An Approach to Adjusting Climatological Time Series for Discontinuous Inhomogeneities

THOMAS R. KARL AND CLAUDE N. WILLIAMS, JR.

National Climatic Data Center, National Environmental Satellite Data and Information Service, National Oceanic & Atmospheric Administration, Asheville, NC 28801

(Manuscript received 3 April 1987, in final form 25 June 1987)

ABSTRACT

A method is described whereby climatological time series of temperature and precipitation can be adjusted for station inhomogeneities using station history information. The adjusted data retains its original scale and is not an anomaly series. The methodology uses the concepts of relative homogeneity and standard parametric (temperature) and nonparametric (precipitation) statistics. The technique has been tested in Monte Carlo simulations, and is shown to produce climatological time series more consistent with the concept of a homogeneous climate record than would otherwise be the case. Additionally, the technique provides an estimate of the confidence interval associated with each adjustment. It has been applied to over 1200 stations in the United States. In many instances the adjustments in temperature time series are substantial (as large as actual climate fluctuations during the twentieth century) often leading to a more consistent pattern of regional climate change than would otherwise be surmised from inspection of the unadjusted data.

1. Introduction

In recent years there has been a proliferation of climate analyses focusing on the problems of climate change. Many investigators have examined long term changes in temperature and precipitation (Jones et al., 1986; Diaz and Quayle, 1980; Wigley et al., 1985; Barnett, 1985; Williams and van Loon, 1976; van Loon and Williams, 1976a, 1976b; etc.) with increasing interest in the potential impacts of inadvertent man-made modification of the climate, e.g., increased concentration of greenhouse gases. Other studies have addressed the specific question of climate variability (Karl et al., 1984a; Karl, 1985; Agee, 1982; Thompson, 1975; Vinning and Griffiths, 1985; Brinkman, 1983) using long periods of record. The modern historical record has also been used by dendrochronologists and paleoclimatologists to calibrate proxy data such as tree ring growth, pollen and marine plankton (Blasing and Duvick, 1984; Fritts et al., 1971; Webb and Wigley, 1985, etc.). Even the diagnostic studies of aperiodic events such as El Niño/Southern Oscillation, sunspot frequency, volcanic dust loading of the atmosphere, etc., often rely on a long climate record to draw inferences with respect to their impact on the climate.

The value of all these studies is no better than the data used to represent the modern historical climate record. Since this is the building block for so much work it is essential that the climatic record is represented by the best sources of data with a minimum of errors and biases. Both recent and not so recent articles suggest that this is often not the situation in many datasets (Karl et al., 1984b, 1986; Karl, 1985; Kukla et al., 1986; Schaal and Dale, 1977; Mitchell, 1953).

As general circulation models (GCM) continue to improve and provide more reliable regional information regarding the expected greenhouse warming (Schlesinger and Mitchell, 1985) it is inevitable that more interest will focus on regional climate and regional climate change. In this regard it is very important that the climate record be as free of station discontinuities and inhomogeneities (due to station changes and observation methods) as possible. Unlike hemispheric or global scale studies of climate change, the number of long-term stations on a regional basis, e.g., 2° latitude by 2° longitude, 50 000 km^2, agricultural reporting districts, climate divisions, etc., is much smaller by at least two orders of magnitude. Inhomogeneities that tend to offset each other when many stations are used, are much less likely to do so when smaller station networks are used to detect regional climate change. This paper focuses on the development and application of a procedure to systematically adjust individual time series of monthly temperature and precipitation for discontinuities due to nonclimate effects, i.e., station relocations, new instruments, changes in instrument heights, changes in observation schedules, etc.

2. Preliminary concepts

a. *Homogeneity*

Conrad and Pollak (1950) provide an exact definition of relative homogeneity: *"A climatological series* *is relatively homogeneous with respect to a synchronous series at another place if the temperature differences (precipitation ratios) of pairs of homologous averages constitute a series of random numbers that satisfies the law of errors."* The underlying assumption behind such a definition is that variations in average weather have similar tendencies over rather large regions. For example, a cold winter in Massachusetts is usually accompanied by a cold winter in New Hampshire, and a very dry winter in central California usually occurs when the winter is very dry in northern California. Exceptions to these tendencies are not systematic in time, but rather are randomly distributed in time. Acceptance of the definition of relative homogeneity reduces the problem of determining whether two series are relatively homogeneous to one of a selection of appropriate inferential statistical methods. Conrad and Pollak (1950) also use the terminology absolute homogeneity if a third station is used and the differences (or ratios) of the three pairs of stations are all random. The goal of the work presented here is not, however, to identify homogeneous or inhomogeneous stations, but rather to use the concept of differences and ratios to adjust the climate record for discontinuous nonclimate biases which may have entered the climate record.

b. The use of differences or ratios

There are two important characteristics associated with the effective use of the method of differences (or ratios). Neighboring stations can detect discontinuities at a specific candidate station (a station which may require adjustments due to nonclimate discontinuities) more readily when the correlation of monthly or seasonal anomalies of temperature or precipitation between the candidate and its neighbors are high and the year-to-year variances of the anomalies are small. Table 1 demonstrates this concept using three stations in the midwestern United States. Des Moines, Iowa (DSM) is used as the candidate (CAN) station and Grand Island, Nebraska (GRI) and Rockford, Illinois (RFD) as its neighbors. The observed correlation of July anomalies during the period 1961 through 1970 between DSM and GRI is 0.81 and between RFD and DSM, 0.84. The variance at GRI is twice that of RFD. It is rather obvious that the higher the correlation between the candidate and its neighbor, the easier it will be to detect a discontinuity, but it is less apparent that the variance also plays an important role. In order to illustrate this, the correlation between RFD and DSM is made identical to that between GRI and DSM, but the observed variance and mean is retained at RFD. The normalized anomalies of July temperatures at RFD are set equal to those at GRI by making use of the Z-score defined as

$$Z_{\mathrm{GRI},i} = (T_{\mathrm{GRI},i} - \bar{T}_{\mathrm{GRI}})/s_{\mathrm{GRI}} \qquad (1)$$

where T_{GRI} is the monthly mean for year i at GRI and $\bar{T}_{\mathrm{GRI}}$ is the 10 year mean and s is the standard deviation at GRI. The value of $T_{\mathrm{RFD},i}$ is found by

$$T_{\mathrm{RFD},i} = Z_{\mathrm{GRI},i}s_{\mathrm{RFD}} + \bar{T}_{\mathrm{RFD}}. \qquad (2)$$

TABLE 1. The mean July temperature at Des Moines, Iowa (DSM) and two neighboring stations with identical normalized departures from the mean (Z-scores) and correlation with DSM (correlation = 0.81), but unlike year-to-year variances. A discontinuity (+0.5°C) is introduced into the record DSM between 1965 and 1966; RFD is the station identifier for Rockford, Illinois and GRI for Grand Island, Nebraska.

	Temperature (°C)								
	Monthly means			Differences		+0.5 Discontinuity differences		−0.5 Discontinuity differences	
Year	DSM	GRI	RFD	GRI-DSM	RFD-DSM	GRI-DSM	RFD-DSM	GRI-DSM	RFD-DSM
1961	23.37	24.92	22.45	1.55	−0.92	1.55	−0.92	1.55	−0.92
1962	23.48	22.14	20.51	−1.34	−2.97	−1.34	−2.97	−1.34	−2.97
1963	24.36	25.39	22.78	1.03	−1.58	1.03	−1.58	1.03	−1.58
1964	25.11	26.33	23.46	1.22	−1.65	1.22	−1.65	1.22	−1.65
1965	23.75	23.75	21.65	0.00	−2.10	0.00	−2.10	0.00	−2.10
Mean_1	24.01	24.51	22.17	0.49	−1.84	0.49	−1.84	0.49	−1.84
1966	25.89	27.22	24.08	1.33	−1.81	0.83	−2.31	1.83	−1.31
1967	22.26	23.78	21.65	1.52	−0.61	1.02	−1.11	2.02	−0.11
1968	23.42	24.00	21.81	0.58	−1.61	0.08	−2.11	1.08	−1.11
1969	24.42	25.11	22.59	0.69	−1.83	0.19	−2.33	1.19	−1.33
1970	24.78	25.83	23.10	1.05	−1.68	0.55	−2.18	1.55	−1.18
Mean_2	24.15	25.19	22.63	1.03	−1.51	0.53	−2.01	1.53	−1.01
Mean	24.08	24.85	22.41	0.76	−1.68	0.51	−1.93	1.01	−1.43
Variance	1.08	2.17	1.07	0.77	0.40	0.69	0.38	0.99	0.56
$\mathrm{Mean}_1 - \mathrm{mean}_2$	−0.14	−0.68	−0.48	−0.54	−0.33	−0.04	0.17	−1.04	−0.83

Next, the differences $T_{GRI,i} - T_{DSM,i}$ and $T_{RFD,i} - T_{DSM,i}$ are calculated, and a discontinuity (either +0.5° or −0.5°C) is introduced at DSM between the years 1965 and 1966. Because of the smaller variance at RFD compared to GRI, the variance of the difference series RFD–DSM is smaller than that at GRI–DSM. As shown in Table 1, the differences of the mean of the differences between RFD and DSM are closer to estimating the magnitude of both the positive and negative discontinuity introduced at DSM than those between GRI and DSM. This effect can be much larger when the variance of two neighboring stations is quite different, i.e., maritime versus continental. Furthermore, since the variance is usually substantially greater during winter compared to summer (differences substantially greater than those depicted in Table 1), consideration of a station's variance is quite important in any objective method which establishes general decision rules for assessing and adjusting for the magnitude and sign of a discontinuity.

Actually, if the mean temperatures at DSM (Table 1) in the first 5 years versus the second 5 years had been used to estimate the discontinuity, the differences between these periods would have closely matched the actual discontinuity. Despite this circumstance, such an approach cannot be used in practice because it precludes any possibility of climate variability and change.

For precipitation, ratios between the candidate (CAN) and a neighbor (NEIGH) are used instead of differences. Actually, the logarithm of the ratios are more physically meaningful quantities to use compared to the ratios themselves. For example, the ratios 0.50 and 2.00 are 0.5 and 1.0 units away from unity, but the logarithm of these values produce values equally distant from 0 (the logarithm of 1.0). This is desirable in the sense that a CAN with twice as much precipitation compared to its NEIGH is treated exactly opposite to the situation when it has only one-half of its NEIGH's precipitation.

c. Seasonal (or annual) versus monthly adjustments

The question arises regarding the appropriate month or month(s) over which the difference or ratio series should be calculated. Mitchell (1961) indicates that it is often difficult with respect to temperature to detect much in the way of changes of the difference series between seasons, and for precipitation it is often difficult to justify ratios for more than a few seasons, i.e., rainfall predominantly derived from convective versus stratiform precipitation. Additionally, it is usually advantageous to use a difference series or ratio series over a season (or seasons) compared to a single month since this usually reduces the variability of the difference series, and provides a better chance of detecting and adjusting for small inhomogeneities. On the other hand, if there is a good reason for suspecting important changes in the differences of one station to another from one month to the next, then monthly difference series should be considered. In our analyses we use seasonal (sometimes annual) difference and log ratio series.

3. The method of adjusting for discontinuities

a. Data

The methodological approach for adjusting time series of temperature and precipitation is specifically developed to make use of station history information. Brower (1985) describes a Historical Climatological Network (HCN) in the United States which has over 1200 stations (Fig. 1). This network, consisting mostly of cooperative station data (most of which are in rural areas, over 70% have populations < 10 000, and over 90% have populations < 50 000), has detailed station histories which can be accessed by electronic computers as well as a relatively large number of long time series (⩾80 years) with monthly temperature and precipitation data. The station history information can be used to ascertain all the information regarding changes in instrument locations (horizontal or vertical), changes in instruments, or changes in observation methods, i.e., tridaily observations, maximum/minimum observations, etc.

