Symposium on Precipitation and Evaporation
Bratislava, Slovakia
September 20-24, 1993

THE GLOBAL HISTORICAL CLIMATOLOGY NETWORK PRECIPITATION DATA

by

Thomas C. Peterson and David R. Easterling
Global Climate Laboratory, National Climatic Data Center
Asheville, NC 28801 USA

Russell S. Vose
Carbon Dioxide Information Analysis Center
Oak Ridge National Lab, Oak Ridge, TN 37831 USA

and

Jon K. Eischeid
Cooperative Institute for Environmental Studies
University of Colorado, Boulder CO 80309 USA

Introduction

Several years ago, in response to growing concern about global climate change, the U.S. National Climatic Data Center and the Carbon Dioxide Information Analysis Center undertook an effort to create a baseline global land surface climate data set called the Global Historical Climatology Network (GHCN, Vose *et al.*, 1992). GHCN was created by merging several large existing climate data sets into one data base. Fifteen separate data sets went into the creation of the GHCN version 1.0 (see Table 1). Merging the data sets was not an easy task: stations with the same name, number, latitude and longitude might have different data in different data sets indicating that they are actually two different stations while stations with different names, numbers, and locations may have identical data and therefore actually be the same station. Our quality control effort had several components, but the main part was examining plots of each time series for unusual features such as extreme outliers and obvious discontinuities.

GHCN version 1.0 was released in 1992. It has 7,533 precipitation stations, but the number of stations varies with time (see Figure 1). A slight majority (55%) have records in excess of 50 years, and a significant proportion (13%) have records in excess of 100 years. The longest period of record for any given station is 291 years (1697-1987 for Kew, United Kingdom). Figures 2, 3, and 4 show the geographic distribution of the stations in 1885, 1935, and 1985. As one might expect, the preponderance of the early data is in North America, Europe, and Australia. It is interesting to note that the coverage for much of Africa is better in 1935 than 1985. Much of the early African data can be attributed to the two special African data sets that went into GHCN. GHCN version 1 is

available from either

National Climatic Data Center
Research Customer Service Group
Asheville, NC 28801 USA
704-271-4994
email: tross@ncdc.noaa.gov

or

Carbon Dioxide Information Analysis Center
Oak Ridge National Laboratory
Oak Ridge, TN 37831 USA
615-574-0390
email: cdp@stc10.ctd.ornl.gov

Plans for GHCN version 2

The original intent of GHCN was to create a high quality, moderately dense, global historical station network suitable for global climate change analysis. But as time went on and we started getting feedback from various researchers, we realized that one problem many climate researchers face, particularly when studying the climate of less developed regions, is obtaining data. To create their research data sets they often had to search out data in several different archives. To eliminate this problem we decided to put almost all monthly land surface station temperature, pressure, and precipitation climate data we can obtain into GHCN. While a researcher examining global climate change might not want to deal with tens of thousands of stations, someone studying regional rainfall may want data from every possible station in his region.

We are currently in the process of putting together version 2 of GHCN. Version 2 will have many more stations owing to a number of additional regional data sets we can now include. Version 2 will also have more early stations in less developed regions of the world because we have been digitizing selected stations from European Colonial Archives. This will improve GHCN's geographic coverage during earlier years. We are also embarking on an effort to improve quality control. And lastly we are working at adjusting the data for inhomogeneities in the time series.

Dealing with inhomogeneities

The GHCN project's work on inhomogeneities started off by evaluating various techniques of detecting and adjusting discontinuities in station time series (Easterling and Peterson, 1992). Alexandersson's (1986) and Potter's (1981) techniques as well as a new technique developed by Easterling and Peterson were able to detect and adjust inhomogeneities when the reference series was good. The key ingredient, though, is the reference series. Creating an unbiased reference series for use in evaluating inhomogeneities in a station's time series is not an easy task. For example, high correlations may be due in part to stations having similar discontinuities. There are, however, ways to avoid this problem when creating reference series and they work quite well with moderately dense temperature networks (Peterson and Easterling, 1993). However, temperature exhibits less spatial variability than precipitation.

