A Method to Adjust Long-Term Temperature Extreme Series for Nonclimatic Inhomogeneities

ROBERT J. ALLEN AND ARTHUR T. DEGAETANO

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

(Manuscript received 23 April 1999, in final form 10 January 2000)

ABSTRACT

A method to homogenize nonclimatic discontinuities in temperature extreme exceedence series is developed and evaluated. The method is based on a set of complementary tests with the application of an individual test depending on the availability of an adequate network of nearby homogeneous reference stations and the presence of significant trends in the resulting difference or original exceedence series. Given a suitable set of neighboring reference stations, a comparison of the differences in exceedences between the inhomogeneous station and neighboring sites is made for the periods before and after the documented discontinuity. In the absence of one or more reference stations, the exceedences at the inhomogeneous station are compared before and after the nonclimatic change. A method by which nonstationary series are detrended and subsequently evaluated is also presented.

When tested using homogenized data series into which an artificial discontinuity of known magnitude was introduced, as many as 80% of the ±1°F discontinuities were detected by the difference series approach. The performance of the single-station exceedence series test was less accurate. Although in a few cases, less than 40% of the ±1°F discontinuities were detected, between 60% and 76% of the ±2°F discontinuities were identified. Using both tests, the probability of falsely detecting a discontinuity (i.e., identifying an inhomogeneity when none existed) was 5%. Provided both methods deemed a documented inhomogeneity significant, the magnitude of the adjustments imposed by both methods was similar.

1. Introduction

Few, if any, climatological records are free of irregularities introduced by such nonclimatic causes as observation time differences, station relocations, and instrument changes. To detect these discontinuities, researchers have previously compared series of differences in mean annual temperature between a suspect station and one or more nearby highly correlated stations. A comprehensive review of such methods is given by Peterson et al. (1998). Therefore, for conciseness, only those methods with direct relevance to the current procedure are discussed here.

In general, previous methods have used the change in the average temperature difference between a homogeneous reference station and a neighboring inhomogeneous station to adjust the latter station's record. Nelson et al. (1979) based their adjustment on data representing the three years before versus the three years after the station change. Karl and Williams (1987) tested discontinuities in difference series corresponding to known station changes. In cases where the null hypothesis of no difference between the series before and after the known move was rejected, a correction factor, based on the difference of means from period 1 (before the discontinuity) to period 2 (after the discontinuity), was applied to the earlier homogeneous period.

Easterling and Peterson (1995) developed a similar test based on mean annual temperature differences, but discontinuities were identified without relying on station history files. Vincent (1998) presented a somewhat different technique to identify nonclimatic step changes and trends in mean temperature series without prior knowledge of station changes. Her approach is regression-based with predictors corresponding to temperatures at neighboring stations and dependent variables represented by the temperature series at the inhomogeneous station. Following application of the regression model, undocumented discontinuities manifest themselves as autocorrelated residuals, while random residuals are characteristic of a homogeneous temperature series.

Discontinuities also exist in precipitation records. Alexandersson (1986) used a series of ratios between precipitation amounts at a suspect station and surrounding homogeneous sites to identify discontinuities. A likelihood ratio technique was then applied to assess the

Corresponding author address: Dr. Art DeGaetano, Northeast Regional Climate Center, Cornell University, 1115 Bradfield Hall, Ithaca, NY 14853.
E-mail: atd2@cornell.edu

significance of these nonclimatic changes. Karl and Williams (1987) also used ratios to adjust precipitation series for known discontinuities.

While such methodologies have been shown to be effective for detecting and adjusting for discontinuities in mean temperature (or total precipitation) series, little effort has been devoted to assessing and adjusting discontinuities affecting extreme temperature time series. The analysis of such series has become a priority in the detection of anthropogenically induced climate change (Houghton et al. 1996). Although presumably established methods for assessing and adjusting mean discontinuities can be modified to accommodate extreme series, a rigorous analysis addressing extreme discontinuities has yet to be presented.

In this study, we devise a method to test and adjust extreme temperature exceedence time series for documented inhomogeneities. Our method is based on Karl and Williams' (1987) technique for adjusting mean temperature series. However, we have substantially modified this existing procedure to account for differences in the underlying causes of inhomogeneities in the extreme versus mean temperature series. For instance, consider the relocation of a station to a site conducive to the pooling of cold air under favorable radiational cooling conditions. Assuming this is the only microclimatic difference between the old and new station locations, the mean correction is likely to be small (and perhaps statistically insignificant), since these conditions affect only a few temperature observations in each year. However, extreme (cold) minimum temperatures are likely to occur under these conditions and thus the resulting inhomogeneity and subsequent adjustment of these temperatures is likely to be significant. Similar effects for extreme (warm) maximum temperatures are likely if a change in instrument shelter design results in different radiative characteristics. In this case, inhomogeneities would be maximized on sunny days, which are likely to be the warmest summer days.

A second, and perhaps more important difference between the homogenization of means and extremes is related to the use of exceedence counts to characterize temperature extreme occurrence. Since annual or monthly mean temperature data are continuous, previous methods have assumed than a nonclimatic inhomogeneity is characterized by a translation of the series average. Thus, following the break, the value for each year is adjusted by an equal mount. This cannot be assumed for the discrete exceedence count data used here. To illustrate this point, consider a series of annual counts of days with a maximum temperature $\geq 32°C$. If a 1°C warming is introduced to the series, the count during a year in which a maximum temperature of 31°C was not reported would not change, since it is only on these days that the warming would increase the original temperature to the level of the threshold. Conversely, in a year when 31°C was the maximum temperature on 10 days, the 1°C warming would increase the exceedence count by 10 days. Clearly, in these cases the adjustment must rely on the number of days with a temperature near the threshold. This tendency must also be taken into account when assessing the statistical significance of a documented station move.

Since the construction of extreme temperature series relies on daily climatological data, the methods previously developed to address inhomogeneities in mean series may also be plagued by problems related to data sparsity. For instance, in the U.S. Historical Climatology Network (HCN) (Karl et al. 1990), monthly temperature data exist for a substantially longer period of record and at about 200 more stations than are associated with daily values (Easterling et al. 1999). Presumably this is also the case with other long-term climatological datasets. The paucity of daily data prior to the mid-twentieth century limits the use of tests such as those described by Karl and Williams (1987) and Easterling and Peterson (1995) during the early stages of most long-term time series. These methods are difficult to implement when data from adjacent stations are scarce or nonexistent. Early in the record and in remote areas, data from the available reference stations may contain too many discontinuities or have relatively low correlation with the suspect station and thus the ability of the tests to identify inhomogeneities is compromised. To account for these data limitations, particularly in the longest data records included in the HCN, we also present an adaptation of our method that does not rely on data from neighboring stations. Clearly, the use of such a test is only intended for cases where data limitations prevent the use of the reference-station-based method.

The fundamentals of our test are described in section 2. In section 3, we present an additional modification to our test that is necessary when either the difference series (neighboring-station test) or exceedence count series (single-station test) is not stationary. Although this occurrence was discussed by Karl and Williams (1987), it is not specifically addressed by their procedure. Tuomenvirta and Alexandersson (1995) present a linear correction term for adjusting nonstationary temperature time series. However, we have chosen to adopt a different method of dealing with such series that is more in line with our test for stationary temperature extreme series. We present an analysis of the performance of our test procedure using idealized examples in section 4 before describing and evaluating a procedure to adjust significant temperature extreme discontinuities in sections 5 and 6.

2. Test fundamentals

Unlike prior tests, which are based on annual, seasonal, or monthly means, the modified test is based on annual extreme temperature threshold exceedences. Here, annual extreme temperature threshold exceedences are defined as the number of days in which the maximum (or minimum) temperature exceeds (or falls

below) the xth percentile of the distribution of all maximum or minimum temperatures at a station. In general, x is some value ≥ 75 (or ≤ 25).

The basis of the test is a comparison of the 75th and 25th percentiles of a time series describing annual extreme temperature threshold exceedences before and after an inhomogeneity. Provided an adequate set of adjacent reference stations exist, the series takes the form of a difference series, with annual values defined by ($E_{di} - E_{ri}$), where E_d is the number of threshold exceedences in year i at the inhomogeneous station and E_r is a weighted-average of exceedences for the reference series sites. To assure an adequate degree of spatial correlation with the inhomogeneous station, each E_r value is based on the station's own xth temperature percentile, rather than the specific temperature corresponding to the E_d threshold. Based on this measure, correlation between the change-in-exceedence-counts-per-unit-time series (Peterson and Easterling 1994) typically exceeds 70%. This is in line with the 75% value reported for similar annual mean temperature series by Easterling and Peterson (1995).

If data limitations preclude the assembly of an adequate set of reference stations, then the series simply reflects the annual E_d values. Identifying a sufficient number of neighboring reference stations does not appear to be problematic using the U.S. HCN, as difference series based on at least 10 yr (5 before and 5 after the documented change) of homogeneous reference station data could be constructed for the majority of stations for the most recent 50 yr of data. This task becomes increasingly more difficult for stations with a record that commences before or during the early 1900s. It is at these sites, and in data sparse regions (primarily the intermountain region of the United States), that the use of the single-station test (E_d series) is indicated. However, as is shown in section 4, these results should be applied with caution.