Given the availability of station history information, these data are used explicitly in the adjustment technique. Any change in instrument location, type of instrument, or averaging methods are treated as a potential station discontinuity, regardless of the magnitude of the change, i.e., a relocation of instruments by 20 m is treated the same as a relocation of instruments by 1 km. Microclimatological differences can often be quite substantial (Kalma et al., 1987; Carlson, 1986; Oke, 1978).

There are other methods which attempt to isolate—and to some extent adjust for—inhomogeneities in a station's climate record without explicitly considering station history information. Details of some of these methods are given by Maronna and Yohai (1978), Potter (1981), Mitchell (1961), and Alexanderson (1984, 1986). These methods are particularly useful for detecting undocumented station changes since they require no station history information. On the other hand, they must estimate the exact timing and number of discontinuities in the climate record which can be difficult, particularly when there are several station changes over the stations in the network. Mitchell (1961) indicates that ambiguous conclusions are possible when several neighboring stations do have inhomogeneities. Since station history information is not explicitly used in these methods, it complements the method, presented here, and can be viewed as a second (or first) step in an attempt to produce homogeneous records. Actually, some type of an iterative scheme making use of both approaches would be ideal, but no doubt a substantial undertaking.

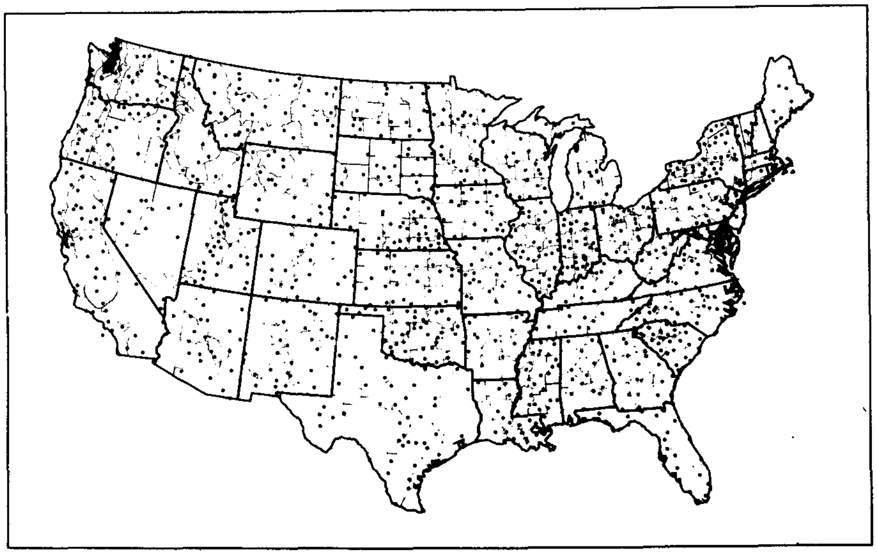

FIG. 1. Spatial distribution of the stations in the historical climatology network.

b. Adjustments and their associated confidence intervals

1) METHODS

There are several approaches that can be taken to adjust a station's records depending upon the homogeneity of the nearby stations with which it is compared. For example, if there is a nearby, already established homogeneous station with a reasonably high correlation of monthly anomalies between the two stations, the established homogeneous (HOM) station can be used to assess the impact of the candidate station's discontinuity. A simple t-test for temperature or the relatively powerful nonparametric Wilcoxon rank-sum test (Devore, 1982) can be used if the assumptions of normality and equal variances of the temperature difference series (d) or the log ratio difference series (r, the logarithm of the ratio of precipitation at the candidate station to the homogeneous station) before and after the discontinuity are thought to be invalid. The temperature or precipitation difference series can be bracketed into two parts $d_1(r_1)$ and $d_2(r_2)$. The statistical significance of the differences before (b) and after (a) the discontinuity in question can then be calculated. If the statistical test indicates that it is unlikely that the quantity $(d_{b,\text{CAN-HOM}} - d_{a,\text{CAN-HOM}})$ for temperature [or $(r_{b,\text{CAN/HOM}} - r_{a,\text{CAN/HOM}})$ for precipitation] has arisen due to chance, the CAN station should be adjusted (and the adjustment tested for statistical significance) to its longest period of record without any discontinuities. The reason why it is desirable to adjust to the longest period without discontinuities (indicated by the CAN history, remembering that in this case the CAN station is compared to a HOM) is attributed to the additional degrees of freedom gained by the statistical test, viz., a greater number of years to compare changes in the difference or log ratio series. In this regard, for each discontinuity, the adjustments, if any, are related back to the climate monitored at the CAN during its longest period of apparent homogeneity.

Based on the station histories in the HCN there are few, if any, homogeneous stations. Each network of nearby stations contains its own station inhomogeneities so it is from these stations that adjustments for potential discontinuities must be assessed. For this reason several stations are used in the difference series and the adjustments are always made such that the most current homogeneous period (which contains at least 5 yr of data; cf. section 3b2 and step 6 of this section) is used as the base period to which all adjustments are made. This is necessary because very few stations have a long homogeneous record as indicated by the station histories. This avoids over reliance on one station which may have undocumented nonclimate changes in its climate record such as the gradual urbanization around the station or changes in the local environment. In this regard, the overall procedure of identifying and producing homogeneous station records can be considered a two-step process. First, documented station changes are assessed, and second, methods such as those discussed in section 3a can be used to identify undocumented biases. The method enumerated here addresses the first of these two steps.

The purpose of this technique is to produce climate time series, as free as possible of discontinuous station inhomogeneities. Ultimately, regional and local climate change can be investigated from these stations. Some measure of uncertainty about the adjustments is desirable. This measure of uncertainty has not been addressed previously, but given this information it could prove quite useful in decisions regarding the ability to detect regional and local climate change above or below some threshold value. The uncertainty also provides some objective measure of selecting the most suitable stations for further analyses.

Another important characteristic of adjusted climate records with respect to regional climate change is the ability to retain the original measurement scale. Nelson et al. (1979) have provided a method where this is possible over climate divisions. This allows estimates to be made, even when using monthly data, regarding changes in quantities such as degree days, streamflow, etc. In studies of large geographic areas of climate change, this is not an issue of concern (Mitchell, 1961; Jones et al., 1986).

The use of at least several nearby stations requires a substantial quantity of electronic computer compatible data, particularly when station discontinuities are relatively frequent as in the United States (about six per century in the HCN). The methods used in this article for the HCN are given in detail. In some instances several decisions were made based on the HCN characteristics, i.e., station density, interstation correlation, variance, etc. Such decisions are noted and others who may choose to use this procedure (or variations of it) will need to consider any unique characteristics of their network. We point out these instances in the enumeration of the adjustment technique which is given as follows:

1) Identify and use as many of the candidate station's nearest neighbors as possible. Care should be taken to assure at least a positive correlation of anomalies with the candidate station. Another important aspect is that there are always at least several stations which have no changes in station history (potential discontinuities) during or on either side of a potential discontinuity at the candidate station. Due to the density of stations in the HCN the use of 20 nearest neighbors were chosen. This almost always resulted in significant positive correlations of anomalies (using standard statistical tests).

2) For the period of simultaneous operation, calculate the correlation of each of the nearest neighbors to the candidate station as well as the year-to-year standard deviation and means of each station over the seasons that will be used to assess and adjust for the impact of a potential inhomogeneity. For temperature and precipitation the four standard meteorological seasons were used for the HCN (winter: December, January, and February; spring: March, April, and May; etc.) but in special circumstances annual values were used. (Repeat step 1 using more or less neighbors if correlations are too low, negative or near zero, or station changes too frequent such that few, if any, stations have at least 5 yr of data without potential discontinuities themselves before and after discontinuities at the candidate station.)

3) For the candidate station, proceed backward in time and identify the year of the first potential discontinuity since the most recent discontinuity or the most recent year of data.

4) Form a difference series between the candidate and each of its nearest neighbors such that the number of years included in the difference series is limited by discontinuities in the CAN and NEIGH station. The number of years may differ from one NEIGH to the next. Figure 2 illustrates this concept.

5) For temperature, calculate the confidence interval using the Student's t-test for some preset significance level of the difference of the various difference series, d, before and after the potential discontinuity. This will result in one confidence interval for each difference series formed in step 4. For precipitation use the rank-sum confidence interval for the differences of the log arithm of the ratios. The calculation of the Student's t confidence interval and Wilcoxon rank-sum confidence interval can be found in Devore (1982) and many other statistical texts.

The "t" interval is given by

$$d_b - d_a + t_{\alpha/2,\mathrm{m+n-2}}[s_p(1/m + 1/n)^{0.5}], \qquad (1)$$

where

$$s_p = \{[(m-1)s_b^2 + (n-1)s_a^2]/(m+n-2)\}^{0.5},$$

and m is the number of years in the series before the discontinuity; n the number of years after the discontinuity; $t_{\alpha/2}$ the critical value of the t distribution for significance level α; and s_b and s_a are the sample standard deviations before and after the potential discontinuity, respectively. The quantity $d_b - d_a$ is used as the offset of the potential discontinuity.

The Wilcoxon rank-sum interval can be obtained by calculating all possible differences for each year i of $r_{b,i}$ with each $r_{a,i}$. This will result in m times n differences. If these differences are ordered from lowest to highest the rank-sum interval is given by

$$r_{(mn-c+1)} r_{(c)}, \qquad (2)$$

where c is the critical constant for the two-tailed, significance level for the Wilcoxon rank-sum test. The equation

$$C = (mn/2) + Z_{\alpha/2}\{[mn(m+n)]/12\}^{0.5} \qquad (3)$$

provides good approximations for C down to sample sizes $\geqslant 5$ yr before *and* after the potential discontinuity. The quantity $Z_{\alpha/2}$ is the standard normal deviate (z-score) for significance level ($\alpha/2$). The offset for the potential discontinuity is defined by the midpoint of the confidence interval $r_{(mn-c+1)}$, $r_{(c)}$.

The rank-sum interval gives up little with respect to the t-interval (Devore, 1982) even when the populations are normal, and in non-normal populations where the tails are heavy the interval may be considerably shorter than the t-interval.

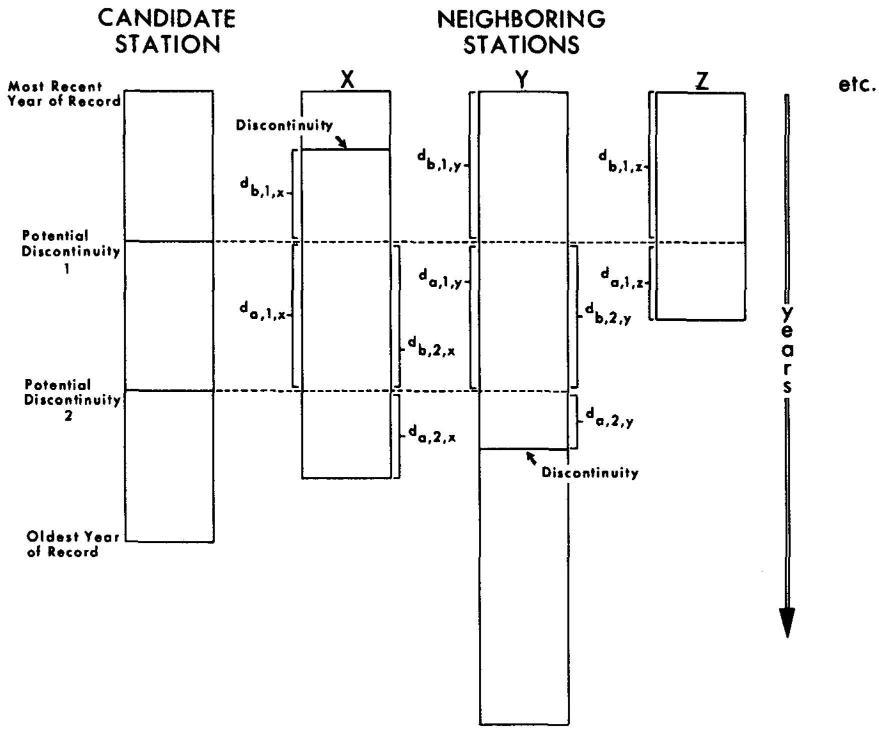

FIG. 2. Schematic of the procedure used to define the number of years in each difference series, d_x, d_y, and d_z bracketing a potential discontinuity at the candidate station.

