CONF-930133--2

DE92 041261

THE GLOBAL HISTORICAL CLIMATOLOGY NETWORK: LONG-TERM MONTHLY TEMPERATURE, PRECIPITATION, AND PRESSURE DATA

Russell S. Vose, CDIAC, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN
Thomas C. Peterson, National Climatic Data Center, Asheville, NC
Richard L. Schmoyer, Statistical Computing Office, Oak Ridge National Laboratory, Oak Ridge, TN
Jon K. Eischeid, Cooperative Institute for Research in Environmental Sciences, Boulder, CO
Peter M. Steurer, Richard R. Heim Jr., Thomas R. Karl, National Climatic Data Center, Asheville, NC

1. OVERVIEW

Interest in global climate change has risen dramatically during the past several decades. In a similar fashion, the number of data sets available to study global change has also increased. Unfortunately, many different organizations and researchers have compiled these data sets, making it confusing and time consuming for individuals to acquire the most comprehensive data.

In response to this rapid growth in the number of global data sets, DOE's Carbon Dioxide Information Analysis Center (CDIAC) and NOAA's National Climatic Data Center (NCDC) established the Global Historical Climatology Network (GHCN) project. The purpose of this project is to compile an improved data set of long-term monthly mean temperature, precipitation, sea level pressure, and station pressure for as dense a network of global stations as possible. Specifically, the GHCN project seeks to consolidate the numerous preexisting national-, regional-, and global-scale data sets into a single global data base; to subject the data to rigorous quality control; and to update, enhance, and distribute the data set at regular intervals. The purpose of this paper is to describe the compilation and contents of the GHCN data base (i.e., GHCN Version 1.0).

2. THE PROBLEM

Greenhouse gas concentrations in the atmosphere have increased markedly during the past several decades (Conway et al. 1988). Most simulations of climate change suggest that detectable, if not significant, modifications of global temperature, precipitation, and pressure patterns may accompany these rising concentrations. A number of studies (e.g., Jones et al. 1986; Bradley et al. 1987; Diaz et al. 1989; Jones et al. 1989; Karl et al. 1989; Vinnikov et al. 1990) have examined the "global climate record" of the past 50–150 years in order to verify the presence (or absence) of an enhanced greenhouse effect. Most of these studies have at best detected only a small long-term trend in global climate. It is conceivable that the greenhouse gas-induced changes, if in fact they have occurred, are extremely small. On the other hand, the present climate record may contain problems that limit its utility in the study of the problem.

A quick review of the literature reveals the existence of dozens of global climate data bases. Typically, they have been produced or are produced by numerous organizations and researchers, and the data bases may or may not be updated at regular or irregular intervals. In addition, the amount of quality control applied to each data base varies considerably. The following is a brief survey of the major long-term climatic data bases. The review is limited to data sets with monthly resolution.

The main source of historical data is a series of publications called the *World Weather Records* (WWR), which has been produced by three different organizations over the past 60 years [i.e., by the Smithsonian Institution (1927, 1934, 1947), the U.S. Weather Bureau (1959, 1967), and the National Environmental Satellite Data and Information Service (NESDIS) (1983, 1991)]. Except for the first version, which contains data through 1920, each issue of the WWR contains 10 years of data (i.e., 1921–1930; 1931–1940; 1941–1950; 1951–1960; 1961–1970; 1971–1980) for hundreds of first-order and cooperative weather stations worldwide.

A source of more recent data is a series of manuscript publications called *Monthly Climatic Data for the World* (MCDW), which has been produced by NESDIS since 1961. These publications consist primarily of first-order weather station data that have been exchanged via the Global Telecommunications System on a real-time basis.

The WWR and the MCDW archives are acquired at regular intervals by the National Center for Atmospheric Research (Spangler and Jenne 1990) and form a subset of a data base called the *World Monthly Surface Station Climatology* (WMSSC). In addition to WWR and MCDW reports, the WMSSC contains data acquired through the efforts of individuals or special projects. In all, nearly 4000 stations are available. The WMSSC is one of the most frequently used data bases for the study of global climate change (WMO 1990). Unfortunately, the WMSSC contains data from only a very small proportion of the roughly 40,000 worldwide stations that currently measure temperature and precipitation at least once a day (WMO 1983). Furthermore, most of the stations in the WMSSC are located at urban or airport sites and, in general, are unevenly distributed across the globe.