At present, potential inhomogeneities are identified using station history files. Based on these data, it is possible to identify changes in station location, observation time, instrument type, instrument height (e.g., roof top versus ground), and to some degree, observer. In all cases, a change in any one of the first four attributes indicates a potential discontinuity that must be tested. Observer changes were not considered, since it was presumed that these changes would have a minimal effect on the homogeneity of the series and since changes in the specific observer at many stations are not given (e.g., at government or university sites). 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 is difficult to consider such changes as potential inhomogeneities. In practice, however, existing techniques such as Easterling and Peterson's (1995) can be used to detect undocumented inhomogeneities in the mean temperature series, which can then be retested for discontinuities in the annual exceedence series using our procedure.

Using these documented changes, the exceedence count time series is divided into two (or more) periods based on the year(s) of the inhomogeneity. The 75th and 25th percentiles of the longer of the two periods ($P75_1$ and $P25_1$) are then calculated. Using these values, the proportion of years in the shorter period that exceed $P75_1$ ($P75_2$) is calculated, as is the proportion of years that fall below $P25_1$ ($P25_2$). A test statistic is computed as

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

If the two periods are similar (i.e., no discontinuities exist) $P75_2 \approx 0.25$. 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: $t_s = 0$.

Once t_s is computed, the statistical significance of the discontinuity is assessed by resampling techniques. Here, the combined series (i.e., the years before and after the discontinuity) is randomly sampled with replacement 1000 times. For each reordering, a new t_s value is calculated creating a distribution of t_s consistent with the null hypothesis of no difference before and after the discontinuity. Distributions of t_s changed little when the combined series were resampled more (10 000) or fewer (500) times.

The significance of the original t_s value is then assessed using this distribution. Acceptance of H_0 indicates no difference between the temperature exceedence series before or after the discontinuity, ignoring possible changes in variability between the two periods. Figure 1a shows a case in which the documented station move is not associated with a discontinuity. Here $t_s = 0$ since $0.25 = P75_2$ and $0.25 = P25_2$. In Fig. 1b it appears that the station move coincided with a change in variability. In this case, t_s also equals 0 since $(0.25 - P75_2) = (0.25 - P25_2)$. The test statistic was constructed in this manner since it is assumed that a station move or other related discontinuity will affect all years in the same manner and not produce a change in variance. Such a change would imply that the move resulted in warm years becoming cooler, while cool years became warmer.

The test statistic becomes increasingly positive when the second period becomes colder than the first. Conversely, t_s becomes increasingly negative as the second period becomes warmer than the first. In Fig. 1c, the station move is associated with a warmer second period. Here, $(0.25 - P75_2) < 0$, whereas $(0.25 - P25_2) > 0$, resulting in a test statistic < 0.

In this modified test, the value of t_s given in Eq. (1) replaces the offset between the difference series before and after the discontinuity that Karl and Williams (1987)

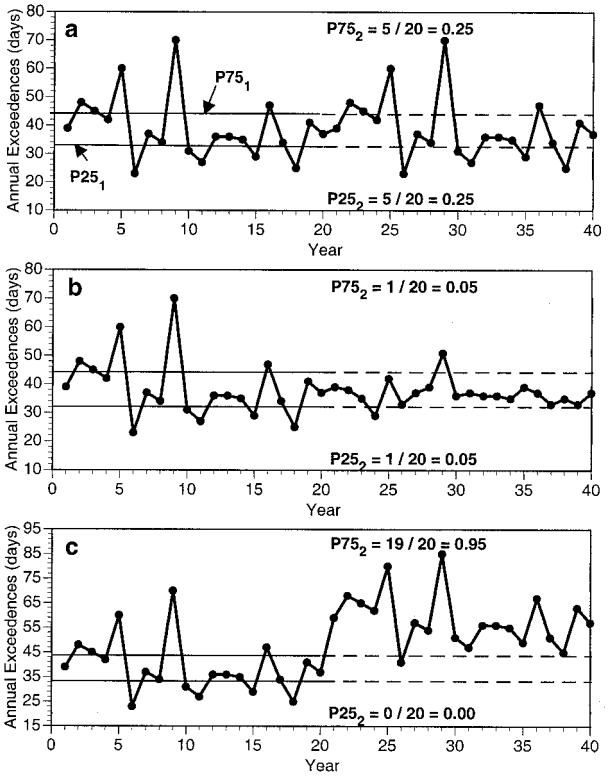

FIG. 1. Examples showing the computation of t_s for cases in which a documented station move (white-gray boundary) is associated with (a) no discontinuity, (b) only a change in variability, and (c) an increase in extreme temperature exceedences. The horizontal solid lines show the 25th and 75th percentiles of the series prior to the move ($P25_1$ and $P75_1$, respectively), while the dashed lines project these values onto the series after the break.

define in their Eq. (1) as $d_b - d_a$. We have also chosen to select neighboring stations in a fashion similar to Karl and Williams (1987). However, in this regard, differences between the tests necessitate some modification to Karl and Williams' procedure for minimizing the confidence interval widths associated with the set of difference series. These differences arise from their use of the Student's t-test and our adoption of a nonstandard test statistic. Initially, we computed confidence interval widths based on our resampled t_s distributions to match Karl and Williams reference-station selection procedure. However, subsequent evaluations of the test procedure using this method of reference-station selection (shown in section 4) were generally less satisfactory than those obtained when the minimization of pooled standard deviation (s_p) alone was used as a criterion for combining neighboring stations. It appears that the discrete nature of the exceedence series was responsible for the less desirable performance of the resampled confidence interval widths.

Once a set of reference stations was selected, the combined series was formed by weighting corresponding values from each series by their respective pooled standard deviation and then summing each set of annual exceedences. Furthermore, since the exceedence series in some cases violated the assumptions of normality (i.e., the left tail of the exceedence count distribution is bounded by zero), the standard parametric tests employed by Karl and Williams (1987) were not strictly valid. Although Karl and Williams employed the Wilcoxon rank-sum test in such nonnormal cases, we have opted to use a Monte Carlo procedure.

3. Nonstationary test

Application of the test requires that the series before and after the suspected discontinuity be stationary (i.e., no significant trend). Based on station history data from HCN, such exceedence series occur in approximately 75% of the cases. Clearly, when one or both time periods have a significant slope, the test incorrectly rejects H_0 too frequently. To assure that both time series are stationary, each must be tested for a significant slope (Wilks 1995) prior to application of the test. While such nonstationarities are of primary concern when data limitations require that the E_d series be tested, difference series can also be compromised. For instance, a nonstationary difference series might result from gradual environmental changes at either the inhomogeneous or neighboring reference station(s) (Karl and Williams 1987). For cases in which a significant slope is detected, an alternative test procedure has been developed.

a. Single significant slope

In cases where the time series displays a single significant slope either before or after the documented inhomogeneity, the series is detrended prior to the application of the test procedure. The detrended series simply represents the residuals obtained from a linear least squares fit of the original time-dependent series. After fitting this regression, 95% confidence intervals for the slope and intercept are computed. For simplicity these intervals are defined as

$$b \pm 2.0\sigma_b, \tag{2}$$

for the slope (b) and

$$a \pm 2.0\sigma_a, \tag{3}$$

for the intercept (a).

The line given by the combination of the smallest intercept and the largest slope is projected to the year of the discontinuity. A second line, given by the largest intercept and the smallest slope, is likewise projected. The two resulting projected values at the year of the discontinuity are used as new base values to translate the original residual series into two new series. This translation simply involves adding the new base value to each residual from the original regression. Using this procedure the nonstationary series is described by two

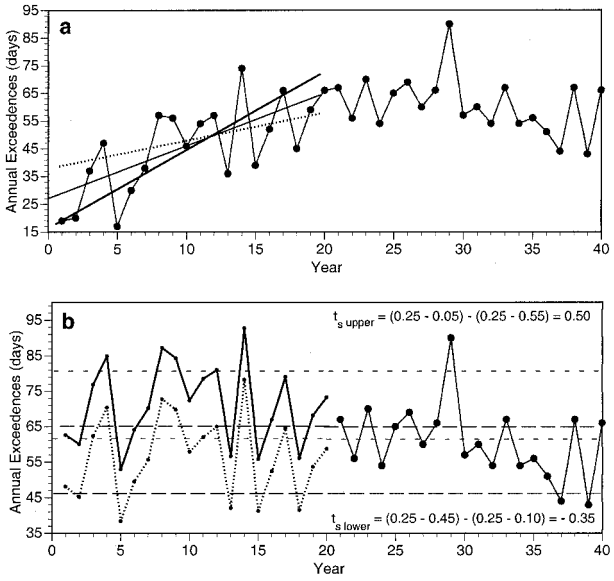

FIG. 2. Illustration of the procedure used to detrend and test a nonstationary difference or exceedence series. (a) The least squares regression line (thin solid) fit to the series prior to the discontinuity is given along with the lines describing the 95% confidence interval about the regression (thick black and dotted). (b) The original time series has been detrended about the upper and lower limits of the confidence interval at the year of the discontinuity (black solid and dotted series, respectively). The 75th and 25th percentiles of these series are given by horizontal black lines (short dash for upper, long dash for lower). The projections of these percentiles to the period after the discontinuity (shaded area) is shown as gray dashed lines.

stationary series at the upper and lower limits of the 95% confidence interval about the original regression.

