

NOTES AND CORRESPONDENCE

On the Development and Use of Homogenized Climate Datasets

DAVID R. EASTERLING, THOMAS C. PETERSON, AND THOMAS R. KARL

NOAA/NESDIS, National Climatic Data Center, Asheville, North Carolina

17 August 1994 and 6 December 1995

ABSTRACT

At the National Climatic Data Center, two basic approaches to making homogeneity adjustments to climate data have been developed. The first is based on the use of metadata (station history files) and is used in the adjustments made to the U.S. Historical Climatology Network monthly dataset. The second approach is non-metadata based and was developed for use with the Global Historical Climatology Network dataset, since there are not extensive station history files for most stations in the dataset. In this paper the two methodologies are reviewed and the adjustments made using each are compared, then the results are discussed. Last, some brief guidelines on the limitations and uses of these data are provided.

1. Introduction

Recently the need to develop homogeneous climate datasets has been expressed in a number of forums (e.g., Bradley and Jones 1985; Folland et al. 1990). Homogeneity of climate data is desirable, and even critical when datasets are used to study questions related to climatic fluctuations and change. Climate data homogeneity (true homogeneity) exists when variations in the data are a result only of variations in weather and climate (Conrad and Pollak 1962). However, it is very rare if not impossible to find a long-term climatic time series that is completely free of inhomogeneities. Furthermore, very few long-term climatic time series can even be considered "relatively homogeneous" (Conrad and Pollak 1962) where the fluctuations in the data are consistent with time series from surrounding stations. Therefore, there are only two possible approaches to developing a homogeneous dataset: 1) choose only stations that have been found to be relatively homogeneous or 2) apply data adjustment techniques to homogenize the data (in this paper we use the term homogeneous to refer to the quality of "relative homogeneity"). If the first choice is made, the dataset would likely have serious deficiencies in spatial coverage since few homogeneous stations are available. However, if the second choice is made, there will be questions regarding the validity of data adjustments.

At the National Climatic Data Center (NCDC), we have developed a number of climate datasets, in collaboration with the Carbon Dioxide Information and Analysis Center of the Oak Ridge National Laboratory, that are high quality (subjected to rigorous quality assurance procedures) and have had a series of data adjustments designed to improve the homogeneity of the data. These datasets include the U.S. Historical Climatology Network (USHCN; Karl et al. 1990) and the Global Historical Climatology Network (GHCN; Vose et al. 1992). As the developers of these datasets, we are often contacted by researchers who have questions regarding various aspects of the data. These questions range from the proper use of the data to the data adjustments and their validity. It is the purpose of this note to review the adjustment methodologies, why adjustments are necessary, and to attempt to provide guidelines regarding the use of these data. Although there are other, similar datasets (e.g., Jones et al. 1985, 1986), we do not intend to imply that our guidelines should be applied to the use of these other datasets; however, in many instances they are likely to be valid.

2. Creation of homogeneous datasets

Inhomogeneities manifest themselves on climatic time series in two basic ways: as a gradual trend or as a discontinuity (sharp change) in the mean or variance. Gradual trends may occur due to urban warming, instrument drift, or other effects that accumulate over time. However, discontinuities are usually due to an abrupt change occurring at the observing station, on the observing platform, or in the way the data are processed. These changes include instrument changes, as

Corresponding author address: Dr. David R. Easterling, NOAA/NESDIS, National Climatic Data Center, Federal Building, Asheville, NC 28801.
E-mail: deasterl@ncdc.noaa.gov

in the change from the liquid-in-glass thermometer (LIG) to the Maximum–Minimum Temperature System (MMTS) in the U.S. Cooperative Network (Quayle et al. 1991), station relocations, or changes in averaging methods for time-averaged quantities (for a more comprehensive review see Bradley and Jones 1985). The bias produced by these changes may be always positive or always negative [e.g., the change to the HO83 hygrothermometer at U.S. first-order stations (Gall et al. 1992)] or the effect may be random throughout a network, as in station relocations. Precipitation measurements are particularly susceptible to large inhomogeneities and biases. Groisman and Easterling (1994) and Groisman and Legates (1994) provide reviews of the problems with precipitation measurements, and propose some solutions to account for these problems. However, it is critical to keep in mind that precipitation measurements, particularly those for solid precipitation, can be biased by as much as 50% or more.