6) Next, the width of the confidence intervals are ranked from narrowest to widest. If a sufficient number of years without potential discontinuities is available at nearby stations all confidence intervals other than the narrowest are recalculated (as in step 5) over the years of common operation with the candidate station straddling the potential discontinuity in question. At this point the two narrowest confidence intervals are identified and their original observations are merged together by a weighted average. That is:

$$d_{xy} = [CI_x/(CI_x + CI_y)]d_x + [CI_y/(CI_x + CI_y)]d_y. \quad (4)$$

Each set of observations is weighted by the width of its respective confidence intervals.

As before, the number of years over which the confidence interval (CI) is calculated must not contain potential discontinuities in any of the stations used (Fig. 2). If the resulting CI is wider than the CI calculated at the previous step, then proceed on to the next step in the process, and use the number of neighbors from the previous weighted average. Otherwise, merge in the next smallest CI and repeat the procedure. If there are fewer than two stations available that span at least 5 yr before and 5 yr after a potential discontinuity, stop and reassess whether enough information is available to correct for a discontinuity. In the HCN, as will be discussed in subsection 3b2, our tests indicate that when fewer than 5 yr of data or only one neighbor is available, satisfactory adjustments cannot be consistently made. If all stations in the network of 20 neighbors are included (this would be very unusual) go on to the next step.

7) The CI calculated in step 6 will have a width that may include a zero offset or no difference in the difference series, $(d_{b,\text{CAN-NEIGH}} - d_{a,\text{CAN-NEIGH}}) = 0$. If this occurs, no adjustment to the candidate series is called for, but the CI of the potential discontinuity in the series is a measure of the uncertainty in estimating the magnitude of the potential discontinuity. For identical confidence levels as the width of the CI becomes smaller, the test will be able to detect smaller potential discontinuities. If the CI does not include zero, then this difference becomes a nonzero offset $(d_{b,\text{CAN-NEIGH}} - d_{a,\text{CAN-NEIGH}})$, and it replaces any previously retained offsets. Subsequently, it is added back to the candidate station for all the years after the potential discontinuity in question, but only after all tests for potential discontinuities have been completed.

For precipitation, essentially the same procedure is followed, but log ratios are used in the difference series. If the CI includes zero, no adjustment is retained; otherwise the adjustment retained is the ratio $(10^{r_{b,\text{CAN/NEIGH}}}/10^{r_{a,\text{CAN/NEIGH}}})$, and it replaces any previously retained offset factors. Finally, it is multiplied by the original candidate time series after all tests for potential discontinuities have been completed.

8) For subsequent potential discontinuities repeat steps 3 through 7. The total uncertainty in the estimate

of the total offset at some constant confidence level after the jth discontinuity is calculated by

$$\mathrm{OFFSET}_j \pm \Big[\sum_{i=1}^{j} (w_i/2)^2\Big]^{0.5}. \qquad (5)$$

The width of the CI is w_i at the ith discontinuity for significance level α. This assumes that the offset for the $i-1$ potential discontinuity is independent of that in the ith discontinuity. Our results (section 3b2) indicate that this is a good assumption.

After all the potential discontinuities have been tested, then the confidence interval at each potential discontinuity can be used as a measure of how much confidence can be placed in the magnitude of any undetected potential discontinuities. It is entirely feasible that a CAN station with many adjustments may have a relatively narrow CI associated with each of its potential discontinuities and may be a more desirable station for studies of regional climate change than another station with few potential discontinuities. This can arise because of the characteristics of the CAN as well as the neighboring stations, i.e., small variances (year-to-year), high correlations with the candidate station, and/or few potential discontinuities. This will be demonstrated in a simulated dataset in subsection 3b2. In general, of course, stations with more discontinuities will have wider or larger CIs.

2) Testing of the adjustment scheme

The most important attributes of any scheme to assess or adjust station discontinuities is its ability to adjust the observations in the correct sense. In this regard, tests were carried out such that the appropriateness of any adjustments of the time series can be judged relative to a homogeneous series. This can be accomplished in two ways. First, stations that are relatively homogeneous for some given interval can be artificially made inhomogeneous by introducing an offset at a particular year. The ability of the adjustment scheme to find the appropriate adjustment can then be ascertained. There are several limitations in using this method to assess the value of any adjustment procedure:

1) Some procedure is required to ascertain whether the stations are relatively homogeneous to begin the tests;
2) The stations are assumed to have no undocumented biases or errors; and
3) The types of tests that can be performed are limited to the actual observations.

A preferred alternative solution is to select several networks of stations with known interstation correlations and variances and use the concepts from autoregressive modeling to simulate the observed data such that it is known to be absolutely homogeneous. Then biases of some magnitude can be introduced into the series using a simulated station history. These inhomogeneous data are then adjusted and the confidence interval of the adjustments estimated. These estimates are then evaluated with respect to the original data to assess the methodology.

Based on the annual mean temperature and precipitation for each of the candidate stations with over 80 yr of data to each of their 20 for all the stations in the northeast portion of the United States (approximately 150 stations), the variance explained by *each* of the candidate's 20 nearest neighbors was calculated and averaged across the 20 stations. This amounted to one value per candidate station or about 150 values. For temperature, the highest average variance explained was 0.72, the median 0.62, the tenth percentile 0.48, and the lowest 0.28, but for precipitation the respective average variances explained were 0.53, 0.43, 0.29, and 0.18. This is an important characteristic with respect to the ability of neighboring stations to detect a potential discontinuity at the candidate station. This information formed the basis for various scenarios of interstation correlations (or departures from the mean at the candidate explained by neighbors), year-to-year variances, and the magnitude of artificially generated discontinuities in the data. Specifically, for each of the average variances explained by the nearest 20 stations as identified by the position statistics above, i.e., for temperature 0.72, 0.62, 0.48 and 0.28; for precipitation 0.53, 0.43, 0.29, and 0.18, a total of 168 time series (21 stations times 8 sets) were generated such that each neighbor station had the same correlation with the candidate station and year-to-year variance as that observed in the original network of stations. This was accomplished using the following relationships:

$$X_{t,\mathrm{CAN}} = (r_{-1}X_{t-1,\mathrm{CAN}} + Z_t)s_{\mathrm{CAN}}, \qquad (6)$$

where $X_{t,\mathrm{CAN}}$ is the candidate stations observation at year t, r_{-1} is the lag one autocorrelation, Z_t is an independent normal random deviate with mean zero (500 for precipitation) and standard deviation $(1 - r_{-1}^2)^{0.5}$. In this regard, any year-to-year persistence in the original series (often reflected as trends or vice versa) is preserved. We purposely neglect higher order autoregressive models because the variance density of the spectrum, beyond the annual cycle, is overwhelmed by white and red noise. The observations for the 20 nearest neighbors are then given by

$$X_{t,\mathrm{NEIGH}} = [(r_{\mathrm{CAN,NEIGH}}X_{t,\mathrm{CAN}}) + Z_t]s_{\mathrm{NEIGH}}, \qquad (7)$$

where $r_{\mathrm{CAN,NEIGH}}$ is the correlation of the candidate station with a specific neighbor and s_{NEIGH} is the year-to-year standard deviation of the neighboring station. By varying the size of s_{NEIGH} and s_{CAN} to one, three, and six times its observed value, three sets of simulations were produced for each of the four scenarios of average interstation correlations for temperature and precipitation.

The next step involved the introduction of the potential discontinuity. This was accomplished by using the normal random deviate with a mean zero but a standard deviation (σ_m) equal to 0.1, 0.5, and 1.0 for temperature and 0.01, 0.03, and 0.05 for precipitation with mean equal to 1. For temperature, Fahrenheit units can be assumed since the data are originally measured in this system, but for precipitation these numbers are dimensionless since they are merely multiplication factors of the simulated data. In total, this produced 36 scenarios for temperature and 36 for precipitation with various combinations of interstation correlations (4), year-to-year variances (3), and standard deviations of the magnitude of the discontinuity (3) (cf. Table 2) ($4 \times 3 \times 3 = 36$). Each series consisted of 100 values.

The question remains as to a reasonable choice for the number and timing of the discontinuities in each of the 36 simulations. Technicians working with the station histories in the HCN indicated that about six discontinuities per 100 years of data was a reasonable number to expect based on their experience with the station histories. The number, of course, could vary substantially from station to station. Using this information, a random number generator (congruence method) was used to simulate the probability of a station discontinuity at each of the stations in the network corresponding to the 100 yr of simulated data. Four stations were assigned a probability of 0.04 for a discontinuity, four a probability of 0.05, five a probability of 0.06, four a probability of 0.07, and four a probability of 0.08. Using these probabilities, the station with the fewest discontinuities had only 1, but the station with the largest number of discontinuities had 11. In order to produce several variations of the number of potential discontinuities at the candidate station, 21 sets of station histories were produced from the simulation by interchanging the history of the candidate station with each of the other station histories. This was accomplished by interchanging the candidate's history with the first neighbor, the first neighbor's history with the second neighbor's history . . . and the 20th neighbor with original candidate station's history. This process was repeated 21 times so that 21 sets of station histories, each set different from the other set, was produced for *each* of the 36 scenarios given in Table 2. In this manner each time a potential station history discontinuity was encountered, going back in time through the simulated station history, the time series from the next data point (or year) to the beginning of the series (the earliest year) was perturbed by an amount equal to P, where P is a standard normal deviate with the mean equal to zero and the standard deviation equal to that specified for the scenario under consideration. The original unperturbed series is always retained so that the adjustments could be evaluated for their accuracy as well as the appropriateness of the confidence intervals.

TABLE 2. Scenarios used to test the adjustment program for precipitation and temperature. Here T implies temperature, P precipitation, sim simulation, obs observed, and ρ^2 the average variance explained by each of the 20 nearest neighbors.