Although strictly speaking the computation of the 95% confidence interval in this manner is likely to underestimate the true 95% confidence interval for the regression (Draper and Smith 1981), the empirical nature of the subsequent significance tests makes this inaccuracy, as well as the assumption that the Student's t value = 2.0 in Eqs. (2) and (3) inconsequential. Certainly, this detrending procedure could have been developed for different confidence interval widths and/or definitions (e.g., the 95% confidence interval for the predicted exceedence count in the year prior to the discontinuity), provided that correspondence between the desired overall level of statistical significance and the empirical percentage of false H_0 rejections were in correspondence.

Figure 2 illustrates this detrending procedure. Here, the series prior to the discontinuity is associated with a significant positive slope, while the period after the discontinuity is stationary. Superimposed on the nonstationary series in Fig. 2a are the least squares regression line as well as the lines used to define the 95% confidence interval at the year of inhomogeneity. In Fig. 2b, the detrended residual series is translated forming two series, one centered on the upper limit of the regression confidence interval with the second at the lower limit.

The test now proceeds 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 is computed and used to calculate $t_{s\ \mathrm{upper}}$ (short-dashed lines in Fig. 2b). The statistical significance of $t_{s\ \mathrm{upper}}$ is then 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 is used, since it is only necessary to detect those cases in which the series after the discontinuity is significantly ($\alpha = 0.125$) higher (i.e., $t_{s\ \mathrm{upper}} < 0$), than the residual series. If this test fails to reject the null hypothesis (H_0: $t_{s\ \mathrm{upper}} = 0$), then a second test is conducted using the lower-residual series. Here, the proportion of years exceeding (falling below) the 75th (25th) percentiles of the lower-residual series are used to compute and statistically evaluate $t_{s\ \mathrm{lower}}$ (long-dashed lines in Fig. 2b). Again, a one-tailed test is used to identify cases in which the series after the discontinuity is significantly ($\alpha = 0.125$) lower than the residual series. In Fig. 2b, $t_{s\ \mathrm{upper}} > 0$ while $t_{s\ \mathrm{lower}} < 0$, leading to overall acceptance of the null hypothesis H_0: $t_s = 0$. However, if the individual null hypotheses had been rejected (i.e., $t_{s\ \mathrm{upper}} > 0$ and $t_{s\ \mathrm{lower}} > 0$ or $t_{s\ \mathrm{upper}} < 0$ and $t_{s\ \mathrm{lower}} < 0$) then the occurrence of a significant discontinuity would be suspected.

The choice of $\alpha = 0.125$ in the above one-tailed tests is based on a suite of iterative trials using randomized exceedence series. Observed series of 20 or more homogeneous (i.e., no documented discontinuities) annual exceedence counts were bootstrapped to remove the effects of any undocumented changes. Using these series, a significant trend was introduced to the earliest 10 or more years, by simply adding a cumulative offset to each observation. For instance the first count was increased by one, the second by two, etc. A constant offset, equal to that assigned to the year prior to the break, was added to the count for each year after the artificially imposed discontinuity. These series were then tested for different levels of α using the above procedure. To coincide with the use of $\alpha = 0.05$ in the stationary case, it was appropriate for the combination of the two one-tailed tests to result in the null hypothesis of no difference between the detrended nonstationary series prior to the break and the subsequent stationary series being incorrectly rejected in 5% of the cases. The use of $\alpha = 0.125$ produced this result, regardless of the magnitude of the imposed slope or the station or exceedence series tested. However, clearly, this level is a function of the confidence interval width used to form the detrended series.

b. Two significant slopes

It is possible to extend the test described in section 3a to cases in which the series both before and after the discontinuity exhibit significant slopes. In this case, the

detrending technique is applied to both temperature exceedence series yielding four detrended residual series. A pair of one-tailed tests is used to compare the upper-residual series prior to the discontinuity (PU) with the lower-residual series after the discontinuity (AL) and separately the remaining two residual series (PL and AU). As in the one-slope case, it was necessary to empirically derive the appropriate α level for these tests. Given the large degree of uncertainty introduced by the use of two sets of detrended series, a relatively large α level of 0.55 was indicated for this combination of tests. Based on $\alpha = 0.55$, if the null hypothesis $t_s = 0$ is rejected in favor of the alternative hypothesis $t_s < 0$ when testing PU versus AL, then the discontinuity is associated with a significant increase in exceedences. Likewise if H_0 is rejected in favor of the alternative hypothesis $t_s < 0$ when testing PL against AU, the inhomogeneity corresponds to a significant decrease in exceedences. Clearly, in this case only the largest discontinuities can be consistently identified.

4. Test comparison and validation

a. Stationary cases

The power and size of both the more conventional difference series and the single-station time series tests were compared by plotting the percentage of H_0 rejections against the magnitude of a known, artificially imposed discontinuity. These artificial inhomogeneities were introduced by increasing or decreasing the extreme temperature exceedence threshold at a known point in the time series. For instance in a series with a +1°F (0.56°C) discontinuity, exceedence counts prior to the discontinuity are based on some threshold temperature T, while those after the discontinuity are based on the threshold $T + 1$°F. Given the nature of the original series (i.e., exceedence counts instead of mean temperatures) this approach was preferred over the more rigorous method of stochastically generating a homogeneous series to which artificial biases are imposed.

To assure that the series were homogeneous prior to testing, the series prior to the discontinuity were formed by resampling (with replacement) the exceedence counts (based on the threshold T) observed during a ≥20-yr period that was free of documented inhomogeneities. The series after the discontinuity reflected the resampling of the same ≥20-yr period, however in this case exceedences were based on the threshold $T \pm d$. To replicate the precision of extreme temperature observations in the United States, d was incremented by multiples of 1°F. However, to evaluate the test procedure for a more subtle discontinuity, a 0.5°F increment was also used. In this case, each annual count represented an average of exceedences based on T and $T \pm 1$°F. This fractional increment was chosen to simulate cases in which a station change resulted in a rounding bias, rather than a full 1°F discontinuity. The 0.5°F threshold increment was also assumed to be characteristic of cases in which a station move resulted in a discontinuity under some meteorological conditions (e.g., clear days), but not others (overcast conditions). However, it is likely that the majority of ''extreme'' days were characterized by similar synoptic conditions (Kalkstein et al. 1990).

For each integer value of d (-4°F $\leq d \leq 4$°F) and $d = \pm 0.5$°F, 1000 artificial time series were constructed. From these series, 1000 t_s values were computed, each of which was evaluated against the null hypothesis H_0: $t_s = 0$. Given this sample of 1000 H_0 evaluations, the frequency of H_0 rejections for known discontinuities was computed.

Figure 3 shows power curves for warm maximum and minimum temperature exceedences at four climatically and geographically diverse stations (Lockport, New York; Lake City, Florida; Williams, Arizona; and Winnibigoshish Dam, Minnesota). The curves show that for both tests (exceedence and difference series), H_0 is falsely rejected approximately 5% of the time in the absence of a discontinuity. This is expected, of course, since the test is conducted at the $\alpha = 0.05$ level. As discontinuities of increasing magnitude are imposed by tallying days greater than or equal to warmer or colder thresholds (e.g., $T \pm 1$°F, $T \pm 2$°F, etc.), the percentage of H_0 rejections shows a fairly symmetrical increase, except for warm maximum temperatures at Winnibigoshish Dam using the single-station test (Fig. 3d). In Fig. 3, the difference series tests (thick solid lines) correctly identify a larger percentage of the known discontinuities than the single-station tests (dotted lines). With the exception of warm maximum temperature extremes at Winnibigoshish Dam, both tests identify all of the ±4°F discontinuities.

In Fig. 3, separate power curves are also shown for difference series constructed using neighboring stations that minimize the resampled t_s confidence interval width, rather than s_p (thin solid lines). There is a clear tendency for the differences series tests based on the minimization of s_p to outperform those based on minimum confidence interval width. Similar comparisons (not shown) were also conducted using one more (or fewer) neighboring stations than was required to minimize s_p. In general, the most powerful results were associated with the original (minimized s_p) set of neighboring stations. However, the differences between the power curves tended to be subtle.