Numerous researchers (e.g., Wernstedt 1972; Bradley et al. 1985; Jones et al. 1986; Eischeid et al. 1991) have compiled their own global and hemispheric data sets for specific applications. These data sets typically combine data from the WMSSC with data from other sources to better sample data-sparse areas. Some data sets also include adjustments to compensate for changes in station location, instrumentation, and urbanization. However, many are subject to little quality control.

3. DATA BASE COMPILATION

The goal of the GHCN project is to compile a new global data base from numerous preexisting digital sources. The current version of the GHCN data base has the following general specifications:

- Contains station data (as opposed to gridded data)
- Contains temperature, precipitation, sea level pressure, and station pressure data with monthly resolution
- Contains time series that are at least 10 years in length, to facilitate the identification of data quality problems
- Contains coverage for global land areas
- Contains basic station inventory information, including country name, station number, station name, latitude, longitude, and elevation
- Contains data that have been quality assured using both manual and automated methods

GHCN Version 1.0 was compiled in roughly four stages, including data set acquisition, quality assurance of station inventories, duplicate station elimination, and preliminary quality assurance of climate data.

3.1 Data Set Acquisition

The GHCN data base was assembled from the various national-, continental-, and global-scale data bases listed in Table 1. Most of the global data sets in Table 1 are derived from the WMSSC and therefore contain many of the same stations (i.e., duplicates). However, each also includes previously undigitized data that either extends the records of WMSSC stations or consists of observations from additional stations. Similarly, most of the national- and continental-scale data sets in Table 1 contain numerous stations that have never been incorporated into a global data base. In addition, certain data sets, notably those from the USSR and the People's Republic of China (PRC), were only recently made available through bilateral data exchanges with NCDC and thus have not been readily accessible outside their respective countries.

TABLE 1

An inventory of sources from which the GHCN data base was compiled.

Data Base Name, Size, and Contributor
60-station temperature/precipitation data base for the PRC Institute of Atmospheric Physics, Chinese Academy of Sciences
277-station temperature/precipitation data base for Mexico A. Douglas, Creighton University
1219-station U.S. Historical Climatology Network data base Carbon Dioxide Information Analysis Center
223-station temperature/precipitation data base for the USSR Research Institute of Hydrometeorological Information, Russia
243-station temperature data base for the USSR Research Institute of Hydrometeorological Information, Russia
622-station precipitation data base for the USSR P. Groisman, National Climatic Data Center
65-station temperature/pressure data base for Southern Hemisphere sites T. Jacka, Australian Antarctic Division
1087-station African precipitation data base compiled by S. Nicholson National Climatic Data Center
967-station African historical precipitation data base (TD9799) National Climatic Data Center
1146-station Non-African historical precipitation data base (TD9799) National Climatic Data Center
5328-station precipitation data base for global land areas J. Eischeid, Cooperative Institute for Research in Environmental Sciences
1872-station temperature data base for global land areas P. Jones et al., University of East Anglia
3806-station World Monthly Surface Station Climatology data base W. Spangler and R. Jenne, National Center for Atmospheric Research
2411-station World Weather Records data bases (1961-70 & 1971-80) National Climatic Data Center
6775-station temperature/precipitation data base for global land areas Climate Analysis Center

3.2 Quality Assurance of Station Inventories

The second step in the compilation of the GHCN data base entailed the quality assurance of all station inventory parameters, or "metadata" (i.e., country names, station numbers, station names, latitudes, longitudes, and elevations). All such metadata quantities were verified or corrected, whenever possible, through comparison with third-party sources.

Most of the data sets contributing to the GHCN data base contained, in one form or another, the name of the country in which each station was located. The accuracy of each country "code" was verified through a digital comparison of each station's coordinates with the boundaries of the appropriate country. Inconsistencies were inspected manually and corrected where necessary. Stations lacking country metadata were assigned country codes using a similar digital comparison of the station's coordinates with the boundaries of all countries in the world. Country code assignments for island and border stations were also inspected manually and corrected where necessary.