	$\rho^2_{\text{CAN,NEIGH}}$				$s_{\text{discontinuity}}$	
	P*	T†	$s_{\text{can,sim}}/s_{\text{can,obs}}$	$s_{\text{neigh,sim}}/s_{\text{neigh,obs}}$	P	T
1	0.53	0.72	1	1	0.01	0.1
2	0.53	0.72	1	1	0.03	0.5
3	0.53	0.72	1	1	0.05	1.0
4	0.53	0.72	3	3	0.01	0.1
5	0.53	0.72	3	3	0.03	0.5
6	0.53	0.72	3	3	0.05	1.0
7	0.53	0.72	6	6	0.01	0.1
8	0.53	0.72	6	6	0.03	0.5
9	0.53	0.72	6	6	0.05	1.0
10	0.43	0.62	1	1	0.01	0.1
11	0.43	0.62	1	1	0.03	0.5
12	0.43	0.62	1	1	0.05	1.0
13	0.43	0.62	3	3	0.01	0.1
14	0.43	0.62	3	3	0.03	0.5
15	0.43	0.62	3	3	0.05	1.0
16	0.43	0.62	6	6	0.01	0.1
17	0.43	0.62	6	6	0.03	0.5
18	0.43	0.62	6	6	0.05	1.0
19	0.29	0.45	1	1	0.01	0.1
20	0.29	0.45	1	1	0.03	0.5
21	0.29	0.45	1	1	0.05	1.0
22	0.29	0.45	3	3	0.01	0.1
23	0.29	0.45	3	3	0.03	0.5
24	0.29	0.45	3	3	0.05	1.0
25	0.29	0.45	6	6	0.01	0.1
26	0.29	0.45	6	6	0.03	0.5
27	0.29	0.45	6	6	0.05	1.0
28	0.18	0.28	1	1	0.01	0.1
29	0.18	0.28	1	1	0.03	0.5
30	0.18	0.28	1	1	0.05	1.0
31	0.18	0.28	3	3	0.01	0.1
32	0.18	0.28	3	3	0.03	0.5
33	0.18	0.28	3	3	0.05	1.0
34	0.18	0.28	6	6	0.01	0.1
35	0.18	0.28	6	6	0.03	0.5
36	0.18	0.28	6	6	0.05	1.0

* $\bar{P} = 500$, $\sigma^2(P) =$ OBSERVED
† $\bar{T} =$ OBSERVED, $\sigma^2(T) =$ OBSERVED

Another approach that could have been used to assess the value of the method would have been to determine the limit of detectability of discontinuities of some specified magnitude, with the magnitudes chosen as various multiples of the variance of the difference or log ratio series. There are several difficulties with this approach, however: first, the detectability of the discontinuity is a function not only of the magnitude of the discontinuity and the variance of the difference series, but also of the sample size (d_b and d_a); second, the treatment of individual discontinuities cannot be considered in isolation of other discontinuities since few series have only one potential discontinuity. That is, one erroneous adjustment, especially if made in the most recent period of record, would jeopardize the appropriateness of all other adjustments.

Figures 3 and 4 summarize the effects of the adjustments. Focusing on Fig. 3, several prominent features appear in all the scenarios regardless of the explained variance of the neighboring stations with respect to the candidate station. First, the higher significance or confidence *levels* have the smallest improve-

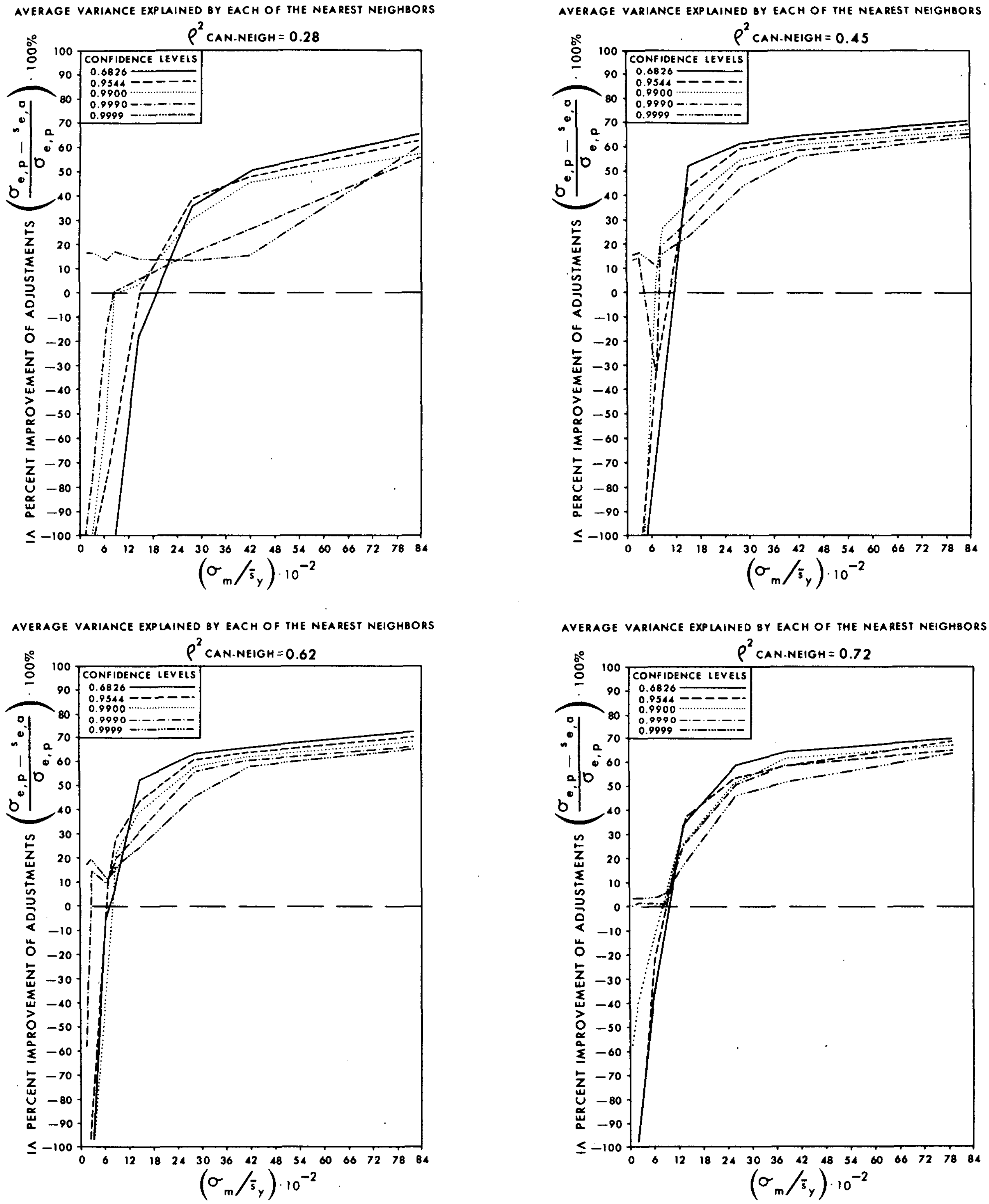

FIG. 3. For the t-interval (used for temperature series), the percent improvement of the adjusted data versus the ratio of the standard deviation of the discontinuity (σ_m) to the year-to-year standard deviation averaged over the network of 20 nearest neighbors ($\bar{s}_y$). Standard deviation of perturbed simulated data denoted by $\sigma_{e,p}$ and the standard error of estimate of the adjusted data denoted by $s_{e,a}$. $\rho_{\text{CAN-NEIGH}}$ is the average of the year-to-year variance explained by each of the 20 nearest neighbors.

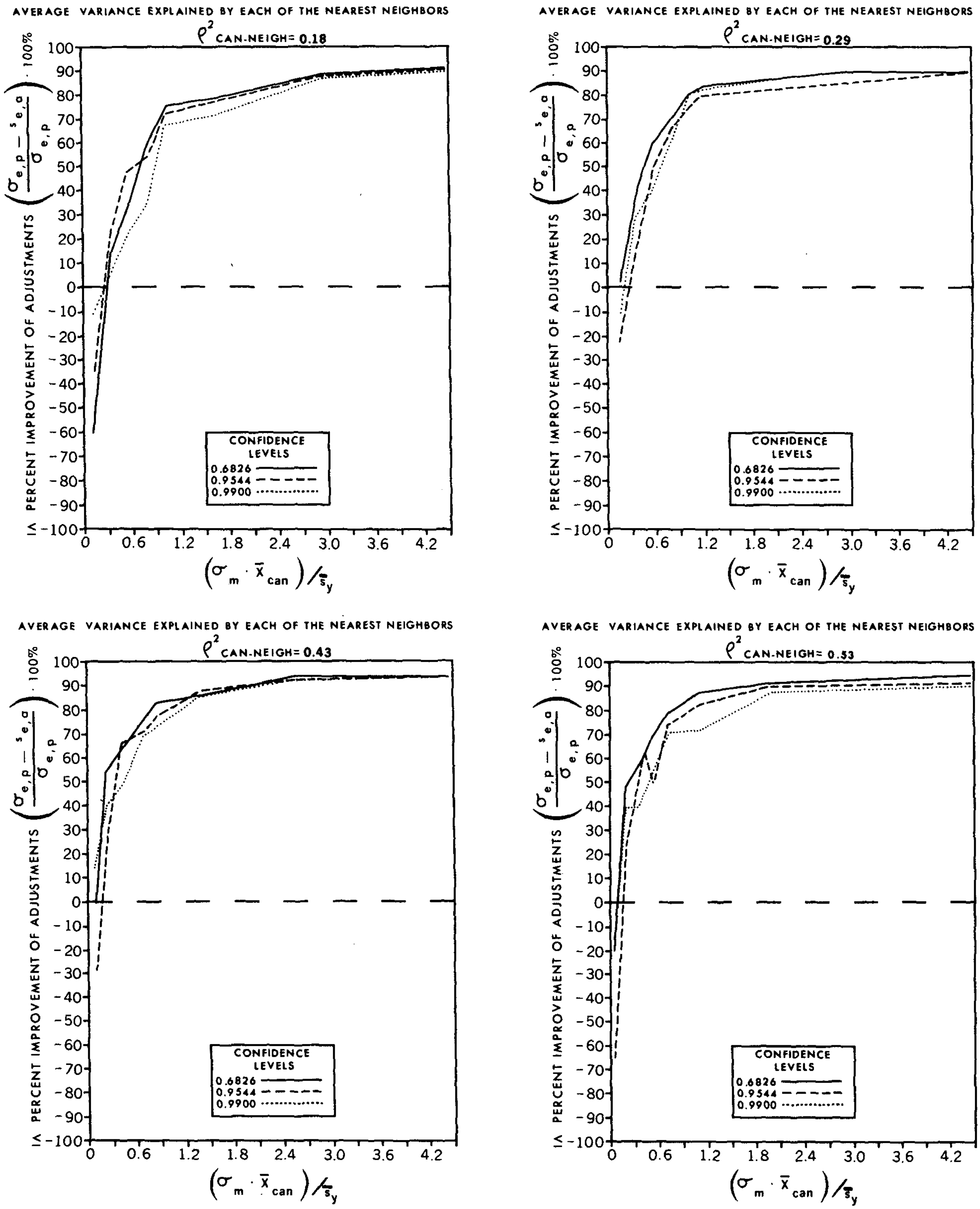

FIG. 4. As in Fig. 3 except for the case of precipitation series and nonparametric Wilcoxon rank-sum interval.

ment, with respect to making no adjustments whatsoever at large values of $\sigma_m/\bar{s}_y$, but they have the largest improvements when $\sigma_m/\bar{s}_y$ becomes small. The quantity $\bar{s}_y$ is simply the square root of the year-to-year variances averaged over the 20 neighbors. Second, the higher significance level does not vary as dramatically in its improvement with respect to making no adjustments compared to the lower significance levels. Third, the lower significance levels have the greatest improvement with respect to no adjustment when $\sigma_m/\bar{s}_y$ is large (>0.25). It is apparent that at the low confidence levels the line connecting identical confidence levels for various ratios of $\sigma_m/\bar{s}_y$ depicts a negative improvement with respect to no adjustment whatsoever as $\sigma_m/\bar{s}_y$ becomes small (<0.10) particularly as $\rho^2_{\text{CAN-NEIGH}}$ decreases. On the other hand, higher confidence level adjustments continue to improve upon the discontinuous series even for many low values of $\sigma_m/\bar{s}_y$ and $\rho^2_{\text{CAN-NEIGH}}$. If some assumptions can be made with respect to the magnitude of the smallest expected potential discontinuities in the climate record, then some guidelines can be developed with respect to the appropriate significance level to choose for the decision to adjust at each potential discontinuity in the station history. If the smallest standard deviation of discontinuities at any candidate station is assumed to be 0.2°F (0.1°C) with mean zero, virtually the limit of the resolution of the reported data, then rules can be adopted such that the significance level chosen for making an adjustment will in general improve upon the unadjusted data. For example, in Fig. 3, if the $\bar{s}_y$ is 0.5, then the ratio of σ_m to $\bar{s}_y$ is 0.40, and it appears the best confidence level for use is 0.68 for all four values of ρ^2. This indicates that even when the discontinuity is small in magnitude and there are relatively poor interstation correlations, if the year-to-year standard deviation of the neighboring and candidate stations is *quite* small, it is advantageous to adjust with a relatively low significance level. On the other hand, when $\bar{s}_y$ is large and σ_m small, then it pays to be very conservative in the adjustment procedure, and the use of a high confidence level would be recommended, i.e., 0.9999. An evaluation of the errors in the adjustments indicates that this is often due to the fact that when the ratio of σ_m to $\bar{s}_y$ is small, one poor adjustment can totally cancel out many good adjustments, particularly if the poor adjustment occurs near the most recent years of record.