Similar power curves are obtained for cold maximum and minimum temperature extremes (days ≤ the 25th percentile). For brevity, four representative sets of cold-extreme curves are shown in Fig. 4. Here again, the difference series tests based on minimum s_p exhibit more power than those using confidence interval width or the single-station tests for most discontinuities. Only subtle differences were noted when one more (or fewer) neighboring stations than required to minimize s_p were used. At Lockport, the difference series test is able to identify a 1°F cold maximum temperature extreme discontinuity

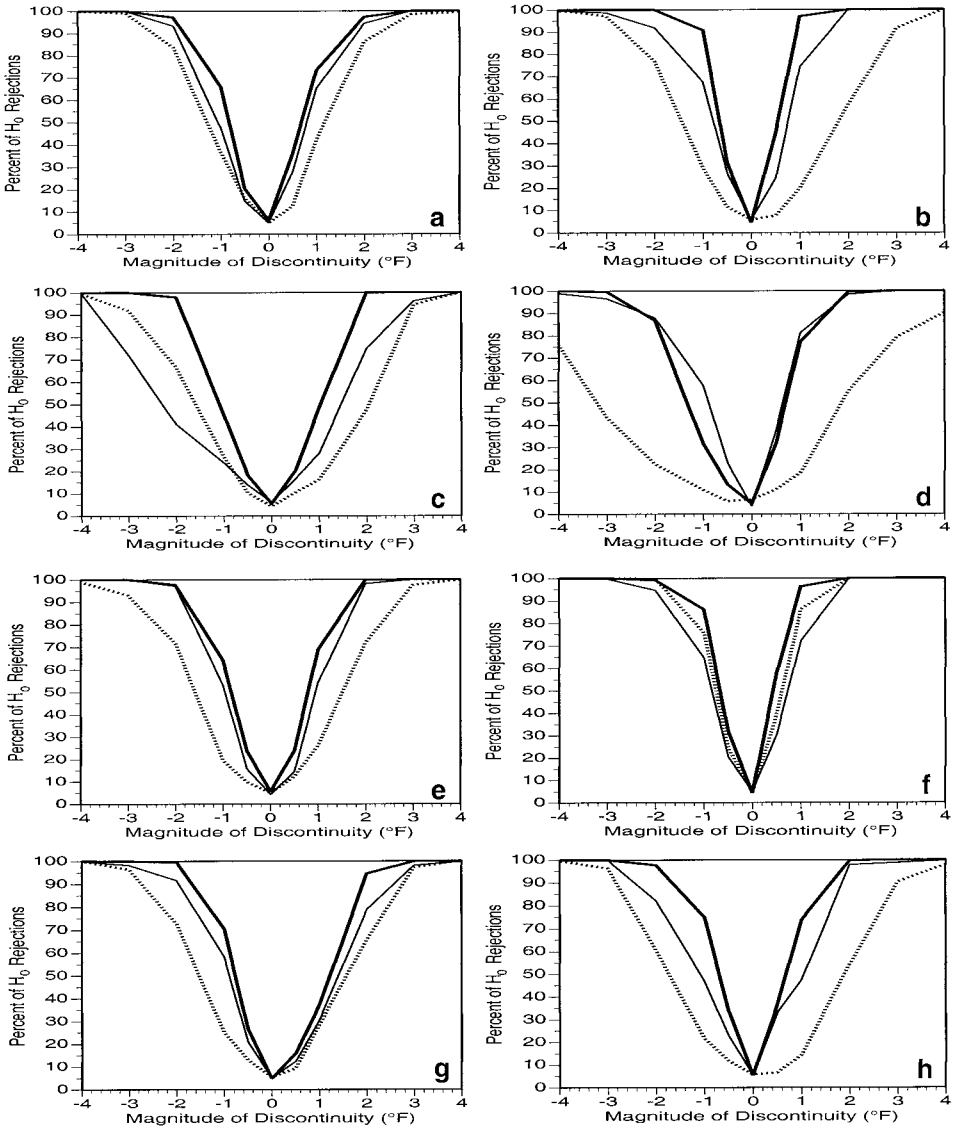

FIG. 3. Percent of H_0 ($t_s = 0$) rejections for series with imposed discontinuities of given magnitudes. The thick solid black curves are based on the difference series, while the dotted curves correspond to the exceedence series alone (i.e., a single-station test). The thin black curves correspond to difference series tests that minimize confidence interval width to select neighboring stations (i.e., akin to Karl and Williams 1987). Separate curves are given for warm maximum temperature exceedences at (a) Lockport, New York; (b) Lake City, Florida; (c) Williams, Arizona; and (d) Winnibigoshish Dam, Minnesota, and warm minimum temperature exceedences (e)–(h) at these stations, respectively.

72% of the time. This is in contrast to the results for Lake City, where the difference series test is only able to detect 33% of the −1°F inhomogeneities.

These differences (and similarities) in test performance can be attributed to the pooled standard deviations of the test series. Figure 5 shows that test performance (as measured by the percentage of 1°F discontinuities that are identified) is exponentially related to the s_p of the exceedence or difference series. Here curves describing the percentage of detected 1°F discontinuities as a function of pooled standard deviation are given for the individual stations. Each curve is fit to data from 20-, 30-, and 40-yr difference and exceedence count series for warm maximum temperatures. Although period of record (for series >20 yr) appears to have little effect on the power of the test, test performance appears to have some dependence on station location, particularly Lake City, Florida. At each individual station, the exponential relationship between s_p and inhomogeneity detection is fairly strong as the average R^2 value exceeds 74%. This strong relationship between test performance and s_p supports the minimization of pooled standard

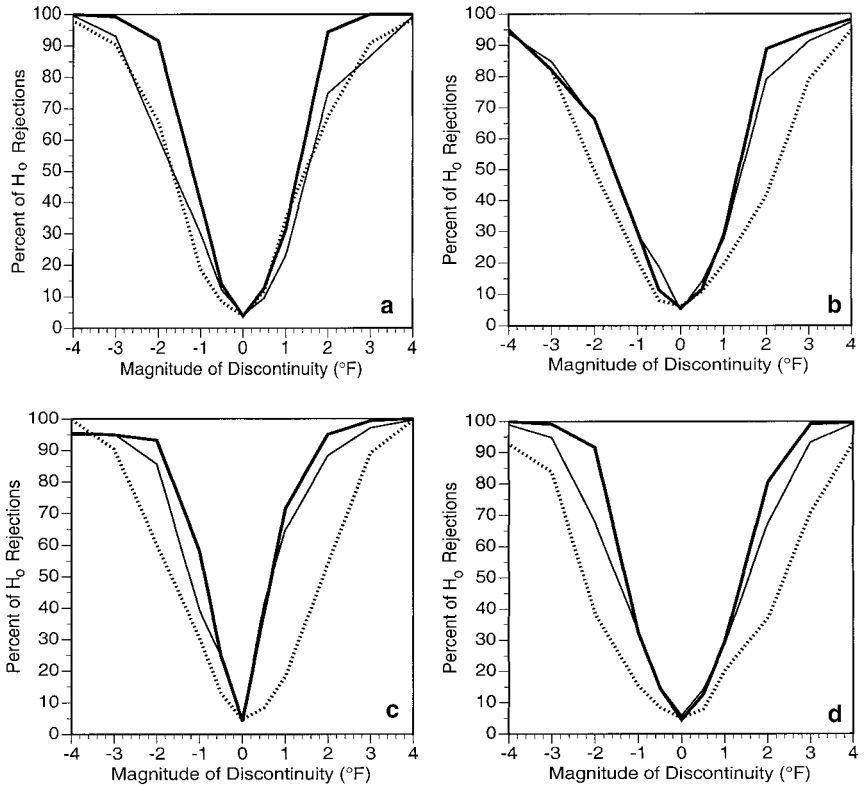

FIG. 4. As in Fig. 3, but for cold maximum temperature extremes at (a) Williams, Arizona, and (b) Winnibigoshish Dam, Minnesota, and cold minimum temperature extremes at (c) Lockport, New York, and (d) Lake City, Florida.

deviation as a criterion for selecting neighboring stations. Furthermore, the relationships given in Fig. 5 are not dependent upon test type. Thus when assessing an actual discontinuity at a suspect station with only distant neighboring reference stations, the relationship can be used to objectively choose the most powerful test.

Excluding Lake City, Florida, a single curve fit to the data from the other sites in Fig. 5 yields an exponential relationship with an R^2 of 72%. While this implies that the relationship between s_p and test power is relatively resilient to station location, the anomalous relationship at Lake City suggests that station location has some influence. This between-station difference in test performance is related to the shape of the tail of the daily temperature distribution. Figures 6a and 6b show the most common case where the percentage of days within ± 4°F of the 90th percentile daily maximum temperature either decreases slowly or remains relatively constant. At Lake City (Fig. 6c), however, the right tail of the daily temperature distribution shows a sharp decrease in the proportion of days with maximum temperatures >94°. Thus, a one degree change in the extreme threshold introduces a larger discontinuity (in terms of number of exceedences) at Lake City (Fig. 6c) than at Lockport (Fig 6a). Alternatively, the flat tail at Winnibigoshish Dam (Fig. 6b), results in a decrease in test power (Fig. 3d). Despite these differences, the general inverse relationship between s_p and test performance holds at each individual station.

Period length is another consideration for assessing the performance of the tests. Figure 7 illustrates a representative example of the deterioration of difference series test performance with decreasing subseries length. The power of the test for subseries (i.e., the periods before and after the discontinuity) with lengths of 40 and 10 yr is similar to that obtained with subseries lengths of 20 and 10 yr. Test power decreases as the 20-yr period is reduced to 10, and declines further for subseries lengths of 10 and 5 yr. However, even at these relatively short subseries lengths, the tests identify almost all of the ± 2°F discontinuities. Test performance declines more drastically when two 5-yr subseries are tested, suggesting that a total difference series length of 15 yr should be used as a minimum, with subseries based on at least 5 yr of data. This agrees with the limits suggested by Karl and Williams (1987).