Adjustment techniques for inhomogeneities have been developed to address both gradual trends and discontinuities. Furthermore, these techniques may be based on metadata (station history) or they may be purely statistically (nonmetadata) based, but in each case the data are adjusted to the most recent homogeneous period of the record. In developing datasets at the NCDC, we use each approach depending on the dataset. The following is a brief review of the methodologies we have developed at NCDC.

In the USHCN, which contains monthly maximum, minimum, and mean temperature and total precipitation for 1221 stations in the contiguous United States, both abrupt discontinuities and trend inhomogeneities are addressed in four basic adjustment schemes. The first step in the processing of the USHCN is rigorous quality assurance where statistical outliers are hand checked against original records. After this, an adjustment is made to the temperature data to account for the time-of-observation bias as described by Karl et al. (1986). Next, we apply an a priori correction for the introduction of the MMTS using the bias values given in Quayle et al. (1991). The MMTS corrections are applied at this stage in the processing because it is a known average bias due to an instrument change; however, a random effect due to a station relocation may also accompany the effects of the instrument change. Thus, we apply the adjustment for the known effect first then allow the next step in the processing to account for any random effect due to a station move. This next step is to adjust individual temperature and precipitation records for known changes that are found in the station history records using statistical techniques and a reference series as described by Karl and Williams (1987). If adjustments cannot be made within the imposed confidence level, values are estimated using a methodology similar to that for the adjustments. If the data cannot be adjusted or estimated then they are des-

ignated as missing. Currently, the final step in the processing is to provide a file of urban adjustments for temperature in a regression-based scheme with population as the predictor (Karl et al. 1988).

The GHCN is a dataset of long-term average monthly temperature, total precipitation, station, and sea level pressure for stations from around the world (Vose et al. 1992). Unlike the USHCN, the GHCN does not have extensive station history (metadata) files to accompany each station. Although metadata files are available for subsets of the data (e.g., USHCN), more often the only metadata accompanying the time series are station location and identification information (e.g., country, latitude, and longitude). To provide a homogeneous temperature dataset for GHCN we have developed a nonmetadata adjustment scheme. The first step is to apply a methodology for creating homogeneous reference series for each candidate station as described in Peterson and Easterling (1994). This approach relies heavily on the first-difference series that is created by subtracting the previous years data from the current year's ($FD_i = T_i - T_{i-1}$). Using the first difference series greatly diminishes the effects of discontinuities on correlation analysis and forming the reference series by isolating the effect of any given discontinuity to the year in which it occurs. The reference series is then used in a statistically based scheme for detecting undocumented discontinuities (Easterling and Peterson 1995a). The homogeneous reference time series is created for each station as a function of a few neighboring stations. Statistical comparisons are made between the reference series and the time series for the station in question (candidate). If statistically significant discontinuities are found, adjustment values are determined and applied to create a homogeneous time series.

The magnitude of the adjustments necessary to create a homogeneous dataset can be relatively large, particularly for precipitation. Furthermore, most rain gauges in use today undercatch "true" precipitation amounts by some factor, which makes adjustments for discontinuities at individual stations particularly difficult. Because of these problems and that the spatial autocorrelation function for precipitation falls off rapidly with distance making statistical comparisons between stations difficult, very few (<20%) of the stations in the current version of the USHCN have any adjustments to precipitation.

For homogeneous precipitation in the GHCN, we are combining the work from individuals in many countries (e.g., Groisman et al. 1990; Hanssen-Bauer and Forland 1994) to identify homogeneous precipitation stations to use or to supply data adjusted using metadata. In some parts of the world, precipitation homogeneity is impossible to verify due to a sparse station network and very limited metadata, but the effects of different gauge configurations are less important when dealing only with liquid precipitation (Karl et al. 1993).