The accuracy of each WMO number was assessed through a digital comparison of each station's metadata with its corresponding entry in the latest version of WMO Publication No. 9, Vol. A (Observing Stations). The WMO Vol. A contains a complete list of all WMO stations that collect meteorological data, their station numbers, names, latitudes, longitudes, and elevations. Stations with names or coordinates that were significantly different from the official WMO entries were inspected manually. Nearly 1000 station numbers were corrected as a result. Numerous station numbers were uncorrectable, however, because they represented stations that did not belong to the WMO network. In these cases, an artificial station number was developed, in part by using the number of the nearest active WMO station in the same country.

All metadata parameters were updated for stations with valid WMO numbers. Specifically, the original names, coordinates, and elevations supplied with each WMO station were replaced with the corresponding values for those parameters in the WMO Vol. A. Stations without valid WMO numbers could not have their metadata verified, because of the lack of reference documentation.

3.3 Elimination of Duplicate Stations

The GHCN data base was created from numerous preexisting archives which contained data for many of the same stations. Each duplicate station was digitally inspected to ensure that the data for each were in fact the same. Frequently, important differences were evident in allegedly identical sets of data. In general, the factors leading to the existence of nonidentical duplicates involve rounding conventions, keypunch problems, and data transcription errors. For example, precipitation data for Beijing were present in two different data sets (the Eischeid et al. global precipitation data base and the 60-station PRC data base). Figure 1 portrays the differences between these duplicate Beijing records. For most years, the precipitation totals are identical. However, during the early 1890s, the late 1920s, and the middle 1930s, they are markedly different. In addition, from 1970–1988, one of the data sets appears to round totals to the nearest millimeter, leading to small but perhaps important differences in precision.

The total number of duplicates to be eliminated was quite large. For example, before duplicates were eliminated, there were >16,000 temperature and 24,000 precipitation stations. After the duplicates were eliminated, there were ~6200 temperature and 8000 precipitation stations. Station numbers were used in the duplicate elimination process to reduce the number of computations. That is, all stations with a given WMO number were compared with one another. Similarly, all non-WMO stations that were nearest to a particular WMO station were compared with one another. Similarity was assessed by computing the mean absolute difference (MAD) between pairs of series. For example, in the case of Beijing (Fig. 1), MAD would be computed by summing the absolute differences between the two January 1890 values, the February 1890 values, etc., and dividing by the total number of months for which both stations had data at the same time. If the mean absolute difference exceeded a given threshold (0.1° C, 1.0 mm, 0.1 mb), then the stations were deemed different (nonzero thresholds were used to allow for rounding differences in each series).

There were roughly two to three duplicates for each unique WMO station. For about 70% of the WMO stations, all of the duplicates were identical. In these cases, the duplicates were merged to form longer series. The remaining WMO stations had duplicates that differed in

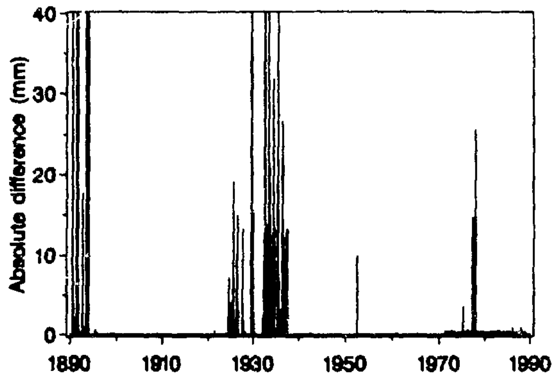

Figure 1. Plot of the monthly differences between duplicate Beijing precipitation records.

some nontrivial manner (e.g., Beijing). In these cases, the duplicate with the longest period of record was selected. This solution is arbitrary but unavoidable unless a time-consuming, manual, interstation comparison was to be applied to determine the station with the correct data. Furthermore, it is not expected to cause significant problems because it is extremely uncommon for duplicates with the same WMO number to be completely different. In most cases the difference is caused by gross data processing errors, though in some extreme cases, the difference may result from nearby stations (e.g., city versus airport) being confused with one another. The redundancy rate among non-WMO stations was much smaller (around 10%).

