $t_s > 0$ , respectively. Thus, in this two-tailed test, the null hypothesis is defined as  $H_0:t_s = 0$ . Periods of less than 5 yr were not tested. Such periods were uncommon given the criteria for selecting stations.

Once  $t_s$  was computed, the statistical significance of the discontinuity was assessed by resampling techniques. Here, the combined series (i.e., the years before and after the discontinuity) were randomly sampled with replacement 1000 times. For each reordering, a new  $t_s$  value was calculated creating a distribution of  $t_s$  consistent with the null hypothesis of no difference before and after the discontinuity. Significant discontinuities were associated with  $t_s$  values within the lowest or highest 2.5% of the resampled distribution.

In its simplest form the testing procedure assumes that the difference series before and after the discontinuity are stationary. When one or both time periods have a significant slope, the basic test incorrectly rejects  $H_0$  too frequently. Prior to application of the test, each sub-series was tested for a significant slope (Wilks 1995). Based on a set of HCN stations included in the temperature extreme dataset, trended differences series (either before or after the discontinuity) occurred in approximately 17% of the cases. In these cases, a more complex testing procedure was implemented.

When only one of the difference series segments was nonstationary, it was detrended, based on the residuals obtained from a linear least squares fit of the original time-dependent series prior to the application of the test procedure. After fitting this regression, the 95% confidence interval for the slope and intercept were computed (Draper and Smith 1981). The lines representing the bounds of this interval were projected to the year of the discontinuity and the intersections used to translate the original residual series into two new series. This allowed the nonstationary series to be described by two stationary series at the upper and lower limits of the 95% confidence interval about the original regression.

The test now proceeded in a manner analogous to the stationary case. First, using the stationary series after the discontinuity, the proportion of years exceeding (falling below) the 75th (25th) percentile of the upper detrended residual series were computed and used to calculate  $t_{s \text{ upper}}$ . The statistical significance of  $t_{s \text{ upper}}$  was assessed based on 1000 bootstraps of the combined residual and after-the-discontinuity series. As opposed to the no-slope case, a one-tailed test was used, since it was only necessary to detect those cases in which the series after the discontinuity was significantly higher (i.e.,  $t_{s \text{ upper}} < 0$ ) than the residual series. If this test failed to reject the null hypothesis ( $H_0:t_{s \text{ upper}} = 0$ ), then a second test was conducted using the lower residual series. Here, the proportion of years exceeding (falling

below) the 75th (25th) percentiles of the lower residual series were used to compute and statistically evaluate  $t_{s, \text{lower}}$ . Again, a one-tailed test was used to identify cases in which the series after the discontinuity was significantly lower than the residual series. Rejection of  $H_0$  in both cases indicated a significant discontinuity.

An analogous detrending and testing procedure was used when a significant slope was present both before and after the discontinuity. In this case, four detrended series were computed and the original test used to compare the two relevant pairs of detrended series. The application of this test was limited to 2% of discontinuities tested.

In a very small percentage of cases (0.4%), the difference series tests could not be conducted, due to either an inadequate number of neighboring stations or a relatively large pooled standard deviation. Stations were not considered as neighbors if they had fewer than five homogeneous (i.e., no documented discontinuities) years in common with the periods before or after the discontinuity being tested, missing data precluded the computation of an annual exceedence count in the overlapping periods, or they were more than 800 km from the site being evaluated. In this case the discontinuity tests were conducted using the exceedence count series at the station with the potential discontinuity. Allen and DeGaetano (2000) refer to this as a single-station test. Although such a test is clearly less powerful than those based on the difference series, Allen and DeGaetano (2000) showed that it was capable of detecting some discontinuities. More importantly, for the single-station test, the probability of falsely rejecting the null hypothesis of no difference between the subseries was comparable to that based the difference series, provided nonstationary series were detrended prior to the test.

In cases where these tests indicated a significant inhomogeneity, an adjustment factor consistent with the results of the test was formulated. As opposed to variables such as mean temperature, the application of a fixed adjustment (or percent adjustment) to all years after the discontinuity is not applicable to extremes. Rather, for extreme exceedences, a more prudent approach involves a variable adjustment for each year. Here, each annual adjustment is based upon the observed number of exceedences of slightly warmer and/or cooler threshold temperatures. In essence, adjustments for extreme occurrences involve a change in the threshold temperature rather than a static change in annual extreme counts.

As an example, assume that the relocation of a station, at which days  $\geq 90^\circ\text{F}$  are considered "extreme," introduces a  $4^\circ\text{F}$  warming to the subsequent record of daily temperatures. Such a change would precipitate an increase in days  $\geq 90^\circ\text{F}$ , since days on which the temperature would have previously (before the move) only reached  $86^\circ\text{F}$  are now likely to meet the  $\geq 90^\circ\text{F}$  threshold. In such a case, adjustment would involve selecting a new higher threshold such that the number of ex-

ceedences of this new limit is comparable to that associated with the original  $90^\circ\text{F}$  value.