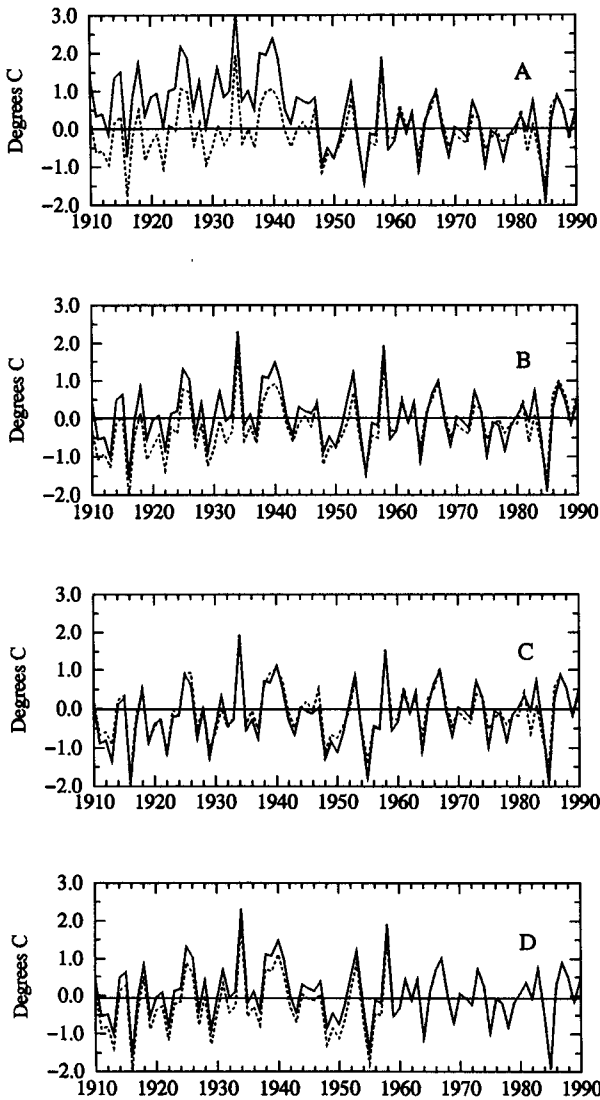


FIG. 1. (a) Annual average temperature for Spokane (solid) and USHCN region 5 (dotted), both unadjusted; (b) same as in (a) except that both were adjusted using USHCN metadata methodology; (c) same as in (a) except that both were adjusted using the nonmetadata methodology; (d) annual average temperature for Spokane adjusted using metadata methodology (solid) and nonmetadata methodology (dotted).

Certain types of adjustments are based on average effects of a change throughout a network of stations. A good example of this is the effect of changing from LIG thermometers to the MMTS in the U.S. Cooperative Network (Quayle et al. 1991). The bias values for this effect were determined through averaging over a large (>400 stations) network. Therefore, applying *only* these adjustments for the instrument change to individual stations is valid only when these data will be used to create area averages (time series or normals, etc.).

A major part of the methodology used for adjustments of both the USHCN and GHCN employs reference series for comparative purposes. The reference series is, in a sense, an area-averaged time series formed using a limited number of nearby stations and is used to form a difference (temperature) or ratio (precipitation) series between it and the candidate time series. For a homogeneous candidate, this difference (or ratio) series is white noise with no trend and only random fluctuations. The adjustment methodology is designed to produce a difference series between the reference series and adjusted candidate series that is white noise, and is based on the assumption that regional trends and fluctuations are reflected at all homogeneous stations in the region.

3. Discussion

Figure 1 shows a number of time series of mean annual temperature anomalies for Spokane, Washington, and for its region (the east slope of the northern Cascades) as defined in Karl et al. (1988). Homogeneity adjustments were performed on the data using two methodologies, the standard USHCN (metadata based) adjustments, and the nonmetadata adjustment scheme outlined above. The time series for the region, computed using the unadjusted data for all USHCN stations (35) in the region is nearly identical to Spokane from the 1960s to the present (Fig. 1a). However, a series of discontinuities due to instrument changes and station moves causes the Spokane time series to be much warmer prior to the 1960s, resulting in a overall linear trend that is strongly negative (Table 1). On the other hand, the regional trend for the unadjusted data is slightly positive, but not statistically significant.

The time series calculated from the data adjusted using the USHCN metadata based methodology (Fig. 1b) shows a distinct change in the trend characteristics both for Spokane and the region. There is now only a slight negative trend in the Spokane time series, but a statistically significant positive trend in the regional time

TABLE 1. Linear regression coefficients for unadjusted and adjusted time series.

	<i>r</i> Correlation	Slope (°C/100 yr)	Std error (°C/100 yr)
Spokane, unadjusted	-0.45	-1.70*	+/-0.30
Spokane, metadata adjusted	-0.04	-0.12	+/-0.11
Spokane, nonmetadata adjusted	0.16	0.50	+/-0.30
Region, unadjusted	0.08	0.20	+/-0.30
Region, metadata adjusted	0.27	0.80*	+/-0.30
Region, nonmetadata adjusted	0.10	0.50	+/-0.30

* Significant at .05 level.

series. The time series for Spokane adjusted using the nonmetadata methodology (Fig. 1c) shows a slight positive trend that is virtually identical to the regional trends.