As an additional check, duplicate elimination was performed on a continent-by-continent basis. That is, all stations in a given continent were compared with one another to ensure that mislocated stations did not remain in the data base. Similarity was again assessed by computing the MAD between pairs of stations, only in finer 5-year periods to isolate "partial duplicates" (i.e., stations that were the same for some years but different for others). Only a small number of mislocated stations were identified as a result (roughly 200 temperature, 500 precipitation, 25 sea level pressure, and 25 station pressure stations).

3.4 Quality Assurance of Climate Data

The accuracy of all temperature, precipitation, sea level pressure, and station pressure data was assessed using a two-part quality assurance (QA) procedure. First, the data were digitally screened for extreme values. All observations exceeding the following thresholds, based in part on historical world record values, were set to missing:

- Mean monthly temperatures <-73.0° C or >58.0° C
- Monthly precipitation totals <0 mm or >9300 mm
- Mean monthly sea level pressures <900 mb or >1080 mb
- Mean monthly station pressures <500 mb or >1080 mb

Second, all time series were plotted and visually inspected for "gross" errors (e.g., months with missing negative signs). When possible, erroneous values were

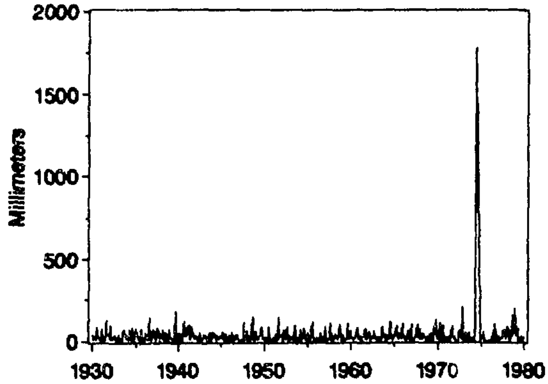

Figure 2. Plot of monthly precipitation totals through time for a station with a "gross" data processing error.

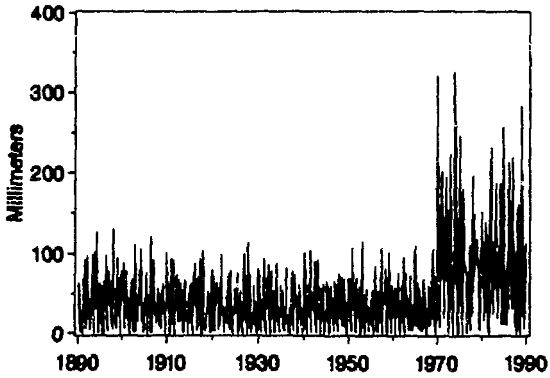

Figure 3. Plot of monthly precipitation totals through time for an inhomogeneous station.

revised and flagged. Others (e.g., Fig. 2) were set to missing. Observations that were unusual but not extreme enough to be obviously erroneous were flagged as suspect. Major discontinuities (e.g., Fig. 3) were also flagged.

4. COVERAGE OF THE DATA BASE

Data collection (as opposed to analysis) was emphasized during the first year of the project. As a result, the GHCN data base is considerably larger than most of its predecessors. For example, the GHCN data base contains roughly twice as many temperature and precipitation stations as the WMSSC (the number of sea level pressure and station pressure stations is roughly the same for both data bases). Furthermore, across all variables many of the stations in the GHCN data base have longer periods of record than their counterparts in the WMSSC.

Specifically, the GHCN data base contains 6,039 temperature, 7,533 precipitation, 1,883 sea level pressure, and 1,873 station pressure stations. All have at least 10 years of data, 40% have >50 years of data, and 10% have >100 years of data. The majority of stations have fairly complete records (72% are missing <10% of their observations). Furthermore, 80% of all records continue into the 1980s or 1990s.

For all variables the number of stations is relatively small and constant until the late 19th century, after which the number increases rapidly (Fig. 4). The sharp increase in the number of stations every 10 years (e.g., 1920, 1930, 1940, 1950, 1960) is due to the inclusion of various versions of the WWR data set in the WMSSC. The decrease in the number of stations after 1971 (for temperature and pressure) results from the inclusion of only three of the six volumes of the 1971–1980 WWR data set in the GHCN data base (i.e., three volumes have yet to be prepared and thus could not be included).