Clearly Figs. 3 and 4 indicate that the difference series approach should be used to assess extreme temperature data series whenever possible. These figures also suggest that the single-station test is of some value for detecting and potentially adjusting inhomogeneities in extreme temperature series in the absence of an adequate network of reference stations. While it is unlikely that such a network would be lacking in the recent clima-

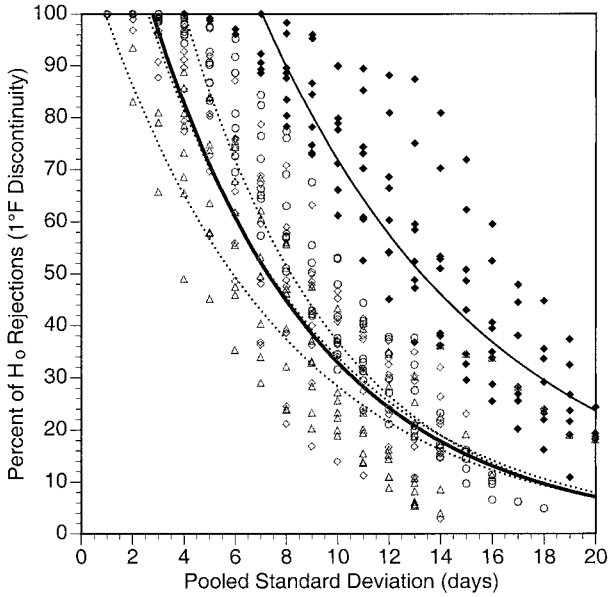

FIG. 5. Percent of H_0 rejections for cases with an imposed 1°F discontinuity vs the pooled standard deviation of the difference (or exceedence) series before and after the discontinuity at Lockport, New York (open diamonds), Williams, Arizona (open triangles), Winnibigoshish Dam, Minnesota (open circles), and Lake City, Florida (solid diamonds). Each curve is based on both exceedence count and difference series and represents a range of record lengths. The thick solid curve is an exponential fit to all the data shown by the open symbols (which exclude Lake City). The dotted and thin solid (Lake City) curves are exponential fits of the data at each individual station.

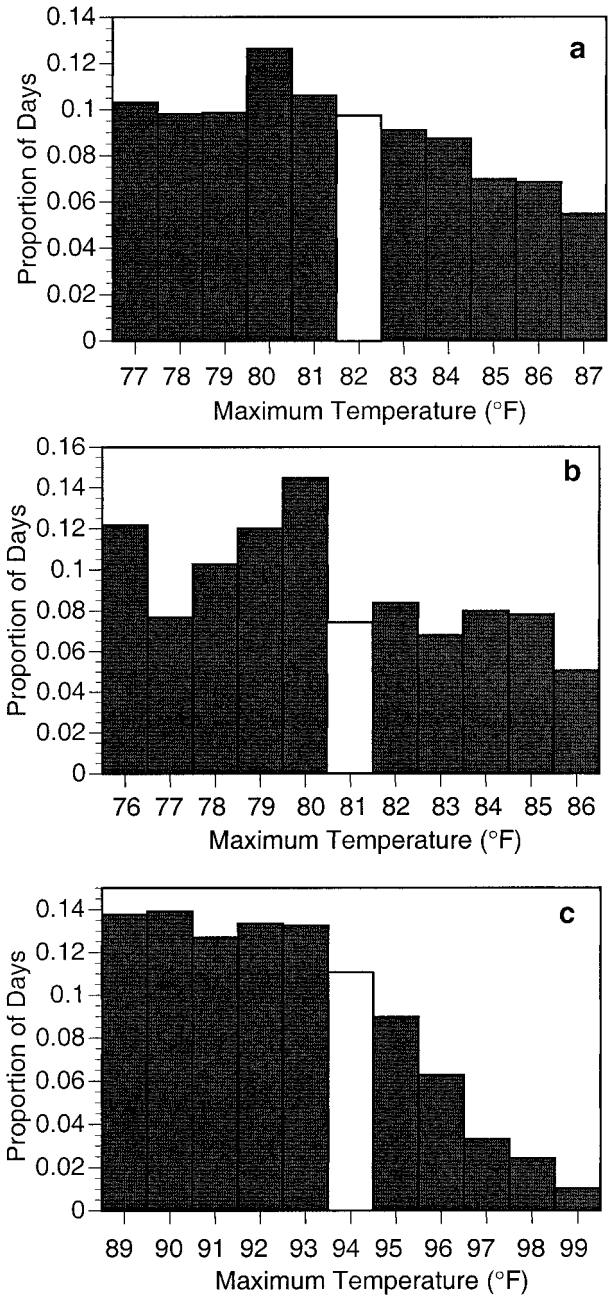

FIG. 6. Proportion of days with a maximum temperature within ±4°F of the 90th percentile daily maximum temperature (white bar) at (a) Lockport, (b) Winnibigoshish Dam, and (c) Lake City.

tological record, given the paucity of digitized daily observations prior to the mid-twentieth century it is probable that the use of the single-station approach would be necessary at those stations for which long (>50 yr) daily climatological records exist. Perhaps the most important feature of the single-station test in Figs. 3 and 4, is that, as expected, Type-I errors (the false rejections of H_0) occur in only 5% of the cases for both warm maximum and minimum temperature extremes. Thus, use of this test is not likely to lead to unwarranted adjustment of the temperature extreme series, except in the unlikely case that a natural climatic step change coincides with a documented nonclimatic discontinuity. Based on Fig. 3, the single-station test was able to detect, on average, 25% of the ±1°F discontinuities in warm maximum temperature extremes and 37% of the ±1°F inhomogeneities in warm minimum temperature extremes. Larger discontinuities (±2°F) on average were detected in 62% of the warm maximum temperature cases and 75% of the warm minimum temperature trials.

For cold temperature extremes, the single-station test on average identified 22% of the ±1°F cold maximum and minimum temperature extreme series discontinuities. Two degree discontinuities were identified in 48% and 56% of the extreme cold minimum and maximum temperature exceedence series, respectively. As with the warm extremes, false rejections of the null hypotheses associated with the cold extremes averaged 5%.

It should be pointed out that attempts to develop both difference series and single-station temperature extreme exceedence tests based on percentiles other than the quartiles were to no avail. Figure 8 compares the power curves for tests using the 50th (median), 90th, and 75th percentiles of the time series to compute t_s. Extreme warm maximum temperatures at Lockport are an ex-

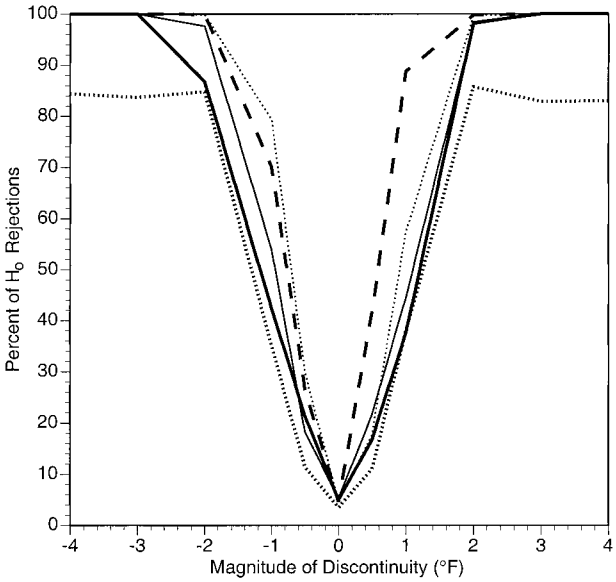

FIG. 7. Difference series test power curves using warm maximum temperature exceedences at Lockport for inhomogeneities preceded and followed by homogeneous periods of 10 and 40 yr (thick dashed); 10 and 20 yr (thin dotted); 10 and 10 yr (thin solid); 10 and 5 yr (thick solid); and 5 and 5 yr (thick dotted).

ample. The use of the median produces a test similar to that of Karl and Williams (1987). In Fig. 8, the median-based test (dashed curve) is clearly outperformed by the tests based on the higher percentiles, particularly when positive discontinuities are imposed (i.e., cooling). The 90th percentile produces a curve similar to that for the 75th percentile, although with some loss of power. This suggests that a warming or cooling of temperature extremes results in a skewing of the distribution of annual exceedence counts, rather than a translation of the mean.

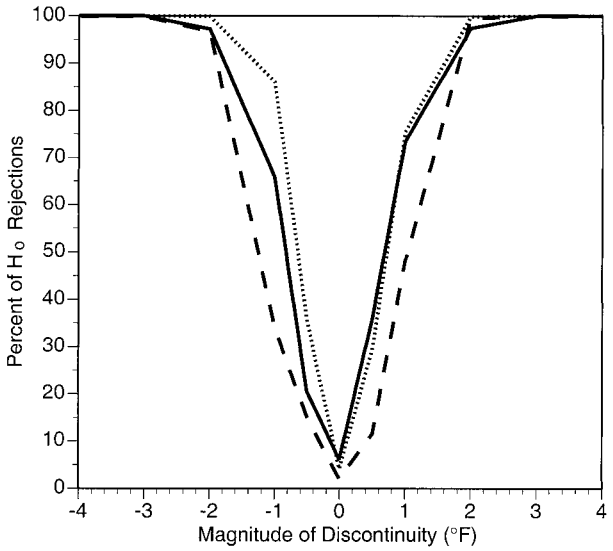

FIG. 8. Single-station test power curves for warm maximum temperature exceedences at Lockport using the 25th and 75th (solid), 90th and 10th (dotted), and 50th (dashed) percentiles to compute t_s.