Figure 4 depicts the characterization of the most appropriate confidence levels for use with precipitation series. Only three confidence levels are represented due to the formulation of significance levels of nonparametric tests with small sample sizes. The abscissa in these diagrams has the ratio $\sigma_m/\bar{s}_y$ multiplied by the mean of the candidate station $\bar{X}_{\text{CAN}}$. The rationale for this stems from the fact that the discontinuities for precipitation are obtained by multiplication of the observed simulated value by σ_m (as opposed to addition for temperature). As described earlier, the ratio method has long been accepted as an effective means of representing differences of precipitation between stations (Kohler, 1949; Conrad and Pollak, 1950). Using this concept, the magnitude of the discontinuity is also a function of the mean precipitation.

The primary difference between Figs. 3 and 4 is the small difference between the various confidence interval's percent improvement for the nonparametric test and the relatively large difference for the parametric t-test. Since the narrowest confidence interval (lowest confidence level) provides slightly better adjustments than the wider confidence intervals (highest significance level) for most values along the abscissa it is used for all adjustments with respect to precipitation.

In each of the 72 scenarios the accuracy of the CIs was determined by calculating the percent of time the original unperturbed data fell into the calculated CI. The results were satisfactory with all of the CIs spanning the original data reasonably close (within a few percent) to the expected frequency. From these results it was apparent that when adjustments were poor, and not properly correcting for the inhomogeneities (less than 0% improvement), it was due to a few very poor adjustments with respect to the original unperturbed data.

Based on the results from the 72 scenarios and 21 sets of station histories for each scenario, it was also determined that at least 5 yr of data before and after each potential discontinuity should be available before an assessment of the magnitude of the potential discontinuity is attempted. As indicated in section 2b1 (step 6), when there are fewer than 5 yr of data available, the adjustment scheme ignores the potential discontinuity and adjusts back to the previous discontinuity. In the HCN this usually occurs in the nineteenth century when many stations begin to drop out of the network or when stations have frequent changes.

Based on Figs. 3 and 4, decision rules for adjustment were developed using the assumption that on the average, the true σ_m that would be expected for a series of station discontinuities would be 0.2°F (0.1°C) for temperature and a 3% change in the precipitation record. In effect this means that about 95% of all station moves or new instruments produce no more than ±0.4°F (0.2°C) bias in temperature and a bias of ±6% with respect to precipitation. These criteria are used as the basis for the decision regarding whether adjustments should be attempted. If the true standard deviation of the moves is larger, then by being overly conservative we choose not to introduce errors at stations which may already be reasonably homogeneous as opposed to moderately reducing large inhomogeneities at stations which have many discontinuities. On the other hand, if the true standard deviation of the moves is less than that assumed the opposite is true. The various significance levels and decision rules for $\rho^2_{\text{CAN-NEIGH}}$, $\bar{X}_{\text{CAN}}$, and $\bar{s}_y$ which were used, based on the HCN, are given in Table 3. The decision rules for adjusting the time series are derived using the relationship

$$\sigma_{mt}/\bar{s}_y < \Phi \tag{8}$$

for temperature, and

$$\sigma_{mp}\bar{X}_{\text{CAN}}/s_y < \Phi \tag{9}$$

for precipitation, where Φ is the point on the abscissa (Figs. 3 and 4) of no improvement or a larger percent improvement using another confidence level; $\bar{X}_{\text{CAN}}$ is

TABLE 3. Confidence levels used in the HCN adjustment program for various means ($\bar{X}$), average interstation variances explained by each neighboring station (ρ^2CAN, NEIGH), and the year-to-year standard deviation averaged across the network ($\bar{s}_y$). (ADJ implies adjustment.)

Temperature Adjustments

ρ^2CAN − NEIGH	$\bar{s}_y$ (°F)	Confidence level
<0.36	<0.9	0.6826
<0.36	≥0.9 but ≤1.8	0.9500
<0.36	>1.8	0.9999
≥0.36 but <0.54	<2.0	0.6826
≥0.54 but <0.54	≥2.0 but ≤3.0	0.9500
>0.36 but <0.54	>3.0	0.9999
≥0.54 but ≤0.67	<2.0	0.6826
≥0.54 but ≤0.67	≥2.0 but ≤3.2	0.9500
>0.54 but <0.67	>3.2	0.9999
>0.67	<2.0	0.6826
>0.67	≥2.0 but ≤3.2	0.9900
>0.67	>3.2	0.9999

Precipitation Adjustments
(All adjusted using a confidence level of 0.6826)

$\rho^2_{\text{CAN - NEIGH}}$	Adjust if
<0.24	$(\bar{s}_y/\bar{X}_{\text{can}})$ <0.1
≥0.24 but ≤0.36	$(\bar{s}_y/\bar{X}_{\text{can}})$ <0.2
>0.36	$(\bar{s}_y/\bar{X}_{\text{can}})$ <0.3

the mean precipitation for one of the four seasons or the annual precipitation; and σ_{mt} and σ_{mp} are the standard deviations of the station discontinuity. For operational purposes, if the characteristics of the candidate station and 20 nearest neighbors fail (8) or (9) for seasonal averaging periods, annual values are used instead. If no adjustment is called for at this point, then the adjustments are not likely to make the data less biased than the original data, and in some instances they could make it considerably more biased. This rarely occurred in the temperatures series, but was not uncommon in the precipitation series, particularly in the warmer seasons in the dry western United States.

As was discussed in subsection 3b1, one of the main advantages of the approach used in this adjustment scheme is that CIs are provided for each adjustment. The extent to which each station can be determined to be free of discontinuities is provided by these CIs. Those stations with the narrowest CIs are those to which the most confidence can be attached. It cannot be inferred, however, that stations with wide CIs are necessarily contaminated with nonclimatic biases. Instead, based on the climate network, it cannot be ascertained as to whether their record is relatively free of discontinuities. Another important characteristic of the method relates to the fact that, unlike some methods (Nelson et al., 1979), adjustments are not made for every potential discontinuity. Instead they are based on the results of Monte Carlo simulations which capture many of the important characteristics of the data.

4. Adjustments: Examples

Several examples are provided which illustrate the impact of the adjustment scheme on the original observations. In these examples, discontinuities for changes in the ending time of the climatological day with respect to maximum and minimum temperature (Mitchell, 1958) were not treated as discontinuities, although they could have been, but the model developed by Karl et al. (1986) was used to adjust for these discontinuities. Each station was made consistent with a midnight-to-midnight observation schedule. On a monthly basis for a relatively long time series this effect is insignificant with respect to precipitation, and it was ignored. Additionally, missing data in the original series were estimated, but not used in the adjustment scheme, by a procedure identical to that described for the adjustments, i.e., estimates were based on the differences or log ratios with neighboring stations. In this regard, a confidence interval of the missing data estimate could also be provided. Furthermore, when potential discontinuities were too numerous to make adjustments (fewer than 5 yr before and after a potential discontinuity) then estimates of the original data were made by treating the data as if they were missing, consistent with the most recent location of the station's instruments or instrument type (or most recent location with at least 5 yr of operation).

a. Temperature

The time series of mean annual temperatures at four New England stations are presented in Fig. 5. All of these stations are within several kilometers of the ocean except for Amherst, Massachusetts. The stations are separated, at most, by approximately 150 km (Amherst to Block Island, Rhode Island) while some stations (New Bedford, Massachusetts to Block Island) are separated by about 50 km. Despite the proximity of these stations, the overall picture of the regional change in climate during the twentieth century is somewhat confusing as New Bedford depicts dramatic warming, Blue Hill, Massachusetts, gradual warming, and Amherst and Block Island, only slight warming. After making adjustments for potential discontinuities, the four stations depict a more coherent record of twentieth century climate change than do the original observations, namely, a progressive warming.

Many of the differences can be attributed to the great number of potential discontinuities at each station. Some of the noteworthy changes include a change in height of the thermometer above the ground at New Bedford, in 1974, from 1.5 to 17 m along with a 10 m decrease in station elevation and a 150 m station relocation. Additionally, the potential discontinuity in 1906 looks suspicious as the station moved about 160 m, its elevation above sea level decreased by 10 m, and

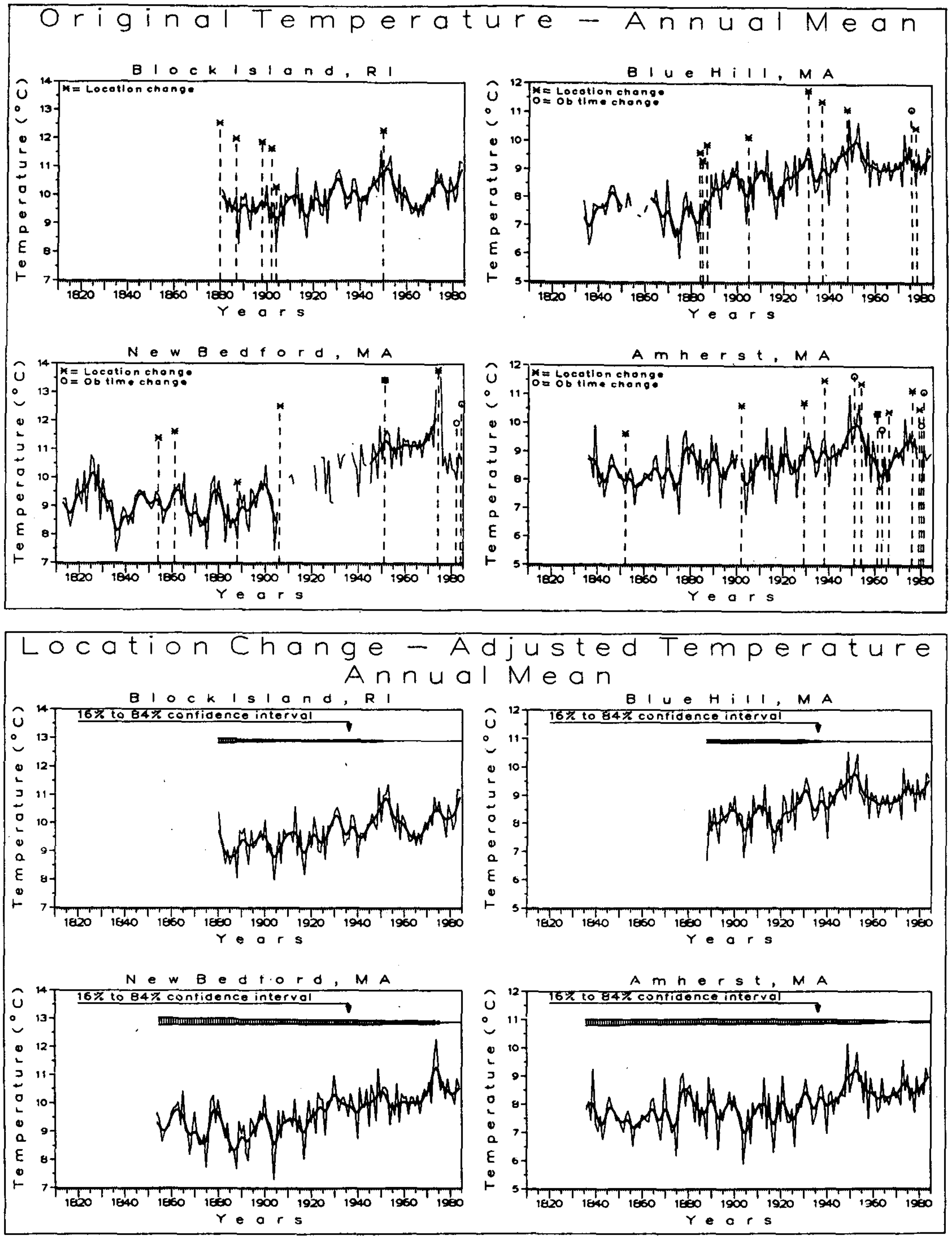

FIG. 5. Original, time of observation (ob) bias adjusted, and location change and time of observation bias adjusted mean annual temperatures. Smooth curve is a nine-point binomial filter truncated at the endpoints. The width of the confidence intervals are given for time of observation bias adjustments and location change adjustments.

it changed the method of calculating mean temperature by using maximum/minimum thermometers in place of tridaily temperature observations from a dry bulb thermometer. At Amherst, there are many potential discontinuities in the record, especially in the 1950s and 1960s. Most of the changes are station relocations and associated changes in its elevation above sea level. Blue Hill, Massachusetts has a number of potential discontinuities in its record, but most of the changes are not large except for a 4 m change in instrument height above the ground in 1887 (5.5 to 1.8 m) and again between 1931 and 1937 (1.8 to 5.5 to 1.8 m).