The new threshold was determined through an array of tests in which the series following the discontinuity was based on sequentially higher or lower threshold values. Progressively higher thresholds were indicated when the inhomogeneity was followed by an increase in warm exceedences or a decrease in cold exceedences. Based on the above example, assume that the  $4^\circ\text{F}$  warming resulted in the rejection of  $H_0$  when the series before and after the move were based on a  $90^\circ\text{F}$  threshold. Since such a warming would lead to an increase in days  $\geq 90^\circ\text{F}$ , the series after the break was recomputed based on days  $\geq 91^\circ\text{F}$  and the test reapplied using days  $\geq 90^\circ\text{F}$  prior to the move and the  $\geq 91^\circ\text{F}$  series after the break. Assuming the null hypothesis was again rejected, the test would be repeated using counts of days  $\geq 92^\circ\text{F}$  after the break. This process of increasing the threshold temperature and retesting proceeded until  $H_0$  was accepted and then continued until the number of exceedences following the break was either significantly less than that based on the original  $90^\circ\text{F}$  threshold or  $H_0$  was accepted for 10 consecutive iterations. This suite of tests generally produced a string of one or more threshold values for which no discontinuity was indicated. The median of those thresholds that resulted in acceptance of the null hypothesis was chosen as the adjustment. If the set of tests failed to give a threshold for which  $H_0$  was accepted the exceedences were adjusted based on the average of the thresholds that changed the sign of  $t_s$ .

A final consideration for the adjustment procedure relates to the order in which adjustments are made in series experiencing more than one discontinuity. Although previous homogenized data sets have adopted a reverse chronological approach (i.e., the most recent portion of the record is left unadjusted), in the HCN-XT the longest homogeneous period was left unadjusted. This approach minimized the quantity of data that was subject to adjustment, while maximizing the ability of the test procedure to detect small discontinuities. Allen and DeGaetano (2000) showed the difference series test was able to detect a higher percentage of artificial discontinuities as the length of one of the homogeneous periods increased. Unlike mean temperatures, where an adjustment would need to be added to current observations, the use of the longest period as a base for homogenization merely means that a different extreme threshold is considered for current data. Such time-dependent threshold changes would result regardless of which subseries was initially homogenized. Once the long base period was identified, adjustments proceeded chronologically with the decision to adjust earlier or more recent periods again based on series length. Once adjusted, sequential series were combined to evaluate and potentially adjust later (or earlier) segments of the series. While this approach is fairly straightforward when the overall data record is represented by difference series, at some stations the early portion of the record

must be adjusted using a single-station test. Here, the periods that required use of the single-station test and those for which the difference series test could be applied were treated separately. Three distinct periods—one requiring the single-station test, another based on the difference series approach, and a third intervening period—were generally present in these cases. If a difference series of five or more years could be formed within the intervening period, then an adjustment was computed based on these years and applied to each year within the intervening period. Otherwise, the adjustment applied to the intervening period was based on the single-station approach. The two final homogeneous periods that resulted (one standardized with the single-station test, the other using a difference series) were tested using the single-station approach. If an adjustment was indicated, it was applied to the single-station series, regardless of length.

In all cases inhomogeneity adjustments were developed only for the 90th and 10th percentile exceedence series. These adjustments were then applied to the more extreme thresholds. For instance, if the 90th percentile threshold required a 2°F adjustment, the 95th and 99th percentile thresholds were also increased by 2°F. This was particularly necessary for the 99th (and 1st) percentile series, since the existence of years with no exceedences compromised the reliability of the inhomogeneity test. While separate adjustments could have been developed for the 95th (and 5th) percentile series, analyses showed that values based on the higher thresholds were comparable (i.e., within  $\pm 1^\circ\text{F}$ ) to those derived separately in over 90% of the cases. Thus for simplicity, a single adjustment for all thresholds was adopted. This is also physically realistic, as it is unlikely that a change in station characteristics would cause a different response in temperatures separated by on average 3°F.

#### *f. Computation of adjusted exceedence series*

The net result of the inhomogeneity adjustment procedure was a set of extreme thresholds that varied with time as the station experienced relocations or instrumentation changes. Therefore, the original series, which was based on a constant threshold, was recomputed to account for the inhomogeneities. During years in which the discontinuity occurred, the extreme (warm) threshold for the subsequent year was used when the documented change occurred prior to July. Otherwise, the threshold for the preceding year was used. An analogous procedure was used for the cold extremes.

Once the new series based on the time-dependent thresholds was assembled, it was also necessary to recompute the observation time adjustments. This was required since the regression-based observation time adjustments are based on the monthly exceedence counts. These counts were likely to have changed when the new thresholds were considered. As before, the observation

time of each series was standardized to that which was most prevalent during the station's history. The HCN station history file was generally used to determine observation time changes. However, a supplemental inferred observation time dataset (DeGaetano 2000) was used in cases where the HCN metadata were missing or when the inferred and documented observation times differed for more than three consecutive years. Adjustments during months with different documented or inferred observation times were computed using the regression-based procedure described above. However, predictors were based on the new time-dependent extreme thresholds rather than the constant base period value.