It is clear from Fig. 1 that data adjustments do cause large changes both in the individual station and area-averaged time series. The preceding discussion also raises a number of questions regarding the differences in the unadjusted and adjusted data. First, why is the trend in the regional time series for the USHCN adjusted data considerably more positive than the unadjusted data, or for that matter the nonmetadata adjusted data and even the Spokane time series? The answer to this is mainly in the time-of-observation bias adjustments. Karl et al. (1986) have shown that changing the observing time from midnight to earlier in the day causes a distinct cool bias in monthly averaged temperatures. However, in the United States this bias occurs mainly at cooperative stations that take observations only once per day. Since the trend has been to earlier observing times due to hydrological forecasting needs, and the USHCN is composed mainly of cooperative stations, adjustments for this bias will cause area-averaged time series to have a more positive trend.

The changes to the time series at Spokane, however, are mainly due to adjustments for station moves and instrument changes. Spokane is a U.S. first-order station and has had a number of documented changes (see Table 2 in Easterling and Peterson 1995a) that caused the earlier years of the unadjusted time series to be much warmer.

If we consider the unadjusted time series in parts, the data observed during a period when no changes occurred at the observing station, or in the way the data were processed can be considered homogeneous since they are representative of the observing site, conditions, and methods in place during that period. However, as soon as a change occurs, the data become inhomogeneous and are now representative of new conditions. Any adjustments to the data can attempt to ensure only that the data are "relatively homogeneous," since the property of "true homogeneity" is irretrievably lost. It then follows that, if a station has undergone changes, as all stations have, and the data have been "homogenized" using a methodology like those described above, the time series for that station does not truly represent any one site, but is more like the composite of a number of sites.

This is particularly evident if we compare data adjusted using different methodologies. Since a discontinuity causes an unknown change in the statistical characteristics of the time series, and adjustments for random effects are generally made through comparison with reference time series, different methodologies can make different adjustments for the same discontinuity. This is shown by comparison of the two time series for Spokane adjusted using the two methodologies described above (Fig. 1d). The two time series are iden-

tical from 1960 to the present. However, they begin to diverge prior to 1960 due to the differences in the adjustments made with the two methodologies. It is clear that since both of the adjusted time series are closer to the regional time series than the unadjusted time series, both are more homogeneous than the unadjusted Spokane time series. However, it is also evident that the adjustments, although very similar, are not identical, resulting in slightly different trends.

The question of which adjusted time series is correct is impossible to answer satisfactorily since the only way to determine this would be by comparison with the "true homogeneous" time series for Spokane. However, if we consider the basis for adjustments, it follows