Spatial coverage is good across most of the globe for all variables (Fig. 5). However, for various historical and demographic reasons, certain areas are characterized by somewhat higher concentrations of stations than others. For example, the distribution of stations in North America and Europe is extremely high. In contrast, the number of stations in the Amazon basin, the Sahara desert, and the polar latitudes is quite low. Similarly, the high-density areas contain more stations with long records (>100 years) than the low density areas.

5. SUMMARY

At this point, GHCN Version 1.0 is considered one the largest and most comprehensive long-term data sets that can be applied to the study of global climate change. It has been constructed using records from numerous preexisting sources, thus consolidating a large amount of climatic data that had previously resided in disjoint repositories. It has also been subjected to several procedures designed to refine the quality of individual station records. In addition, it will continue to evolve in the coming years. Planned improvements entail the inclusion of additional data, the correction of erroneous data, the adjustment of data inhomogeneities, the addition of new variables, and the production of gridded data sets.

The GHCN data base is available, free of charge, as a numeric data package (NDP-041) from the Carbon Dioxide Information Analysis Center at Oak Ridge National Laboratory. It can be ordered by telephone, fax, electronic mail, or anonymous FTP:

Address: Carbon Dioxide Information Analysis Center
Oak Ridge National Laboratory
Post Office Box 2008
Oak Ridge, Tennessee 37831-6335
Phone: (615) 574-0390
FAX: (615) 574-2232
E-mail: BITNET: CDP@ORNLSTC
INTERNET: CDP@STC10.CTD.ORNL.GOV
OMNET: CDIAC
FTP: CDIAC.ESD.ORNL.GOV (128.219.24.36).

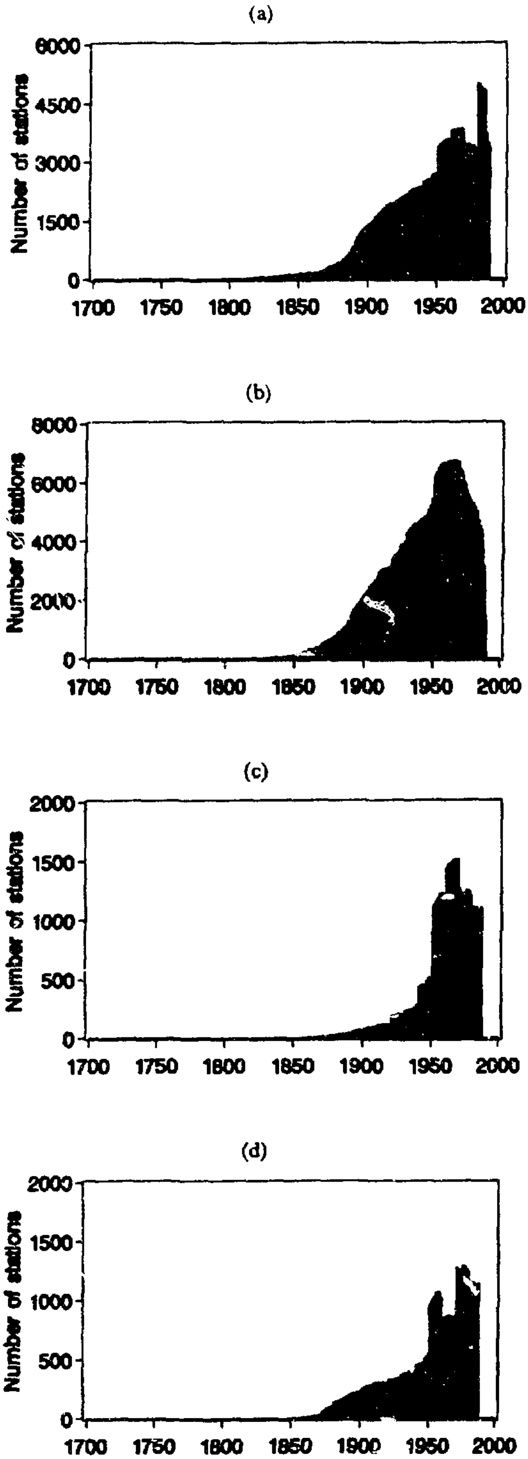

Figure 4. Number of GHCN (a) temperature, (b) precipitation, (c) sea level pressure, and (d) station pressure stations at various points in time.