#### *g. Urban classification*

While previous investigators have developed methods to adjust for urban influences in high quality temperature databases (i.e., Karl et al. 1988), the degree of urbanization of each HCN-XT station is simply identified. This approach is intended to increase the utility of the dataset beyond global change research to applications concerning urban-induced changes in temperature extreme occurrences that have impacts on such areas as human health and energy demand.

Data from the Defense Meteorological Satellite Program Operational Linescan System (OLS) were used to characterize the degree of urbanization associated with each station included in the database. Owen et al. (1998) developed a method to objectively define climate stations as urban, suburban, or rural using OLS data. These data represent the frequency that lights were detected within a 1-km grid cell relative to the number of cloud-free grid observations during 1994 and 1995 (Elvidge et al. 1997). Each grid was classified based on two thresholds. The urban threshold was suggested by Imhoff et al. (1997a) and Imhoff et al. (1997b), while Owen et al. (1998) established the rural threshold.

An analysis of local (3 km  $\times$  3 km) and regional (21 km  $\times$  21 km) samples of 1-km grid cells around each station was used to determine whether the station was classified as urban, suburban, or rural. This determination was based on decision logic outlined by Gallo et al. (1999). Stations with any urban grid cells within the local sample were immediately classified as urban. Rural sites contained no urban grids in the local sample and fewer than 25% grid cells classified as urban or suburban in the regional sample. Likewise, suburban stations were associated with no urban grids in the local sample and regional samples containing between 25% and 50% urban or suburban grids. A larger percentage of urban or suburban grids in the regional sample was characteristic of an urban site.

This type of satellite-based urbanization classification is not included in the HCN metadata. In this dataset, urbanization adjustments to the temperature means are based on population (Karl et al. 1990). Thus, the ur-

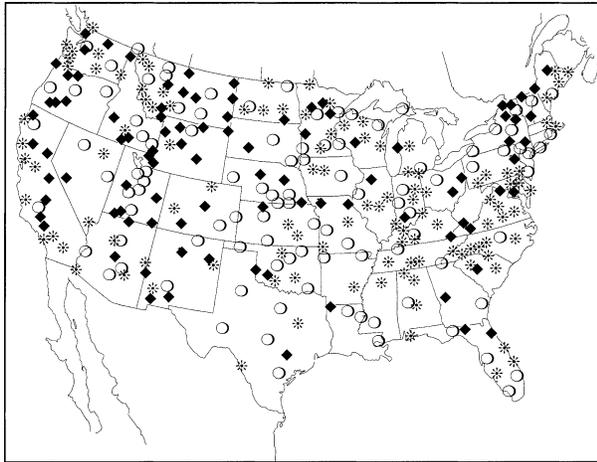


FIG. 9. Locations of rural ( $\blacklozenge$ ), suburban ( $\circ$ ), and urban ( $*$ ) HCN-XT stations.

banization categories included with the HCN-XT data provide supplemental information for studies using the HCN.

The number of stations within each urbanization category was similar. Using this procedure, 136 stations were classified as urban, 126 as suburban, and 99 as rural. The spatial distribution of these station types is shown in Fig. 9.

### 3. Dataset evaluation and applications

Twelve distinct data files comprise the HCN-XT. These correspond to station homogenization and extreme occurrence data files for each of the four threshold categories (warm and cold extremes for daily maximum and minimum temperatures). The datasets are available online ([http://www.nrcc.cornell.edu/hcn\\_XT.html](http://www.nrcc.cornell.edu/hcn_XT.html)). Information regarding the formats of the data files is also included on the Web site.

#### a. Illustrative examples

Based on the comments of an anonymous reviewer, examples of homogenized exceedence series at four sites are presented to give users a qualitative measure of the homogeneity of the series. In each case the adjusted and unadjusted exceedence counts are also presented as differences from the corresponding counts observed at a reference series of neighboring stations. The reference series were constructed in a manner akin to Easterling and Peterson (1995). The stations chosen for illustration were perceived to be difficult to homogenize (by the reviewer) based on the number of nonclimatic changes, climate, and geographic location.

Crater Lake, Oregon, is an isolated high elevation (1976 m) station. The neighboring stations that were available for the homogenization process represented much lower elevations (<230 m). A station move in

1984 was judged to introduce a discontinuity into the cold maximum temperature extreme series. Adjustment required the extreme threshold be raised by  $1^{\circ}\text{F}$ . This increased the adjusted series in the later years producing a visually (and statistically) homogeneous series (Fig. 10a). Crater Lake also experienced several observation time changes through its record. For these cases, exceedence counts based on an afternoon hour were standardized to an 0800 LT observation. This resulted in slight increases in cold extreme exceedences.

Newberry, Michigan, is also isolated based on its location in the Upper Peninsula. Neighboring stations included in the reference series were located more than 400 km from Newberry. Warm minimum temperatures are used in this illustration, with a station move in 1984 introducing a  $1.5^{\circ}\text{F}$  decrease in the extreme threshold. Another move in 1992 brought the threshold back to the pre-1984 level. Newberry also experienced four changes in observation time, switching from an afternoon hour to midnight throughout its history. Despite these changes, only minor adjustments were necessary (Fig. 10b).