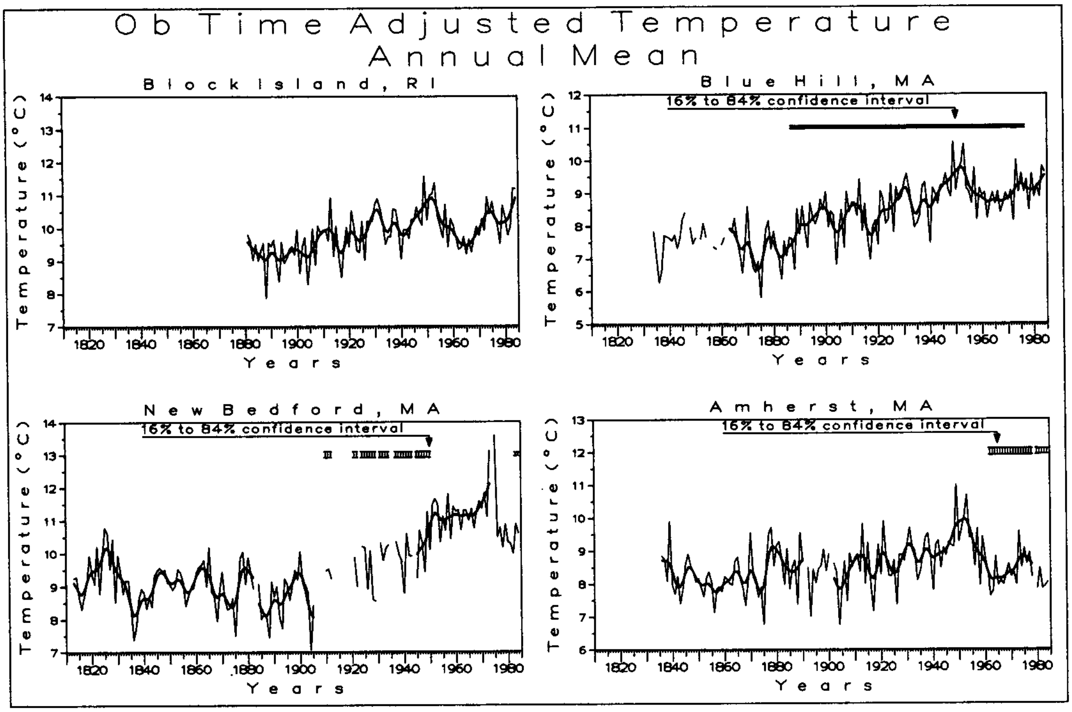

FIG. 5. (*Continued*)

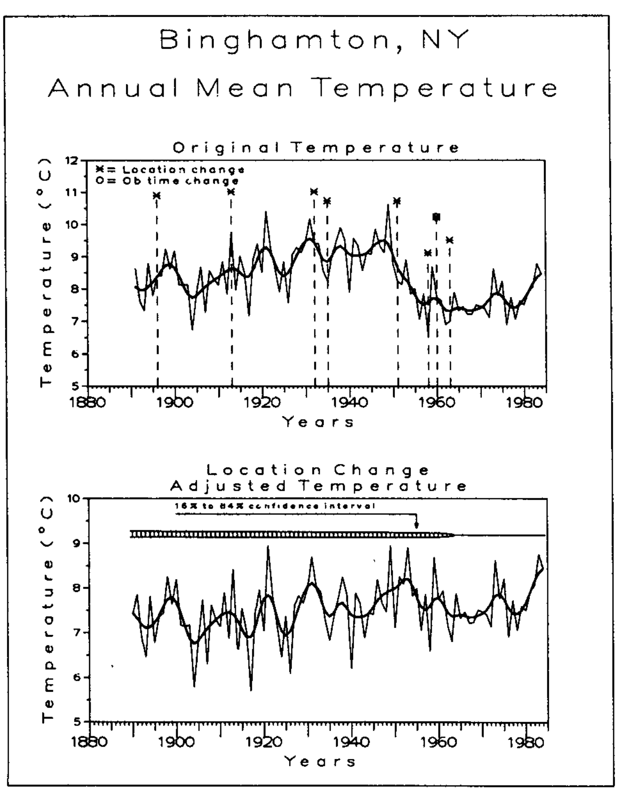

FIG. 6. As in Fig. 5 except no time of observation bias adjustments were needed.

Block Island has had one substantial move since 1904 as the station relocated 1 km in 1951 and changed its elevation by about 25 m.

An example of a rather large adjustment required at the beginning of a climate fluctuation is given in Fig. 6. In 1951 the Binghamton, New York station moved from the city to the airport, about 15 km north-northwest of the city location, and as a result its elevation changed by 225 m and the height of the thermometer above the ground was reduced from 17 to 1.2 m. This resulted in an exaggerated cool epoch in the climate record during a time when the region was already undergoing a relatively cool sequence of mean annual temperatures.

An example of some of the more subtle impacts of station inhomogeneities can be seen by examination of the original and adjusted data at Setauket, New York (Fig. 7). This station has a reasonably good station history, and it has been operated by the same family since 1885 with only one potential discontinuity which occurred in 1960. The station moved only 300 m, but a standard cotton region shelter was also introduced at this time and the ending time of the climatological day changed from 2100 LST to 1700 LST. Previous to 1960 a "window box" attached to a window on a north-facing unheated porch had been used to house the thermometers. The effect of moving the thermometer from the wall to an open field 300 m away from the house had a major impact on the minimum temperature and the range, and a lesser impact on the maxima

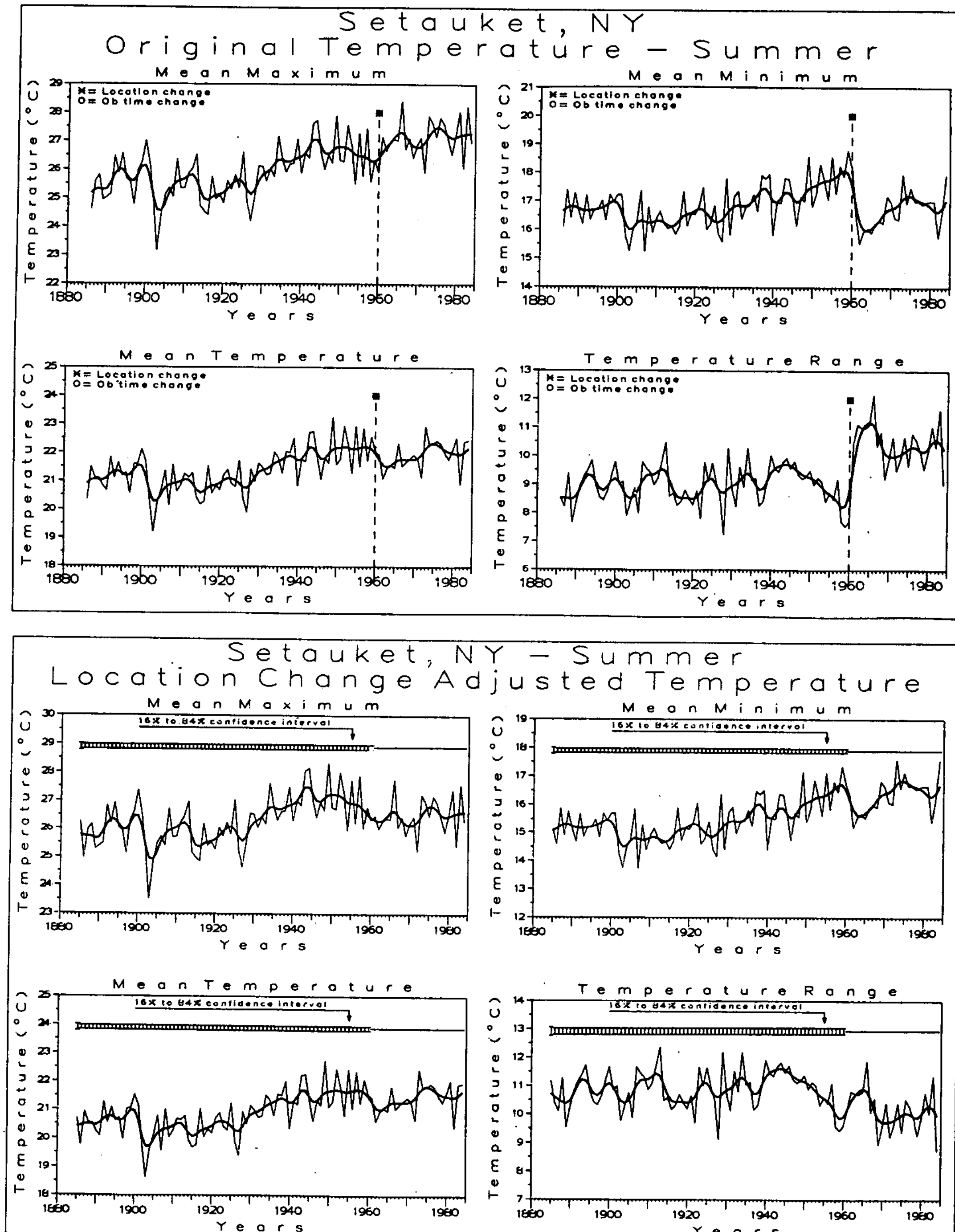

FIG. 7. As in Fig. 5 except summertime (June, July and August) original, time of observation (ob) bias adjusted, and location change and time of observation bias adjusted temperature for the maximum, minimum, mean, and range (maximum minus minimum).

and mean. The maxima and range were reduced, and the minima and mean were increased when the window box shelter was in use. The temperature range, based on the original data, was increasing during the period 1941–80, contrary to what Karl et al. (1984) report for the northeast United States during summer. Subsequent to the adjustments however, the temperature range at Setauket fell in line with the other stations as Karl et al. (1984) report.

b. Precipitation

Three different adjustment problems are presented in Fig. 8 with respect to precipitation. At Wellsboro 3S, Pennsylvania, a gradual bias took place with respect