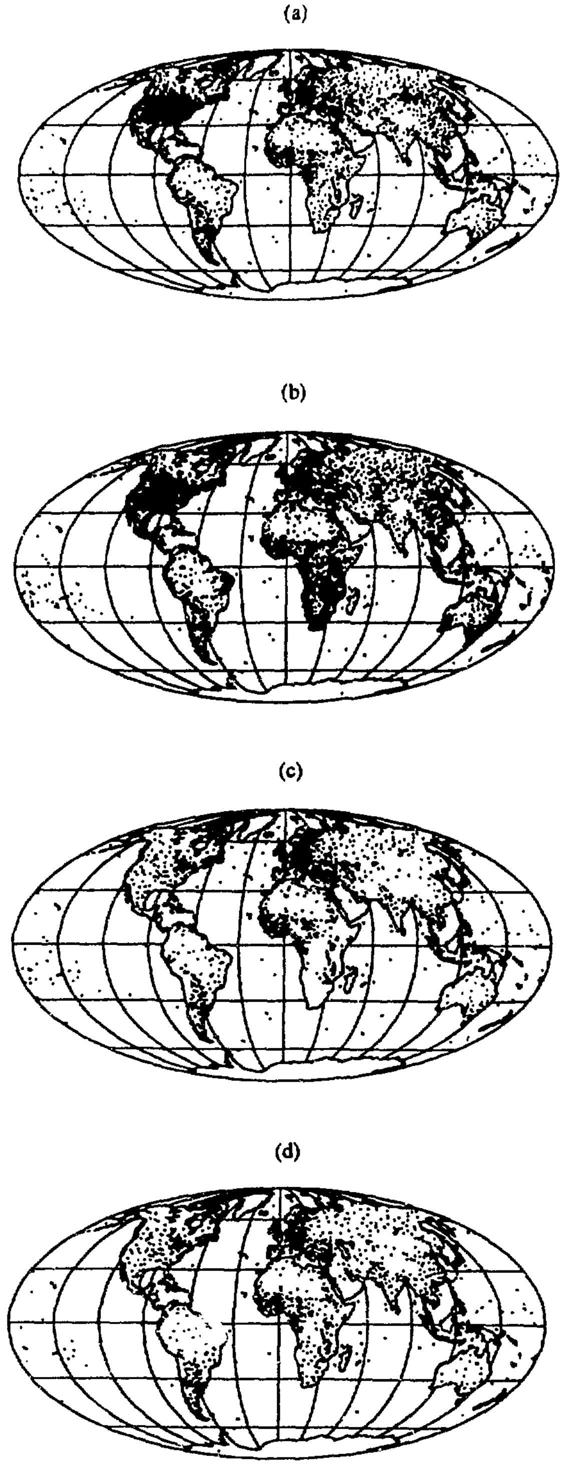

Figure 5. Global distribution of GHCN (a) temperature, (b) precipitation, (c) sea level pressure, and (d) station pressure stations. Not all stations were operating during the same time period (i.e., stations have different record lengths).

In the Spring of 1993, the data base will also be available from NCDC in Asheville, North Carolina through NOAA's Global Climate Perspective System. Contact Dr. Bruce Baker (BBAKER@NCDC.NOAA.GOV) for details.

Acknowledgements: CDIAC is sponsored by the U.S. Department of Energy, Office of Health and Environmental Research, Global Change Research Program. Oak Ridge National Laboratory is managed by Martin Marietta Energy Systems, Inc. for the United States Department of Energy under contract DE-AC05-84OR21400. Publication No. 3964, Environmental Sciences Division, ORNL.

REFERENCES

Bradley, R. S., P. M. Kelly, P. D. Jones, C. M. Goodess, and H. F. Diaz. 1985. A climatic data bank for Northern Hemisphere land areas, 1851-1980. DOE Technical Report No. TR017. Carbon Dioxide Research Division, U.S. Department of Energy, Washington, D.C.