At a more southern location, Fairhope, Alabama, experienced several shifts in observation time throughout its record. In the 1950s, observations were taken at midnight. This schedule switched to the afternoon in 1960 and then to a morning hour in 1977. The station also relocated in 1987, but this change did not require adjustment of the extreme warm maximum temperature threshold. Only the period during which observations were taken during the afternoon required adjustment. Visual inspection of the Fairhope exceedence series suggest a possible discontinuity prior to 1955, as the annual counts appear relatively high (Fig. 10c). No documented station changes could be identified during the mid-1950s. However, the period from 1952 to 1955 was characterized by a prolonged drought in Alabama. Presumably, these conditions are related to the high number of temperature extremes during this period based on the preliminary analysis conducted in the next subsection.

Eureka, California, was used as a final example for two reasons. According to the HCN station history file, Eureka is free of relocations and instrument changes. Thus, homogeneity adjustments were not applied to the data series. Furthermore, since observation time has been constant for all but the first two years of the 49-yr record, there is little reason to doubt the homogeneity of this site. The station also potentially presents a challenge to the homogenization based on its coastal location and pronounced maritime climate. Visual examination of the difference series reveals a rather striking discontinuity in the record beginning in 1982, despite the apparent homogeneity of the station based on the HCN metadata (Fig. 10d). Further evaluation of the Eureka station history using metadata provided online by the Western Regional Climate Center (<http://www.wrcc.dri.edu/inventory/sodca.html>) showed that