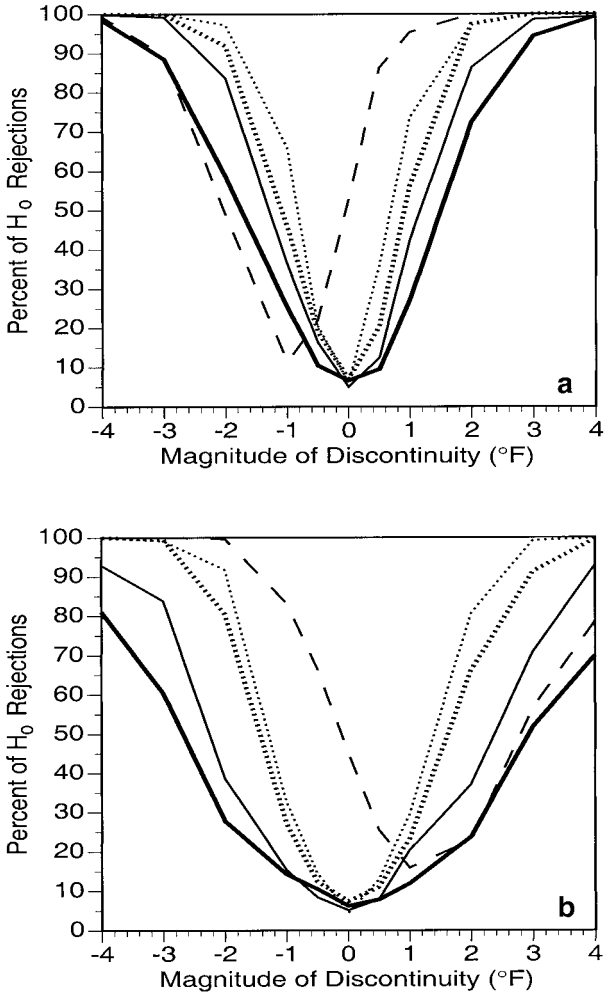

FIG. 9. Comparison of power curves associated with stationary (thin dotted) and nonstationary (heavy dotted) difference series and stationary (thin solid) and nonstationary (heavy solid) single-station exceedence series tests. The dashed curve represents the effect of ignoring the nonstationarity in the difference series. The curves are based on (a) warm maximum temperature exceedences at Lockport and (b) cold minimum temperature exceedences at Lake City, Florida.

b. Nonstationary tests

Representative power curves for cases in which a significant slope exists in one of the exceedence or difference series (i.e., either the series before or after the inhomogeneity) are shown in Fig. 9. For reference, the power curves for analogous cases in which a slope was not imposed are also shown. Similarly, the power curves that result if the presence of the significant slope is ignored are given as well. At both stations a general decrease in power is associated with the nonstationary cases, resulting from the uncertainly introduced by the presence of the time-dependent trend. At Lockport (Fig.

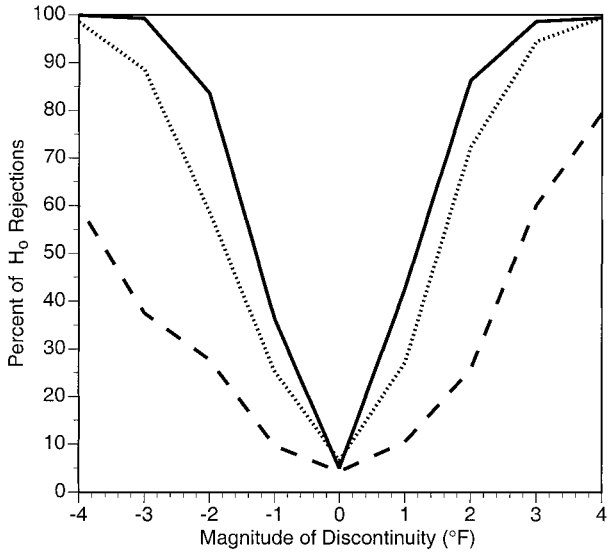

FIG. 10. Comparison of power curves for series with significant slopes before and after the discontinuity (dashed), a single significant slope (dotted) and no significant trends (solid). Each curve is based on warm maximum temperature exceedences at Lockport.

9a), using the difference series (dotted) curves, 70% of the ±1°F warm maximum temperature discontinuities are detected on average in the stationary case, whereas only 51% of these discontinuities are identified in the nonstationary example. A similar decrease is associated with the single-station exceedence series (solid) curves. For cold minimum temperature exceedences at Lake City (Fig. 9b), both the stationary and nonstationary difference series tests perform poorly, identifying only 29% of the ±1°F discontinuities on average. For the single-station series, slightly more ±1°F discontinuities are identified in the stationary case. Here again, however, the performance of both tests is poor.

At both stations the consequences of ignoring a nonstationarity in one of the difference or exceedence series is evident. Here the existence of a significant slope translates the power curve to the left or right depending on the sign of the slope. This translation results in a substantial increase in the number of false H_0 rejections, as well as an asymmetric decrease in the number of actual discontinuities that are detected.

The power of both tests decreases further when both series (before and after the discontinuity) are nonstationary. Using warm maximum temperature exceedences at Lockport as an example, Fig. 10 indicates that even for discontinuities as large as 4°F, the procedure is only able to identify between 60% and 80% of the inhomogeneities. Fortunately the occurrence of two successive nonstationary periods (based on documented station changes) in the U.S. HCN database is rare.

5. Series adjustment

Once the existence of an extreme temperature discontinuity is identified, it is necessary to formulate an adjustment factor, consistent with the results of either the difference series or single-station exceedence series test. 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. Thus 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 ≥90°F are considered ''extreme,'' introduces a 4°F warming to the subsequent record of daily temperatures. Such a change would precipitate an increase in days ≥90°F, since days on which the temperature would have previously (before the move) only reached 86°F are now likely meet the ≥90°F threshold. In such a case, adjustment would involve selecting a new higher threshold such that the number of exceedences of this new limit is comparable to that associated with the original 90°F value.

Determining this new threshold involves an array of tests in which the series following the discontinuity is based on sequentially higher- or lower-threshold values. Progressively higher thresholds are indicated when the inhomogeneity is followed by an increase in warm exceedences or a decrease in cold exceedences. Based on the above example, assume that the 4°F warming results in the rejection of H_0 when the series before and after the move are based on a 90°F threshold. Since such a warming would lead to an increase in days ≥90°F, the series after the break is recomputed based on days ≥91°F and the test reapplied using days ≥90°F prior to the move and the ≥91°F series after the break. Assuming the null hypothesis is again rejected, the test would be repeated using counts of days ≥92°F after the break. This process of increasing the threshold temperature and retesting proceeds until H_0 is accepted and then continues until the number of exceedences following the break is significantly less than that based on the original 90°F threshold. This suite of tests generally produces a string of one or more threshold values for which no discontinuity is indicated, one of which corresponds to the most appropriate adjustment. Given the symmetry of the power curves (i.e., Figs. 3, 4) along with the marginal power of the inhomogeneity tests for ±1°F discontinuities, the median of the thresholds that result in acceptance of the null hypothesis is chosen as the adjustment in cases where more than one adjustment results in acceptance of H_0. If the set of tests fails to give a threshold for which H_0 is accepted, the exceedences are adjusted based on the average of the thresholds that change the sign of t_s.

Figure 11 provides further empirical justification for the use of the median threshold. Here +4°F, and sep-

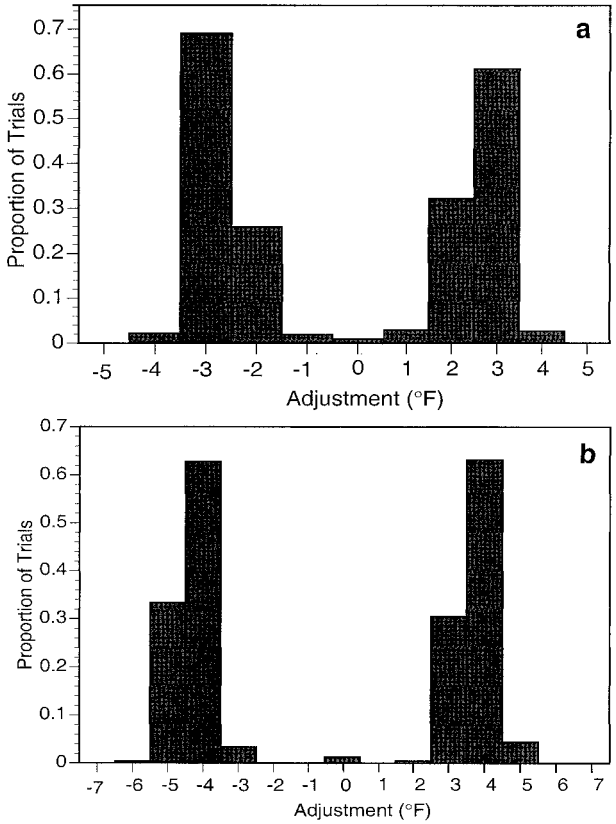

FIG. 11. Adjustments indicated for series with known −4°F (left side) and +4°F (right side) discontinuities imposed using (a) the first threshold for which H_0 is accepted and (b) the median threshold.