Bradley, R. S., H. F. Diaz, J. K. Eischeid, P. D. Jones, P. M. Kelly, and C. M. Goodess. 1987. Precipitation fluctuations over Northern Hemisphere land areas since the mid-19th century. *Science* 237:171-175.

Conway, T. J., P. Tans, L. S. Waterman, K. W. Thoning, K. A. Masarie, and R. H. Gammon. 1988. Atmospheric carbon dioxide measurements in the remote global troposphere, 1981-1984. *Tellus* 40B:81-115.

Diaz, H. F., R. S. Bradley, and J. K. Eischeid. 1989. Precipitation fluctuations over global land areas since the late 1800s. *Journal of Geophysical Research* 94:1195-1210.

Eischeid, J. K., H. F. Diaz, R. S. Bradley, and P. D. Jones. 1991. A comprehensive precipitation data set for global land areas. DOE/ER-6901T-H1. Carbon Dioxide Research Division, U.S. Department of Energy, Washington, D.C.

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. *Journal of Climate and Applied Meteorology* 25:161-179.

Jones, P. D., P. M. Kelly, C. M. Goodess, and T. Karl. 1989. The effect of urban warming on the Northern Hemisphere temperature average. *Journal of Climate* 2:285-290.

Karl, T. R., J. D. Tarpley, R. G. Quayle, H. F. Diaz, D. A. Robinson, and R. S. Bradley. 1989. The recent climate record: What it can and cannot tell us. *Reviews of Geophysics* 27:405-430.

Karl, T. R., C. N. Williams, and F. T. Quinlan. 1990. United States Historical Climatology Network (HCN) serial temperature and precipitation data. NDP-019/R1. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

National Environmental Satellite Data and Information Service. 1961-present. Monthly Climatic Data for the World. U.S. Department of Commerce, Washington, D.C.

National Environmental Satellite Data and Information Service. 1981-1983. World Weather Records, 1961-1970 (Vols. 1-6). U.S. Department of Commerce, Washington, D.C.

National Environmental Satellite Data and Information Service. 1987-1991. World Weather Records, 1971-1980 (Vols. 1-3). U.S. Department of Commerce, Washington, D.C.

National Environmental Satellite Data and Information Service. 1989. TD9799: African Historical Precipitation. U.S. Department of Commerce, Washington, D.C.

National Environmental Satellite Data and Information Service. 1989. TD9799: Non-African Historical Precipitation. U.S. Department of Commerce, Washington, D.C.

Shiyan, T., Fu Congbin, Zeng Zhaomei, and Zhang Qingyun. 1991. Two long-term instrumental climatic data bases of the People's Republic of China. NDP-039. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Smithsonian Institution. 1927. World Weather Records, -1920. Miscellaneous Collections, Vol. 79. Washington, D.C.

Smithsonian Institution. 1934. World Weather Records, 1921-1930. Miscellaneous Collections, Vol. 90. Washington, D.C.

Smithsonian Institution. 1947. World Weather Records, 1931-1940. Miscellaneous Collections, Vol. 105. Washington, D.C.

Spangler, W. M. L., and R. L. Jenne. 1990. World Monthly Surface Station Climatology (And Associated Data Sets). National Center for Atmospheric Research, Boulder, Colorado.

U.S. Weather Bureau. 1959. World Weather Records, 1941-1950. U.S. Department of Commerce, Washington, D.C.

U.S. Weather Bureau. 1967. World Weather Records, 1951-1960 (Vols. 1-6). U.S. Department of Commerce, Washington, D.C.

Vinnikov, K. Ya., P. Ya. Groisman, and K. M. Lugina. 1990. Empirical data on contemporary global climate changes (temperature and precipitation). *Journal of Climate* 3:662-677.

Wernstedt, F. L. 1972. World Climatic Data. Climatic Data Press, Lemont, Pennsylvania.

World Meteorological Organization. 1983. World Climate Data Program - 59: Plan for the World Climate Data Program. Geneva, Switzerland.

World Meteorological Organization. 1990. World Climate Data Program - 11: Report of the Expert Group on Global Baseline Datasets. Asheville, NC.