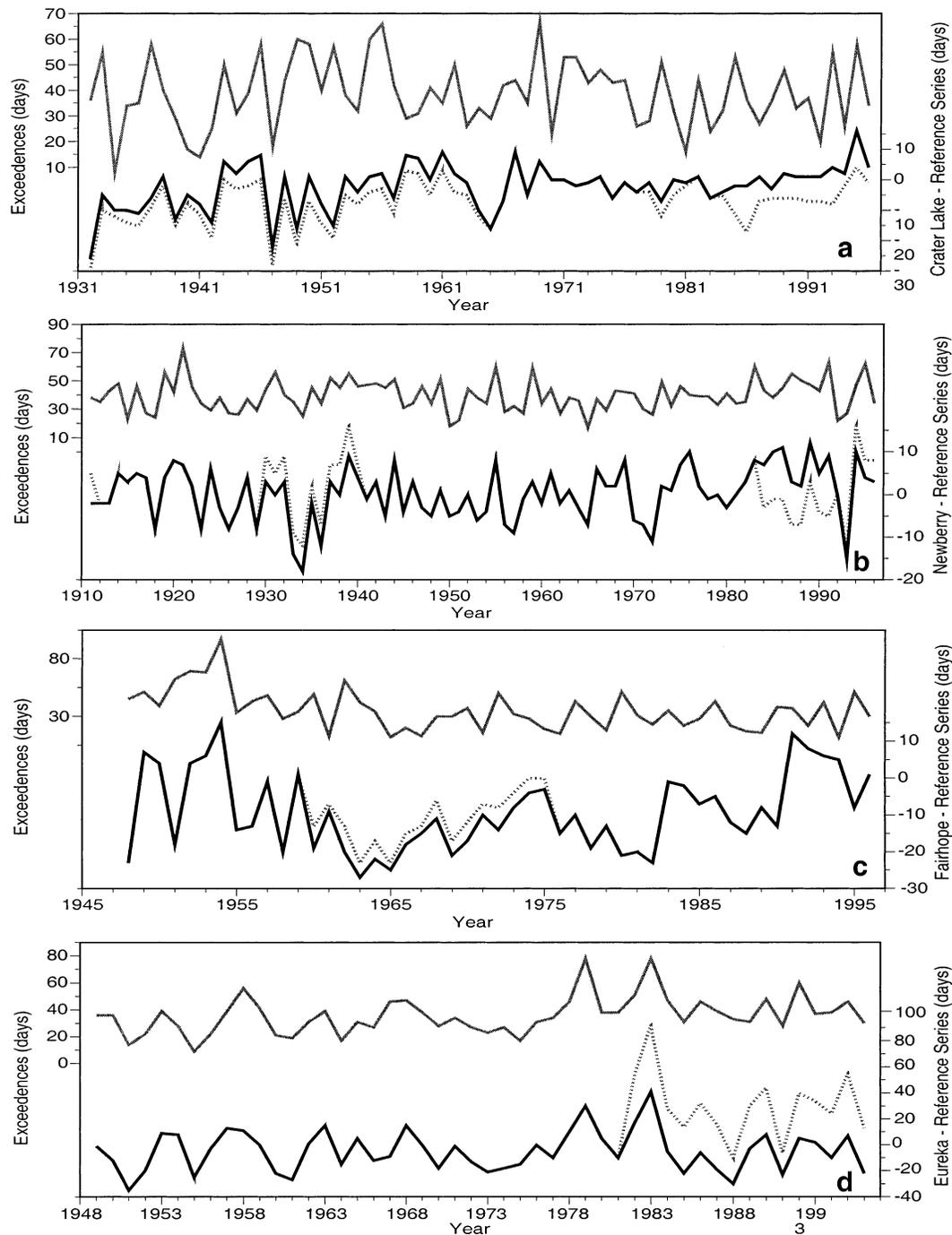


FIG. 10. Examples of homogenized exceedence count series (uppermost series) and adjusted (solid) and unadjusted (dotted) difference series for (a) cold max temperatures at Crater Lake, Oregon; (b) warm min temperatures at Newberry, Michigan; (c) warm max temperatures at Fairhope, Alabama; and (d) warm max temperatures at Eureka, California.

the station was indeed relocated during 1982. Accounting for this change eliminated the discontinuity that was apparent in the unadjusted data (Fig. 10d).

This example highlights an important caveat. The data series adjustments are based only on documented station

location or instrument changes. Undocumented changes may introduce nonclimatic discontinuities into the data. The detection and adjustment for such inhomogeneities is beyond the capabilities of the adjustment procedures developed and applied in this work.

### b. Applications

The original and homogenized temperature extreme time series from the HCN-XT stations were subjected to two analyses to quantify the effects of homogenization. In the first case, temperature exceedence trends over the period 1950–96 were evaluated. This analysis followed the methods of Karl and Williams (1987). A more thorough discussion of this evaluation and its results are given in DeGaetano and Allen (2002). Of interest here are those cases where homogenization of the series resulted in a change in the direction of the trend. On average, over 16% of the series experienced such a trend reversal. Similar percentages were noted for each of the exceedence types (i.e., warm maximum, cold minimum, etc.). Over other time periods more than 20% of the series experienced trend reversals upon homogenization.

Figure 11 shows four cases in which the trends of the adjusted and unadjusted trends were opposite, and both significantly ( $\alpha \leq 0.10$ ) different from zero. For warm maximum temperatures exceeding the 90th percentile, the unadjusted trend at Tallahassee, Florida, is significantly positive ( $\alpha = 0.01$ ). Following adjustment for discontinuities in 1960 and 1986, the trend in extremely warm maximum temperatures reverses (Fig. 11a), while significance at the  $\alpha = 0.01$  level is retained. The change in the warm maximum temperature exceedence trends at Chasm Falls, New York (Fig. 11b), is affected by moves in 1964 and 1985. Chasm Falls also experiences several observation time changes. The net effect of these adjustments over the 1950–96 period is a positive trend ( $\alpha = 0.10$ ) as opposed to a decreasing trend ( $\alpha = 0.05$ ). In Figs. 11c and 11d, negative trends ( $\alpha = 0.05$ ) in warm minimum temperature exceedences at Minden, Nebraska, and Geary, Oklahoma, become positive ( $\alpha = 0.10$  and  $0.05$ , respectively) following adjustment. Relocations in 1960, 1970, and 1981 precipitated the adjustments at Geary, while the discontinuity at Minden occurred in 1988. A single observation time change occurred in 1968 at Minden, while several changes in observation schedule were documented at Geary.

The effect of homogenization becomes less pronounced when the station-specific time series are averaged into regional or national composite trends. In such cases, the net effect of homogenization depends on several factors including the number of stations that comprise the composite and the existence of systematic data biases such as the adoption of new instrumentation at the majority of sites or a network wide shift in the preferred observation schedule. Nonetheless, the effect of homogenization is not entirely negligible when such composites are considered. For instance, averaged over stations in the central ( $95^{\circ}$ – $110^{\circ}$ W longitude) United States, the unadjusted warm maximum temperature extreme exceedence series (1960–96) decreases by 0.07 exceedences per squared year. Upon homogenization,

this series increases by 0.03 exceedences per squared year. When averaged over all continental United States HCN-XT stations, a similar change in slope (0.07 exceedence per squared years) results from homogenization.

The HCN-XT can also be used to relate the interannual variations in extreme occurrence to other meteorological conditions. For instance, warm maximum temperature exceedences are negatively correlated with Palmer drought severity index (PDSI). Using all HCN-XT stations, the correlation between the annual number of exceedences of the 90th percentile and JJA average PDSI of the corresponding climate division is  $-0.46$ . Similar correlation is found for exceedences of the 95th and 99th percentiles. However, when warm minimum temperature exceedences are considered, the correlation drops to  $-0.16$ .

Figure 12 compares annual exceedence count (95th percentile) boxplots for years having average JJA PDSI values  $\leq -2.0$  with those of the remaining years. Stratification of the extreme exceedence counts is based on both the adjusted and nonadjusted 1930–96 time series. At each station, the exceedence distributions are skewed toward higher values during the drought years. Adjustment increases the drought versus nondrought differences at two stations (Selma and Dufur), but lessens the skew at the other sites. Figure 13 shows the adjusted and unadjusted series at Dufur and El Dorado. Adjustment of the Dufur record is relatively extensive giving rise to changes in both the drought and nondrought distributions. At El Dorado, the adjustments are more subtle and confined to the 1930s and 1940s. Nonetheless, the adjustments reduce the median exceedence counts during the drought years by 5 days and cause a reduction in the spread of the nondrought boxplot.