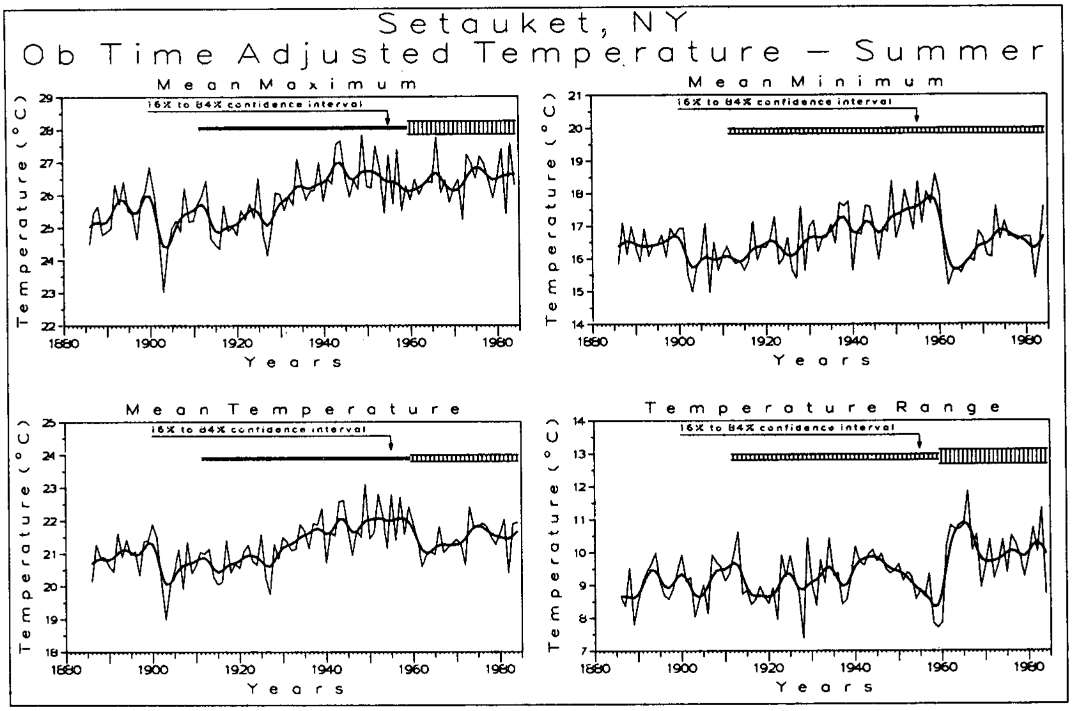

FIG. 7. (*Continued*)

to the precipitation record prior to 1900. Rochester, New York, had a decade (1870s) of high precipitation values relative to the rest of the network, which occurred between two station and instrument relocations (about 10 m in the vertical and several hundred meters in the horizontal direction). At Syracuse, New York, an unusually large amount of precipitation was reported for six consecutive years (1971–76) during a 10-yr period when the instruments were not changed or moved, based on the station history.

Not surprisingly, the adjustment scheme did not properly adjust for the gradual bias as Wellsboro (Fig. 8). The overall amount of precipitation prior to 1900 still contains a suspicious looking trend toward lower precipitation. The overall quantity of precipitation during this period is somewhat closer to the observed precipitation subsequent to the 1906 station move. This is a good example of the need to check for undocumented biases. If this station is not removed from the network prior to using the scheme it will improperly adjust nearby stations prior to 1900.

The adjustment scheme considerably reduces the precipitation at Rochester during the 1870s, and the decade reverses from extremely wet to moderately dry in comparison to other decades. The relative dryness is more consistent with other stations operating at this time. The adjustment scheme appears to have improved the consistency of the climate record.

An interesting example of an apparent undocumented change at Syracuse is depicted during the years between the 1968 and 1978 station changes. Beginning about 1971 and persisting until 1977, Syracuse received substantially more precipitation than any of its nearby neighbors in the HCN as well as other stations with shorter records not included in the HCN. Figure 9 depicts the gradual increase of precipitation at Syracuse relative to several nearby stations (since 1970) and two climate divisions. A climate division average is based on all available National Weather Service data within a given portion of a state. One climate division (Central Lakes Climate Division) contains Syracuse while the other (Great Lakes Climate Division) is just north of Syracuse and is adjacent to the Great Lakes. Other nearby stations in New York are only tens of kilometers away from Syracuse. Skaneateles is about 25 km southwest, Baldwinsville approximately 15 km west, Brewerton Lock 23 10 km north, and Canastota about 15 km east of Syracuse. In Fig. 9, the gradual change to relative heavy precipitation at Syracuse is several years after a station relocation of the recording rain gauge of only several hundred meters, and it appears to end just prior to another relocation of similar proportions in 1978. Perhaps by serendipity, the undocumented anomalous measurements are adjusted downward. Nonetheless, the precipitation during the years 1969 and 1970 probably should not have been adjusted nor included in the period of years which were used to calibrate the size of the adjustment necessary. In this regard, for the adjusted data the specific years 1969 and 1970 are underestimates of the actual precipitation and the remaining years 1971 to 1977 are still probably slightly overestimated.

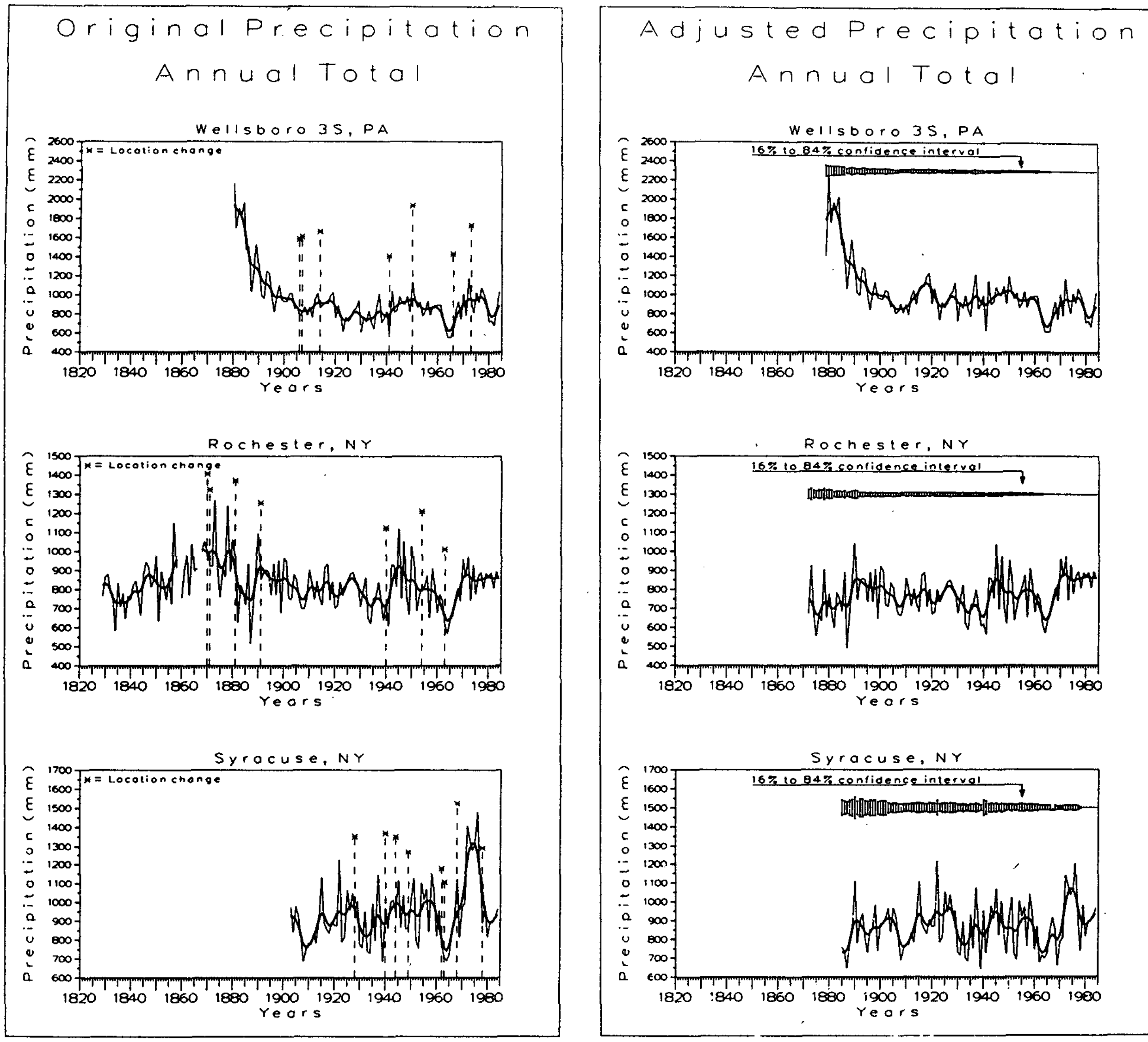

FIG. 8. As in Fig. 5 except original location change adjusted total annual precipitation.

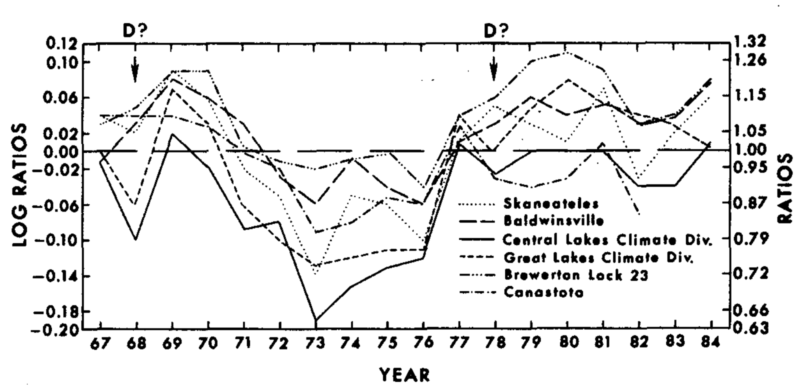

FIG. 9. Log ratio and ratio of total annual precipitation at nearby N.Y. stations divided by total annual precipitation at Syracuse, N.Y. "D?" implies a potential discontinuity.

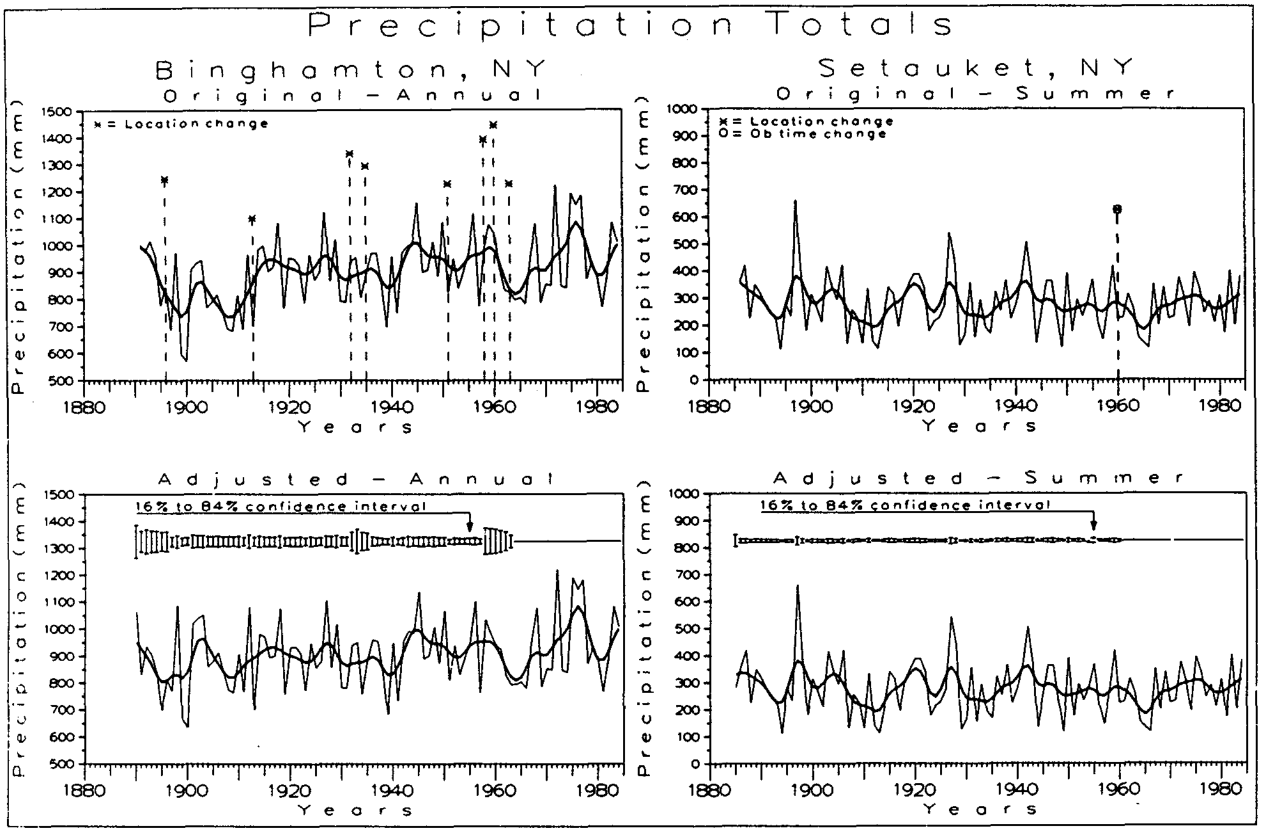

FIG. 10. As in Fig. 8 except for summer seasonal total at Setauket, N.Y.