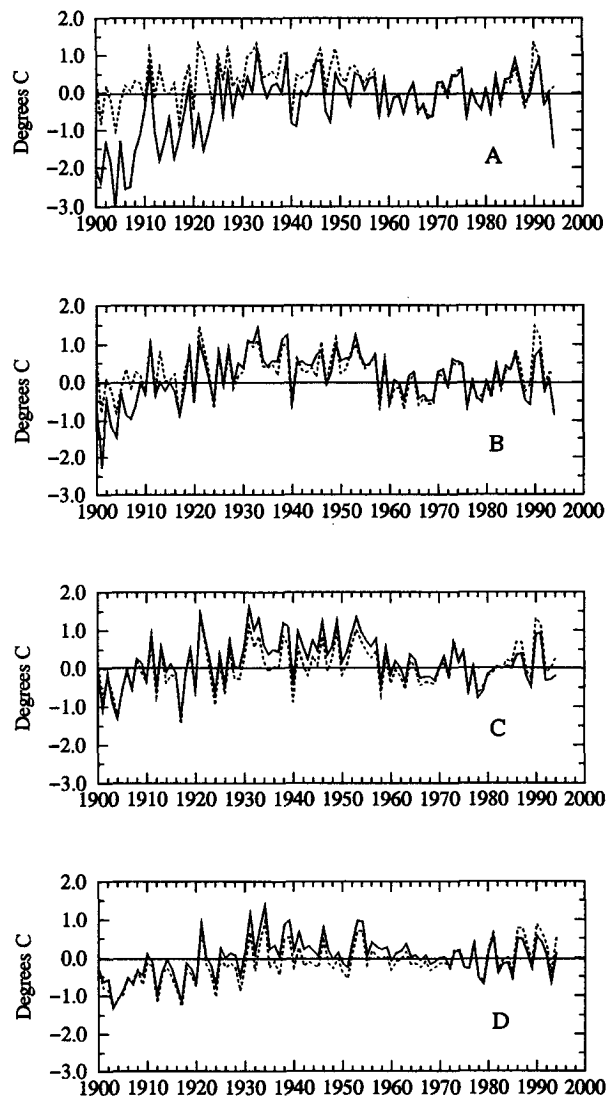


FIG. 2. Unadjusted (solid) and metadata-adjusted (dotted) area-averaged annual temperature from the USHCN for the (a) North Carolina mountains, (b) southern Appalachians, (c) eastern United States, (d) continental United States.

that a set of recommendations can be made regarding a hierarchy of preferred adjustment methodologies. The metadata approach used to adjust the USHCN data uses two basic assumptions: that the bias associated with known effects (e.g., the time-of-observation bias) can be estimated and applied with a reasonable degree of confidence, and that knowing the date of changes that can cause a discontinuity allows us to test that change for statistically significant biases. Thus, an approach based on prior knowledge is preferable. However, in many cases station history files (metadata) are incomplete or missing altogether, which requires a nonmetadata approach for homogeneity adjustments. In reality, even the most extensive station history files probably do not contain information on all changes at a station that can cause a discontinuity (e.g., simply replacing a broken thermometer). Therefore, it is likely that the optimal approach to homogeneity adjustments when metadata are available is a combined approach where both metadata and nonmetadata methods are applied. However, the question remains as to the proper application strategy if both are to be used.

As for the use of individual station data that have been subjected to homogeneity adjustments, we recommend that the researcher keep in mind the points of the preceding discussion. In particular, the adjusted time series represents an attempt to make the data appear as if they have been observed at one site under consistent methods and conditions when in reality those conditions may have existed only briefly during the most recent period. Due to the need for reference series comparisons to make most data adjustments, the adjusted time series often has statistical characteristics more consistent with the regional time series than with the unadjusted data for that same station.

Finally, it is worthwhile to consider the effects of homogeneity adjustments on time series for various-sized regions. Easterling and Peterson (1995b) showed that the differences between adjusted and unadjusted area-averaged time series decrease as the area size increases, particularly as the area size approaches the hemispheric scale. This is also shown by Figs. 2a–d. Figure 2a shows the unadjusted and metadata-adjusted area-averaged time series of mean annual temperature for the mountains of North Carolina constructed using the nine stations in the USHCN for the area. The two time series are similar from the 1950s to the present but become more dissimilar back in time such that they are different by about 1°C in the early 1900s. This difference earlier in the record could be due to a lack of data for some stations; however, this is not the case here since eight of the nine stations had data starting in 1900. Figure 2b shows the same two time series for the Southern Appalachian Mountains, a region including the North Carolina mountains (see the regionalization map in Karl et al. 1988). Here we see that the adjusted and unadjusted time series are different, but not as different as in the smaller region. Figures 2c and 2d show

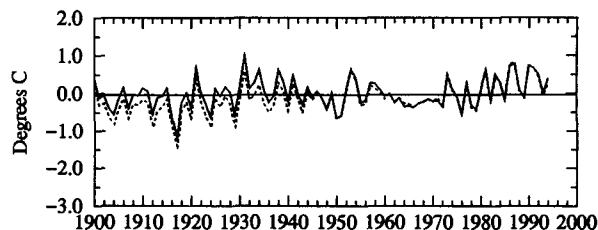


FIG. 3. Unadjusted (solid) and nonmetadata-adjusted (dotted) annual minimum temperature from stations with populations of over 50 000 for the United States and southern Canada. Data are from the Global Historical Climatology Network version 2 (Vose et al. 1995).

the area-averaged time series for the United States east of the Mississippi River and the United States as a whole. It is clear from these two figures, and the above discussion, that the difference between adjusted and unadjusted area-averaged time series diminishes as the area size increases to the point that they are virtually indistinguishable for large areas with many stations.

The reason this difference decreases as the number of stations averaged in increases is that many of the adjustments are for changes that are essentially random. However, when a systematic change occurs at nearly all the stations at approximately the same time, its effect is still present after large-scale averaging. An example of this is a subset of stations from the United States and southern Canada from cities with populations of over 50 000. A high percentage of these stations moved to airports in the 1930s and 1940s. This change shows up clearly on minimum temperatures despite large-scale averaging (Fig. 3).