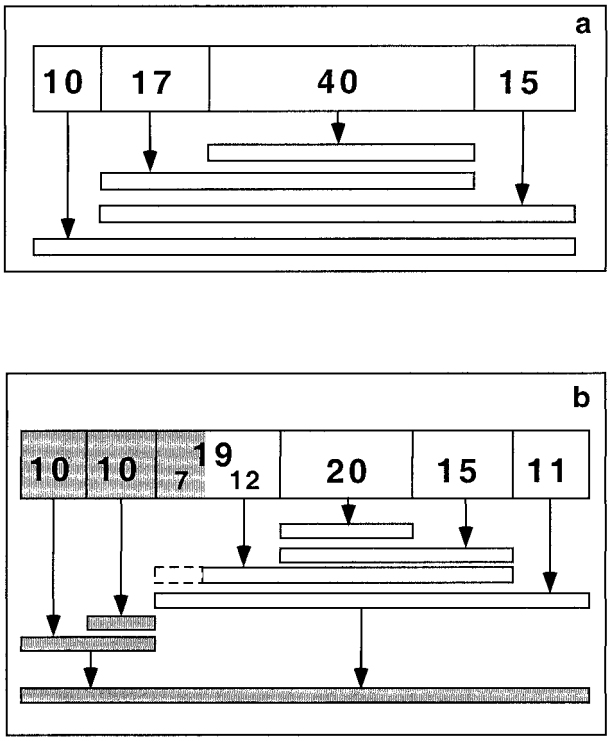

FIG. 12. Illustration of the order of adjustment for series having more than one discontinuity. The wide bars represent documented breaks in the time series, with the numbers indicating the length of each subseries. The length of each arrow corresponds to adjustment order (shortest first), with the combined adjusted series indicated by the narrow bars. (a) The entire series is represented by differences or exceedence counts. (b) The shaded areas are portions of the series in which reference stations could not be identified and thus adjustment was based on the single-station method.

arately −4°F, discontinuities were introduced into 1000 otherwise homogeneous exceedence count time series. Using the above suite of tests, 1000 adjustment factors were obtained and evaluated against the known ±4°F discontinuities. Figure 11a shows that basing the adjustment on the threshold associated with the first H_0 acceptance results in an underestimate of the true discontinuity. In these cases, an adjustment less than the correct 4°F adjustment is indicated in over 95% of the trials. When the threshold corresponding to the median of those thresholds for which H_0 is accepted is considered as the adjustment, the proper ±4°F adjustment is selected most frequently. Unfortunately the use of the median threshold leads to cases in which a fractional adjustment is indicated. In these instances, adjusted annual counts are based on the average exceedences for the whole thresholds bracketing the fractional median value.

A final consideration for the adjustment procedure relates to the order in which adjustments are made in series experiencing more than one discontinuity. Karl and Williams (1987) adopted a reverse chronological approach in which the most recent homogeneous period was used as a standard and earlier periods were adjusted to reflect these current conditions. It can be argued that this is the preferential order for applying adjustments since in this manner the historical record becomes homogeneous with the most recent (and presumably ongoing) climatological record.

Despite this advantage, it is also reasonable to base adjustments on the longest stationary homogeneous period within a station's record. This approach minimizes the quantity of data that is subject to adjustment, while maximizing the ability of the test procedure to detect small discontinuities. Based on Fig. 7, the difference series test is able to detect a higher percentage of artificial discontinuities as the length of one of the homogeneous periods increases. This also corresponds with the data presented in Fig. 5, since generally pooled standard deviation decreases with record length. Thus, this approach was adopted in this study. Once this long standard period has been identified, adjustments proceed chronologically with the decision to adjust earlier or more recent periods again based on series length. Once adjusted, sequential series are combined to evaluate and potentially adjust later (or earlier) segments of the series. This process is illustrated in Fig. 12a. Despite the use of this approach here, the test and adjustment procedure

can be applied in a reverse chronological fashion if desired.

While this approach is fairly straightforward when the overall data record is represented by a reference-station difference or single-station exceedence series, in practice the adjustment of most long climatological records will require the use of both types of series. In these cases, the periods requiring use of the single-station test (either due to lack of reference stations or perhaps minimization of the pooled standard deviation of the series) and those for which the difference series test can be applied are treated separately. Thus, three distinct periods, one requiring the single-station test, another based on the difference series approach, and a third intervening period, will be present (Fig. 12b). If a difference series of 5 or more yr can be formed within the intervening period, then an adjustment is computed based on these years and applied to each year within the intervening period. Otherwise the adjustment applied to the intervening period is based on the single-station approach. The two final homogeneous periods that result (one standardized with the single-station test, the other using a difference series) are adjusted using the single-station approach. If an adjustment is indicated, it is applied to the single-station series, regardless of length.

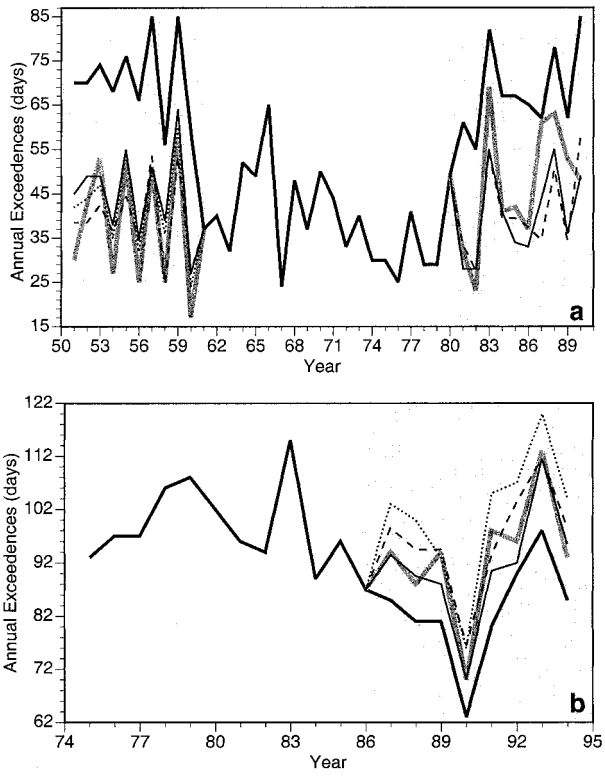

FIG. 13. Comparison of formulated inhomogeneous temperature exceedence series (thick black solid) with the actual homogeneous series (solid gray) and the difference series (thin solid), single-station (dotted), and mean translation (dashed) adjusted series of (a) days with maximum temperature ≥80°F at Mohonk Lake, New York, and (b) days with maximum temperature ≤65°F at Calhoun Research Station, Louisiana. The imposed station discontinuities are indicated by the change in shading.

6. Adjustment examples

To further evaluate the combined testing and adjustment procedure, these methods were applied to stations at which a move was simulated. In these cases, station relocations were imposed by substituting the observed record of threshold exceedences at a neighboring station. Therefore, these examples represent the exaggerated relocation of a station to the site of an existing neighboring station, that was not subsequently considered as a reference station. Although actual station moves are considerably more subtle, these simulations allowed the adjusted temperature extreme series to be compared with the series that would have resulted without the relocation (i.e., the series at the original station). The contrived relocations were also useful to illustrate the benefits of the iterative threshold adjustment procedure over previous techniques that rely on translation of the series mean based on the difference (Karl and Williams 1987) or ratio (Alexandersson 1986) of the series before and after the discontinuity. Additional examples are also given for actual station changes. These correspond to cases presented by Karl and Williams (1987) and Easterling and Peterson (1995).

a. Contrived relocations

Figure 13 shows two cases in which a station move was fabricated by substituting the threshold exceedence series of a neighboring station for portions of an otherwise homogeneous series at the original site. In Fig. 13a, days ≥80°F at Mohonk Lake, New York, and Charlotteburg, New Jersey (approximately 85 km to the south), form the inhomogeneous series. Data from Charlotteburg are used prior to 1961 and after 1980, while the Mohonk Lake values form the 1961–80 base period. This series was then adjusted using the single-station exceedence count and differences series tests separately. The adjusted series were then compared to the observed homogeneous series at Mohonk Lake, as well as the series adjusted using the difference of the two subseries means.

Using the difference series test, annual exceedences of 83°F are tallied during the pre-1961 period to adjust the series. After 1980, a 83.5°F threshold is used. These higher thresholds reduce the number of exceedences thus counteracting the artificial warming introduced by the move to Charlotteburg. The single-station test applies a slightly greater adjustment in the earlier period based on its use of a 83.5°F extreme threshold for both subseries. Using the difference between the means of the two subseries as an adjustment, each observation in the earlier subseries is reduced by 32 exceedences, while those in the later period are reduced by 28 exceedences.

Overall, each of the methods tracks the actual Mohonk Lake series fairly well. However, during individual years, fairly large discrepancies exist between the adjusted and original series, owing to the exaggerated distance of the move. It is during these years, which are particularly evident in the early 1950s (Fig. 13a), that the threshold-based adjustments tend to be superior to those based on the difference (or ratio) of the series means. This increase in performance can be attributed to a tendency for the difference between the Mohonk Lake and Charlotteburg series to decrease as the number of exceedences increases. During warmer (more exceedence) years, many days exceeding the original 80°F threshold would have also exceeded the new 83.5°F threshold, thus limiting the effect of the adjustment. Whereas in cooler years, fewer days are likely to have attained the new higher threshold, increasing the adjustment. These variable adjustments can be better simulated using the threshold method as opposed to a constant shift of the mean.