Other examples of potential discontinuities are depicted in Fig. 10. The relatively innocuous move at Setauket, New York, is not adjusted for, but perhaps more surprising are the lack of any adjustments at Binghamton, where the temperature record was found to be quite inhomogeneous.

5. Summary and conclusions

A scheme has been developed which adjusts climatological time series of temperature and precipitation for discontinuities due to station changes. The procedure has been tested and developed by the utilization of simulation studies, and it has been implemented in over 1200 stations in the United States. The method can be applied to maximum and minimum as well as mean monthly or seasonal temperatures and total precipitation. It requires station history information and a network of nearby stations in order to assess any potential discontinuity at a specific station. The method produces an adjusted time series in the same scale as the original measurements (not an anomaly series) and confidence intervals either side of the data whenever a potential discontinuity exists in the time series.

The adjustment procedure was guided by a series of simulation studies which tested the adequacy of the approach. The results indicated that the error of estimating the adjustments needed for station discontinuities is a function of the station's correlation with a network of nearby stations, the magnitude of the year-to-year variance of the network, and the number and expected size of any of the potential discontinuities attributed to station changes. The simulation studies indicated that in some circumstances, depending on the network characteristics, adjustments to data based on changes before and after a potential discontinuity can actually make the data more biased than if no adjustments had been applied.

In order to illustrate the impact of the adjustments, several time series were analyzed. Comparison of the original and the adjusted temperatures indicate that by using the adjusted data, a substantially cleaner and consistent picture of regional climate change evolves than would otherwise be the case by using the original data. Examples can easily be found where the impact of changes in instrument locations, e.g., roof top to surface, window box to cotton region shelter, etc., are largely removed and the time series is made more suitable for use in studies of regional climate change. Comparison of original and adjusted precipitation time series, however, indicates that important differences in regional trends are not as readily discernable. On the average, the precipitation time series should be an improvement over the original series, but occasionally some obviously poor adjustments can be found, i.e., at Wellsboro 3S, PA.

As with many techniques, certain caveats must be given for the described method:

1) Accurate and complete station history information is essential to this method. Unfortunately, even with the detailed station histories compiled for the HCN, some stations do not have complete station histories as shown in section 4;

2) station histories rarely include information on environmental changes around the station;

3) the method may adjust too often if the assumption regarding the standard deviation of the biases introduced by station changes is significantly too high, but this can be easily changed; and

4) stations with nonclimatic progressive changes due to urbanization may lead to inappropriate adjustments at nearby stations.

The latter problem is mitigated to some extent in the HCN since 70% of the stations have populations < 10 000 in the 1980 census and 90% have populations < 50 000. Also, the method uses as many nearby stations as possible which reduces such effects and the adjustments usually do not span more than 20 yr so that only a portion of the urban warming at a single station may be included in the adjustments. The visual prescreening of difference plots prior to the use of this method, as described by Jones et al. (1986), can be used to help identify these stations. Unfortunately, the visual prescreening process, like the other methods which do not explicitly use station histories, is also hampered when there are a substantial number of discontinuities in the record, and when much of the network has potential urban effects. Probably the best solution is to avoid the use of urban stations in the adjustment method and to use an iterative procedure with methods that do and do not use station histories.

One possible improvement to the method, that has not been tested, would be to vary the confidence level (α) in the decision to adjust or not adjust; that is, make α a function not only of interstation correlation, variances, and means, but on the position of the potential discontinuity in the time series. During the years near 1984 (or most current year) keep α low (<0.01), but near the beginning of the record, α could be relaxed and made higher. This might improve the technique because a poor adjustment near the most recent year of record causes problems throughout the rest of the time series. Particularly for precipitation time series, one poor adjustment in the most recent years can lead to some miserable adjustments throughout the record because the offsets are multiplicative rather than additive, such as in the temperature time series adjustments.

Acknowledgments. This work was supported by the Carbon Dioxide Research Division of the Department of Energy under Contract AC02-DE-FG02-85ER60372, Interagency Agreement DE-AI05-850-R21501, and the National Oceanic and Atmospheric Administration.

REFERENCES

Agee, E. M., 1982: A diagnosis of twentieth century temperature records at West Lafayette, Indiana. *Climatic Change,* **4,** 399–418.

Alexanderson, H., 1984: A Homogeneity Test Based on Ratios and Applied to Precipitation Series. Report No. 79, Meteorologiska Institutionen, Kungl Universitete, Sweden.

——, 1986: A Homogeneity Test Applied to Precipitation Data. *J. Climatol.,* **6,** 661–675.

Barnett, T. P., 1985: Long-term changes in precipitation patterns. *Detecting the Climate Effects of Increasing Carbon Dioxide.* M. C. MacCracken and F. M. Luther, Eds., DOE/ER-0235. [Available from National Technical Information Service, Springfield, VA 22161.]

Blasing, T. J., and D. Duvick, 1984: Reconstruction of precipitation history in the North American corn belt using tree rings. *Nature,* **307,** 143–145.

Brinkman, W. A. R., 1983: Variability of temperature in Wisconsin. *Mon. Wea. Rev.,* **111,** 172–180.

Brower, W. A., 1985: U.S. historical climatology network. *Third Conf. on Climate Variations and Symposium on Contemporary Climate: 1850–2100.* Amer. Meteor. Soc.

Carlson, T. N., 1986: Regional-scale estimates of surface moisture, availability and thermal inertia using remote thermal measurements. *Remote Sensing Reviews,* **1,** 197–247.

Conrad, V., and L. W. Pollak, 1950: *Methods in Climatology.* Harvard University Press, 459 pp.

Devore, J. L., 1982: *Probability and Statistics for Engineering and the Sciences.* Brooks/Cole, 640 pp.

Diaz, H. F., and R. G. Quayle, 1980: The climate of the United States since 1895: Spatial and temporal changes. *Mon. Wea. Rev.,* **108,** 249–266.

Fritts, H. C., T. J. Blasing, B. P. Hayden and J. E. Kutzbach, 1971: Multivariate techniques for specifying tree growth and climate relationships for reconstructing anomalies in paleoclimate. *J. Appl. Meteor.,* **10,** 845–864.

Jones, P. D., S. C. B. Raper, R. S. Bradley, H. F. Diaz, P. M. Kelly and T. M. L. Wigley, 1986: Northern Hemisphere surface air temperature variations 1851–1984. *J. Climate Appl. Meteor.,* **25,** 161–179.

Kalma, J. D., G. P. Laughlin, A. A. Green and M. T. O'Brien, 1987: Minimum temperature surveys based on near-surface air temperature measurements and airborne thermal scanner data. *J. Climatol.,* **6,** 413–430.

Karl, T. R., 1985: Perspective on climate change in North America during the twentieth century. *Phys. Geogr.,* **6,** 207–229.

——, R. E. Livezey and E. S. Epstein, 1984a: Recent unusual mean winter temperatures across the United States. *Bull. Amer. Meteor. Soc.,* **65,** 1302–1309.

——, G. Kukla and J. Gavin, 1984b: Decreasing diurnal temperature range in the United States and Canada from 1941 through 1980. *J. Climate Appl. Meteor.,* **23,** 1489–1504.

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

Kohler, 1949: Double-mass analysis for testing the consistency of the records and for making required adjustments. *Bull. Amer. Meteor. Soc.,* **30,** 188–189.

Kukla, G., J. Gavin and T. R. Karl, 1986: Urban warming. *J. Climate Appl. Meteor.,* **25,** 1265–1270.

Maronna, R., and V. J. Yohai, 1978: A bivariate test for the detection of a systematic change in mean. *J. Amer. Statist. Assoc.,* **73,** 640–645.

Mitchell, J. M., Jr., 1953: On the causes of instrumentally observed secular temperature trends. *J. Meteor.*, **10**, 244–261.

——, 1958: Effects of changing observation time on mean temperature. *Bull Amer. Meteor. Soc.*, **39**, 83–89.

——, 1961: The Measurement of Secular Temperature Change in the Eastern United States. Research Paper No. 43, U.S. Weather Bureau, Washington, DC.

Nelson, W. L., R. F. Dale and L. A. Schaal, 1979: Non-climate trends in divisional and state mean temperatures: A case study in Indiana. *J. Appl. Meteor.*, **18**, 750–760.

Oke, T. R., 1978: *Boundary Layer Climates.* Wiley and Sons, 372 pp.

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean precipitation series. *Mon. Wea. Rev.*, **109**, 2040–2045.

Schaal, L. A., and R. F. Dale, 1977: Time of observation bias and "climate change." *J. Appl. Meteor.*, **16**, 215–222.

Schlesinger, M. E., and J. F. B. Mitchell, 1985: Model projections of the equilibrium climate response to increased carbon dioxide. *Projecting the Climate Effects of Increasing Carbon Dioxide,* M. C. MacCracken and F. M. Luther, Eds., U.S. Dept. of Energy DOE/ER-0237.

Thompson, L. M., 1975: Weather variability, climatic change, and grain production. *Science,* **188**, 535–541.

van Loon, H., and J. Williams, 1976a: The connection between trends of mean temperature and circulation at the surface. Part I: Winter. *Mon. Wea. Rev.*, **104**, 365–380.

——, and ——, 1976b: The connection between trends of mean temperature and circulation at the surface. Part II: Summer. *Mon. Wea. Rev.*, **104**, 1003–1011.

Vinning, K. C., and J. F. Griffiths, 1985: Climatic variability at ten stations across the United States. *J. Climate Appl. Meteor.*, **24**, 363–370.

Webb III, T., and T. M. L. Wigley, 1985: What past climates can indicate about a warmer world. *The Potential Climatic Effects of Increasing Carbon Dioxide,* M. C. MacCracken and F. M. Luther, Eds., DOE/ER-0237. [Available from National Technical Information Service, Springfield, VA 22161.]

Wigley, T. M. L., J. K. Angell and P. D. Jones, 1985: Analysis of the temperature record. *Detecting the Climate Effects of Increasing Carbon Dioxide,* M. C. MacCracken and F. M. Luther, Eds., DOE/ER-0235. [Available from National Technical Information Service, Springfield, VA 22161.]

Williams, J., and H. van Loon, 1976: The connection between trends of mean temperature and circulation of the surface. Part III: Spring and autumn. *Mon. Wea. Rev.*, **104**, 1591–1596.