A more thorough analysis of the relationship between drought and temperature extreme occurrence is beyond the scope of this study. However, a cursory look at a larger network of stations quantitatively supported the adjustments. The most pronounced differences between PDSI distributions during high and not-high exceedence count years was apparent in the Southeast and Midwest. Only minor differences were evident in the Northeast and West.

## 4. Summary

The development of a 361-station homogenized daily temperature extreme dataset is discussed. The dataset lists occurrences of maximum and minimum temperatures that exceed the 90th or 10th percentiles of the daily temperature distribution for a subset of the daily U.S. HCN stations. The extreme series that comprise the dataset have been serially completed and adjusted for nonclimatic inhomogeneities using methods that have been specifically designed for temperature extreme data. The data estimation routine eliminates the underestimation of extreme temperature exceedence counts that is char-

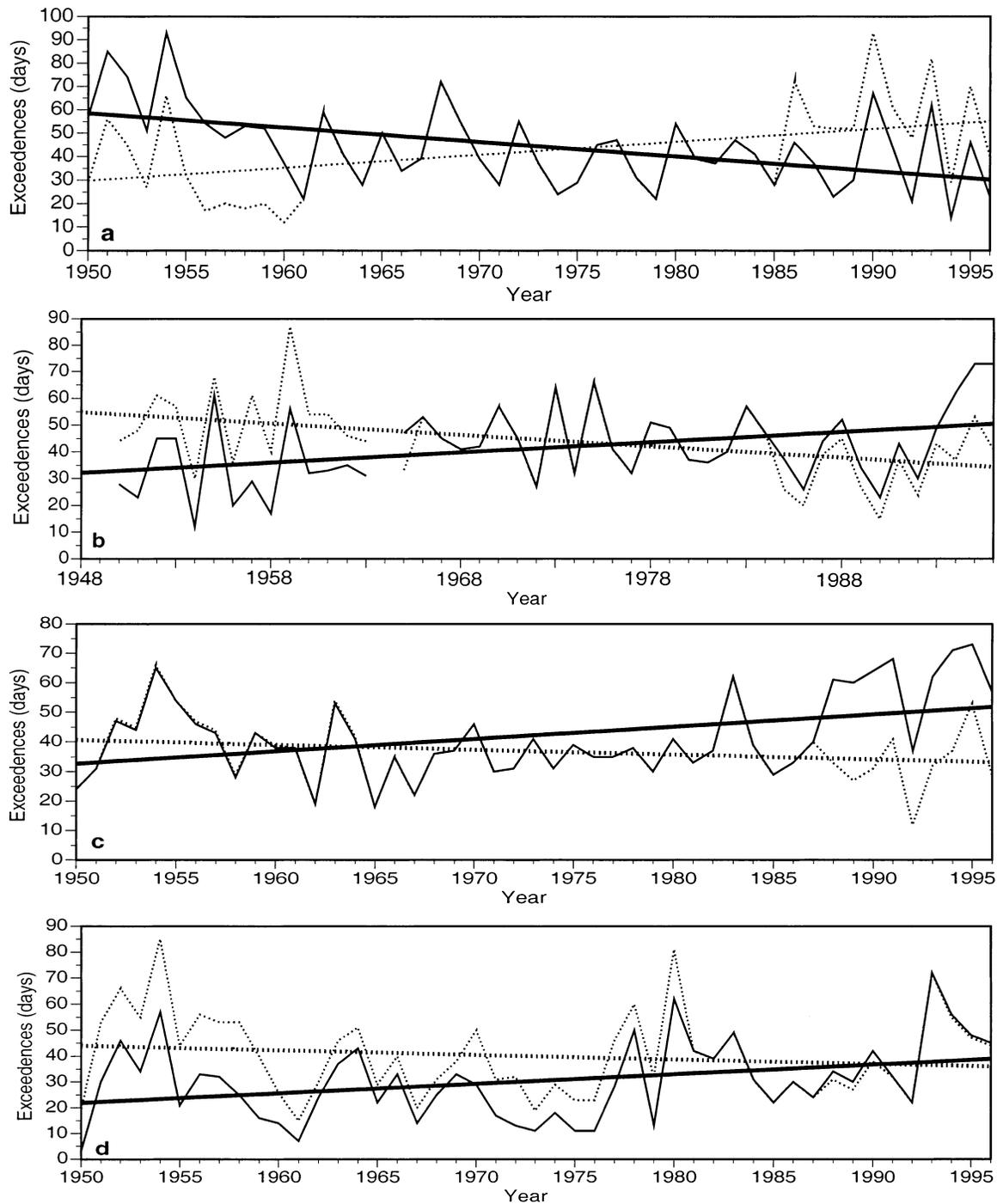


FIG. 11. Adjusted (solid) and unadjusted (dotted) exceedence count time series for (a) Tallahassee, Florida; (b) Chasm Falls, New York; (c) Minden, Nebraska; and (d) Geary, Oklahoma showing a reversal in slope for the two series (both slopes significant at  $\alpha = 0.10$ ): (a), (b) warm max and (c), (d) warm max/min temperature exceedences.