Numerically, the median error (actual − adjusted) based on the threshold adjustment method of −0.5 day yr^{-1} exhibits little bias, while translation of the series mean gives a slightly larger 1.0 day yr^{-1} median error. Median absolute errors corresponding to the threshold and mean adjustments are 7.0 and 8.0 day yr^{-1}, respectively.

A second formulated station relocation is based on days ≤65°F (25th percentile daily maximum temperature) at Calhoun Research Station, Louisiana (Fig. 13b). Data from Alexandria, Louisiana, were substituted for those at Calhoun after 1986 to simulate an exaggerated station move (Alexandria is 129 km to the south of Calhoun). Since Alexandria experiences fewer days with a maximum temperature ≤65°F, warmer thresholds are indicated by the difference series (66.5°F) and single-station (67.5°F) tests. Similarly, the mean of the exceedence series is 13.5 day yr^{-1} higher prior to the discontinuity.

Again, each adjusted series tracks the actual Calhoun series quite well. However, the mean adjusted series (and the single-station threshold adjustment) consistently overestimates the actual number of cool maximum temperature occurrences. Median errors (adjusted − actual) of −2.5 and −0.8 day yr^{-1} result from the mean and difference series threshold adjustments, respectively, again indicating some bias in the mean adjustment method. The median absolute error of 3.0 day yr^{-1} for the threshold-based adjustment compares favorably with the 6.0 day yr^{-1} value for the mean adjusted series. Again, the distance between Calhoun and Alexandria in this formulated example represents a station move that is at least an order of magnitude greater than that of an actual station relocation.

b. Actual examples

Karl and Williams (1987) showed that a subtle 300-m change in instrument siting during 1960 resulted in a dramatic decrease in minimum temperature at Setauket, New York. This was the only documented discontinuity at Setauket since 1885. Figure 14a shows that this 1960 relocation was also followed by a marked decrease in days with minimum temperatures ≥66°F. This discontinuity was deemed significant by both tests, with a revised threshold of 63.5°F indicated by the difference series test and a slightly higher 64°F threshold given by the single-station test. Visually, these adjustments produce a more homogeneous time series at Setauket, despite the magnitude of the extreme correction being several degrees lower than that indicated for summer mean minimum temperature (Karl and Williams 1987).

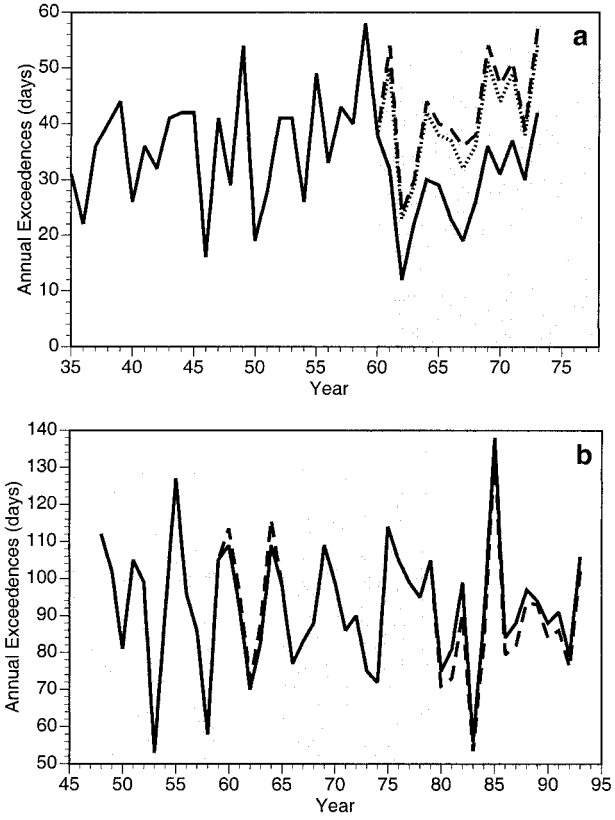

FIG. 14. Observed (solid) and adjusted (dashed for difference series approach, dotted for single-station method) exceedence series for (a) days with minimum temperature ≥66°F at Setauket, New York, and (b) days with minimum temperature ≤28°F at Spokane, Washington. Station discontinuities are indicated by the change in shading.

Easterling and Peterson provide a second example of an actual station discontinuity. In this case a new thermometer was installed at Spokane, Washington, in 1959. Following this instrument change the observation location was moved 1.1 km in 1965, and relocated 0.5 km again in 1979. Easterling and Peterson analyzed only the 1959 instrument change and found that this discontinuity introduced a 0.61°C (1.1°F) cooling to the mean annual temperature record.

When this series was analyzed for changes in the

number of days with a minimum temperature ≤28°F (the 25th percentile minimum temperature), the single-station test did not detect discontinuities for any of the three moves. Use of the difference series approach, however, indicated that the threshold should be lowered by 0.5°F to 27.5°F during the more recent 1979–93 periods (Fig. 14b). Subsequent tests suggested that the threshold for the 1959–65 period should be raised to 28.5°F and that no change from the 28°F threshold was necessary for the earliest subseries. Clearly, the differences introduced by the station changes at Spokane are more subtle than that which occurred at Setauket. Nonetheless the difference series approach was able to detect a discontinuity associated with the 0.61°C change in mean annual temperature reported by Easterling and Peterson.

7. Summary

The approach of Karl and Williams (1987) for adjusting discontinuities in mean temperature series is modified for use with extreme temperature exceedence series. In addition, two alternative procedures are developed. The first method relies only on the data series at the inhomogeneous station and thus provides a means of adjusting temperature extreme discontinuities in cases where an adequate network of nearby homogeneous reference stations is unavailable. The second procedure is used in those instances when either the single-station exceedence or difference series are nonstationary.

Overall the performance of the difference series approach is superior to that of the single-station method for detecting imposed discontinuities. However, the single-station method exhibits sufficient precision to warrant its cautious use in cases where homogeneous reference series are absent. In fact, adjustments based on the two methods are quite similar, with neither method consistently producing the lowest adjustment errors. Neither test has a propensity to incorrectly adjust randomly generated homogeneous time series. Inhomogeneities are also detected with reliable precision in nonstationary cases by both tests, provided only one subseries (i.e., either the period before or after the discontinuity) exhibits a significant trend. Only relatively large discontinuities could be identified when both periods were nonstationary. However, even in this case, false rejections of the null hypothesis (i.e., no difference in the exceedence series before and after the break) do not occur more frequently than is indicated by the size of the test. Despite this favorable result, it should be noted that natural climate step changes (as opposed to changes in trend) that coincide with the documented inhomogeneity would be considered nonclimatic and adjusted by the single-station procedure. Since neighboring stations would likely exhibit the same climate-induced step change, the difference test would not indicate adjustment in this unlikely case.

Clearly, limitations in data availability and station density preclude the exclusive use of difference series tests with long-term climatological series that are based on daily data. The suite of tests developed in this study provides a method that can be used when such restrictions are present. Although in itself the difference series approach provides a more desirable test, overall the single-station and difference series approaches complement each other providing an improved method for homogenizing extreme temperature exceedence series. Given the ability of these methods to produce homogeneous series of temperature extreme exceedences, it is conceivable that this approach could be used to homogenize long-term daily datasets. While a formal evaluation of this use is beyond the scope of our present work, we plan to further investigate this application. Presumably by computing separate adjustment thresholds for different monthly or seasonal extreme percentiles, it would be possible to develop different adjustments for individual temperature ranges comprising the entire daily temperature distribution.

Acknowledgments. This work was supported by the NOAA–NASA Grant NA76GP0351 and NOAA Grant NA46WP-0227. A prerelease version of the daily HCN data were obtained through the generosity of Dale Kaiser of the Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory. The comments of two anonymous reviewers contributed to the clarity and improvement of our manuscript.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.,* **6,** 661–675.

Draper, N. R., and H. Smith, 1981: *Applied Regression Analysis.* Wiley, 709 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, precipitation, and snow data for 1871–1997. Oak Ridge National Laboratory Rep. ORNL/CDIAC-118, NDP-070, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 241 pp. [Available from Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831.]

Houghton, J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., 1996: *Climate Change 1995: The Science of Climate Change.* Cambridge University Press, 572 pp.

Kalkstein, L. S., P. C. Dunne, and R. S. Vose, 1990: Detection of climate change in the western North American Arctic using a synoptic climatological approach. *J. Climate,* **3,** 1153–1167.

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 F. T. Quinlan, 1990: United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data. Oak Ridge National Laboratory Rep. ORNL/CDIAC-30, NDP-019/R1, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 274 pp. [Available from Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831.]

Nelson, W. L., R. F. Dale, and L. A. Schaal, 1979: Non-climatic trends

in divisional and state mean temperatures: A case study in Indiana. *J. Appl. Meteor.,* **18,** 750–760.

Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol.,* **14,** 671–679.

——, and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.,* **18,** 1493–1517.

Tuomenvirta, H., and H. Alexandersson, 1995: Adjustment of apparent changes in variability of temperature times series. *Proc. Sixth Int. Meeting on Statistical Climatology,* Galway, Ireland, Steering Committee for International Meetings on Statistical Climatology, 443–446.

Vincent, L. A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate,* **11,** 1094–1104.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 464 pp.