It should be clear from the preceding discussion that using data that have had homogeneity adjustments can produce somewhat different results than using unadjusted data, particularly for smaller areas, and that adjustments may not remove all inhomogeneities. Therefore, care must be taken when using either the adjusted or unadjusted time series for an individual station, and it is our opinion that the optimal use of adjusted climatological time series is in the development of more robust regional analyses.

Acknowledgments. Partial support for this work was provided to the National Climatic Data Center by the U.S. Department of Energy through Interagency Agreements DE-AI05-90ER60952 and DE-AI05-900R21956.

REFERENCES

- Bradley, R., and P. D. Jones, 1985: Data bases for isolating the effects of increasing carbon dioxide concentration. *Detecting the Climatic Effects of Increasing Carbon Dioxide*, M. C. MacCracken and F. M. Luther, Eds., U.S. Dept. of Energy, Carbon Dioxide Research Division, 29–53.

- Conrad, V., and C. Pollak, 1962: *Methods in Climatology*. Harvard University Press, 459 pp.
- Easterling, D. R., and T. C. Peterson, 1995a: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369–377.
- , and —, 1995b: The effect of artificial discontinuities on recent trends in minimum and maximum temperatures. *Atmos. Res.*, **37**, 19–26.
- Folland, C. K., and Coauthors, 1990: Observed climate variation and change. *Climatic Change: The IPCC Scientific Assessment*, J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, Eds., Cambridge University Press, 199–235.
- Gall, R., K. Young, R. Schotland, and J. Schmitz, 1992: The recent maximum temperature anomalies in Tucson: Are they real or an instrumental problem? *J. Climate*, **5**, 657–665.
- Groisman, P. Ya., and D. R. Easterling, 1994: Variability and trends of precipitation and snowfall over the United States and Canada. *J. Climate*, **7**, 184–205.
- , and D. R. Legates, 1994: The accuracy of United States precipitation data. *Bull. Amer. Meteor. Soc.*, **75**, 215–227.
- , K. Ya. Vinnikov, and K. M. Lugina, 1990: Empirical data on contemporary global climate changes (temperature and precipitation). *J. Climate*, **3**, 662–677.
- Hanssen-Bauer, I., and E. J. Forland, 1994: Homogenizing long Norwegian precipitation series. *J. Climate*, **7**, 1001–1013.
- Jones, P. D., S. C. B. Raper, B. D. Santer, B. S. G. Cherry, C. M. Goodess, P. M. Kelly, T. M. L. Wigley, R. S. Bradley, and H. F. Diaz, 1985: A grid point surface air temperature data set for the Northern Hemisphere. U.S. Dept. of Energy, Carbon Dioxide Research Division, Tech. Rep. TR022, 251 pp.
- , —, B. S. G. Cherry, C. M. Goodess, and T. M. L. Wigley, 1986: A grid point surface air temperature data set for the Southern Hemisphere. U.S. Dept. of Energy, Carbon Dioxide Research Division, Tech. Rep. TR027, 73 pp.
- 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.
- , —, P. J. Young, and W. W. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. *J. Climate Appl. Meteor.*, **25**, 145–160.
- , H. F. Diaz, and G. Kukla, 1988: Urbanization: Its detection and effect in the United States climate record. *J. Climate*, **1**, 1099–1123.
- , C. N. Williams Jr., F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. Environmental Sciences Division Publ. 3404, 389 pp. [Available from Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831.]
- , R. G. Quayle, and P. Ya. Groisman, 1993: Detecting climate variations and change: New challenges for observing and data management systems. *J. Climate*, **6**, 1481–1494.
- Peterson, T. C., and D. R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol.*, **14**, 671–679.
- Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes, 1991: Effects of recent thermometer changes in the cooperative station network. *Bull. Amer. Meteor. Soc.*, **72**, 1718–1724.
- Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. R. Heim, T. R. Karl, and J. K. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea-level, and station pressure data. Environmental Sciences Division Publ. 3912, 189 pp. [Available from Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831.]
- , T. C. Peterson, R. L. Schmoyer, and J. E. Eischeid, 1995: The Global Historical Climatology Network, a preview of version 2. *Ninth Conf. on Applied Climatology*, Dallas, TX, Amer. Meteor. Soc., 59–64.