acteristic of existing techniques. Although the detection of inhomogeneities in the temperature extreme exceedence series is based on existing methods for mean temperature, adjustment of the series involves iterative adjustment of the extreme threshold rather than a translation of the series. The adjusted counts better reflect

the natural interannual variations in extreme occurrence, since little if any adjustment is applied during years in which temperatures rarely approach the unadjusted extreme threshold. Conversely, a relatively large adjustment may be indicated in years that frequently approach the original threshold. Such year-to-year variability

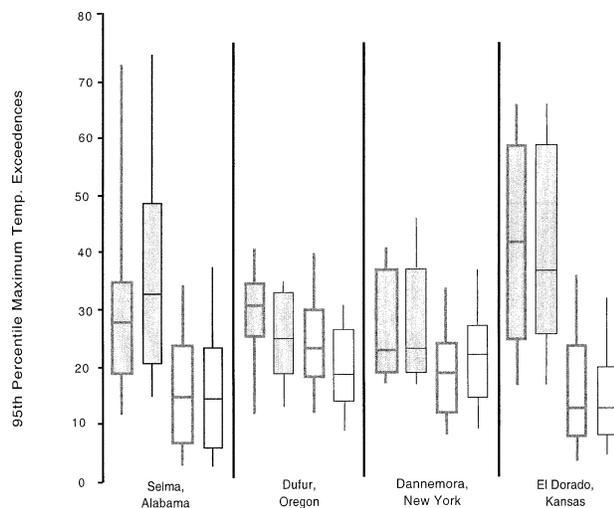


FIG. 12. Boxplots of annual 95th percentile max temperature exceedences during drought (shaded) and nondrought years (unshaded). The distributions are shown for the unadjusted (thick gray) and adjusted (thin black) data series.

would be lost using a static adjustment. Similarly, observation time adjustments are also a function of the number of unadjusted extreme thresholds exceedences.

The dataset also includes a set of files that document the adjustment history of the series. This allows users to identify the original homogeneous (based on metadata) periods that drove the adjustment procedure and perhaps select stations based on the stringency of the parameters (e.g., record length or number of neighboring stations) used to homogenize the data record. A measure of median cross-validation estimation error is also given for each homogeneous subseries. A separate station metadata file is not included as this information is available as part of the HCN. However, the stations are identified by their coordinates and categorized as urban, rural, or suburban based on satellite data.

It is expected that this dataset can provide a benchmark for studies examining temporal trends in any number of extreme temperature-related parameters. These could include counts of single-day threshold exceedences, exceedence runs, etc. An example of such an analysis using both the adjusted and unadjusted data series shows that across the HCN between 15% and 20% of the time series experience a change in slope following adjustment. The data could also be used to derive temperature extreme degree-day statistics (Sen et al. 1998). However, the analysis of this variable would require that a separate time of observation adjustment be developed.

The dataset also provides a foundation for studies concerned with the interannual variability of temperature extreme occurrence. Such studies could have implications in areas ranging from climate impact analysis to seasonal climate forecasting. A change in the distribution of exceedence counts during drought and non-

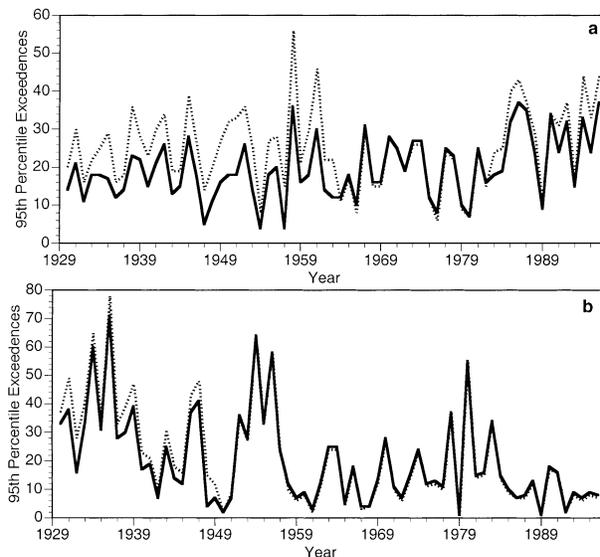


FIG. 13. Adjusted (solid) and unadjusted (dotted) 95th percentile max temperature exceedence series for (a) Dufur, Oregon, and (b) El Dorado, Kansas.

drought conditions based on the adjusted and unadjusted data is also shown.

**Acknowledgments.** This work was supported by NASA/NOAA Grant NA76GP0351 and NOAA Grant NA76WP0273. Thanks are extended to Steve Hudson for developing the observation time adjustment routines and to two anonymous reviewers for their constructive comments. The daily HCN data were obtained through the generosity of Dale Kaiser of the Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory.

#### REFERENCES

- Allen, R. J., and A. T. DeGaetano, 2000: A method to adjust long-term temperature extreme series for non-climatic inhomogeneities. *J. Climate*, **13**, 3680–3695.
- , and —, 2001: Estimating missing daily temperature extremes using an optimized regression approach. *Int. J. Climatol.*, **21**, 1305–1319.
- Balling, R. C., and S. E. Idso, 1990: Effects of greenhouse warming on maximum summer temperatures. *Agric. For. Meteorol.*, **53**, 143–147.
- DeGaetano, A. T., 1996: Recent trends in extreme maximum and minimum temperature occurrences in the northeastern United States. *J. Climate*, **9**, 1646–1660.
- , 1999: A method to infer observation time based on day-to-day temperature variations. *J. Climate*, **12**, 3443–3456.
- , 2000: A serially complete simulated observation time metadata file for U.S. daily Historical Climatology Network stations. *Bull. Amer. Meteor. Soc.*, **81**, 49–67.
- , and R. J. Allen, 2002: Trends in twentieth century temperature extremes across the United States. *J. Climate*, in press.
- , K. L. Eggleston, and W. W. Knapp, 1995: A method to estimate missing daily maximum and minimum temperature observations. *J. Appl. Meteorol.*, **34**, 371–380.
- Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis*. Wiley, 708 pp.

- Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369–377.
- , T. R. Karl, J. H. Lawrimore, and S. A. Del Greco, 1999: United States Historical Climatology Network Daily Temperature and Precipitation Data (1871–1997). ORNL/CDIAC-118, NDP-042/R1, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 23 pp.
- , J. L. Evans, P. Ya. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje, 2000: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.*, **81**, 417–425.
- Eischeid, J. K., P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.*, **39**, 1580–1591.
- Elvidge, C. D., K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis, 1997: Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogramm. Eng. Remote Sens.*, **63**, 727–734.
- Gaffen, D. J., and R. J. Ross, 1999: Climatology and trends of U.S. surface humidity and temperature. *J. Climate*, **12**, 811–828.
- Gallo, K. P., T. W. Owen, D. R. Easterling, and P. F. Jamason, 1999: Temperature trends of the U.S. Historical Climatology Network based on satellite-designated land use/land cover. *J. Climate*, **12**, 1344–1348.
- Imhoff, M. L., W. T. Lawrence, C. D. Elvidge, T. Paul, E. Levine, M. V. Privalsky, and V. Brown, 1997a: Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States. *Remote Sens. Environ.*, **59**, 105–117.
- , —, D. C. Stutzer, and C. D. Elvidge, 1997b: A technique for using composite DMSP/OLS “city lights” satellite data to map urban area. *Remote Sens. Environ.*, **61**, 361–370.
- Kalkstein, L. S., and R. E. Davis, 1989: Weather and human mortality: An evaluation of demographic and inter-regional responses in the U.S. *Ann. Amer. Assoc. Amer. Geogr.*, **79**, 4–64.
- 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.
- , and D. R. Easterling, 1999: Climate extremes: Selected review and future research directions. *Climatic Change*, **42**, 309–325.
- , C. N. Williams Jr., P. M. Young, and W. M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum minimum and mean temperatures for the United States. *J. Climate Appl. Meteor.*, **25**, 145–160.
- , H. F. Diaz, and G. Kukla, 1988: Urbanization: It’s detection and effect in the United States climate record. *J. Climate*, **1**, 1099–1123.
- , C. N. Williams Jr., and F. T. Quinlan, 1990: United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data. ORNL/CDIAC-30, NDP-019/R1, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 274 pp.
- Kunkel, K. E., R. A. Pielke Jr., and S. A. Changnon, 1999: Temporal fluctuations in weather and climate extremes that cause economic and human health impacts: A review. *Bull. Amer. Meteor. Soc.*, **80**, 1077–1098.
- Mearns, L. O., R. W. Katz, and S. H. Schneider, 1984: Extreme high-temperature events: Changes in their probabilities with changes in mean temperature. *J. Climate Appl. Meteor.*, **23**, 1601–1613.
- Owen, T. W., K. P. Gallo, C. D. Elvidge, and K. E. Baugh, 1998: Using DMSP/OLS light frequency data to categorize urban environments associated with U.S. climate observing stations. *Int. J. Remote Sens.*, **19**, 3451–3456.
- Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- Ross, R. J., and W. P. Elliot, 1996: Tropospheric water vapor climatology and trends over North America. *J. Climate*, **9**, 3561–3574.
- Sen, Z., M. Kadiolu, and K. Koçak, 1998: Comments on “Recent trends in maximum and minimum temperature threshold exceedences in the northeastern United States.” *J. Climate*, **11**, 2147–2149.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 464 pp.
- Zwiers, F. W., and V. V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO<sub>2</sub> doubling. *J. Climate*, **11**, 2200–2222.