The global nature and varying station density of GHCN further complicates reference series problems. Alexandersson (1986) created reference series from a network of 56 stations in a small part of southwestern Sweden where he found many stations with correlations higher than 0.9. But it is a very different task to build reference series for the 46 precipitation stations GHCN has in all of Mali and Niger which combined cover 2.4 million square kilometers. The greater distance between the stations makes the correlations lower. Also, strong gradients in precipitation further lower the correlations between stations. In addition, having a very high percentage of the precipitation resulting from convection cause even lower correlations. The combination of station density and regional characteristics of precipitation will prevent us from creating adequate reference series to evaluate precipitation time series for inhomogeneities in some parts of the world.

Another approach to inhomogeneity problems with precipitation is to use extensive metadata such as gauge exposure, wind shields, and wind speeds to determine adjustments for inhomogeneities as outlined in Grois-

man (1993). Groisman's approach focuses mainly on the problem of under catch of solid precipitation. Unfortunately, the GHCN project lacks the global metadata and human resources to apply this time consuming technique globally.

Because of the complexity of homogenizing global precipitation, our effort to produce a homogeneous GHCN precipitation subset will use a four pronged approach: (1) We will detect and adjust for inhomogeneities using a computational technique similar to Alexandersson's in regions such as northern Europe and northeastern North America where high station density and low spatial variability of precipitation allow the technique to be used. (2) We will incorporate adjusted regional precipitation data sets, such as those produced by Groisman, where possible. (3) We will numerically and/or visually examine single time series of precipitation to remove stations with obvious discontinuities. And (4) we will acknowledge that there are some regions of the world where we will be unable to adequately evaluate the homogeneity of precipitation data.

Acknowledgements

This work is supported by the National Climatic Data Center, NOAA's Climate and Global Change Program's Global Climate Perspective System, and the U.S. Department of Energy through Interagency Agreements : DE-AI05-900R21956 and DE-A105-90ER60952. CDIAC is sponsored by the U.S. Department of Energy, Office of Health and Environmental Research, Global Change Research Program. Oak Ridge National Lab is managed by Martin Marietta Energy Systems, Inc. for the U.S. Department of Energy under contract DE-AC05-84OR21400.

References

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

Easterling, D. and T. Peterson, 1992: Techniques for detecting and adjusting for artificial discontinuities in climatological time series: A review. *Fifth International Meeting on Statistical Climatology*, June 22-26, Toronto, Ont., Canada, J28-J32.

Groisman, P. 1993: Towards unbiased estimates of North American precipitation. *Eighth Symposium on Meteorological Observations and Instrumentation*, January 17-22, Anaheim, California, J43-J47.

Peterson, T. and D. Easterling, 1993: Creation of meteorological reference series. *Eighth Symposium on Meteorological Observations and Instrumentation*, January 17-22, Anaheim, California, J31-J34.

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

Vose, R. S., Richard L. Schmoyer, Peter M. Steurer, Thomas C. Peterson, Richard Heim, Thomas R. Karl, and J. Eischeid, 1992: *The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data*. ORNL/CDIAC-53, NDP-041. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Table 1. Inventory of data sets from which the GHCN data base was compiled

Data Set	Source/Contributor	Stations[a]
60-station temperature/precipitation data base for the People's Republic of China	Institute of Atmospheric Physics Chinese Academy of Sciences Beijing, China	60
277-station temperature/precipitation data base for Mexico	A. Douglas Creighton University	277
U.S. Historical Climatology Network Serial Temperature and Precipitation Data	Carbon Dioxide Information Analysis Center	1219
223-station temperature/precipitation data base for the Union of Soviet Socialist Republics	Research Institute of Hydrometeorological Information Obninsk, Russia	223
243-station temperature data base for the Union of Soviet Socialist Republics	Research Institute of Hydrometeorological Information Obninsk, Russia	243
622-station precipitation data base for the Union of Soviet Socialist Republics	P. Ya. Groisman National Climatic Data Center	622
65-station temperature/pressure data base for selected sites in the Southern Hemisphere	T. H. Jacka Australian Antarctic Division	65
African precipitation data base compiled by S. Nicholson	National Climatic Data Center	1087
TD9799: African Historical Precipitation Data	National Climatic Data Center	967
TD9799: Non-African Historical Precipitation Data	National Climatic Data Center	1146
A Comprehensive Precipitation Data Set for Global Land Areas	J. K. Eischeid et al. Cooperative Institute for Research in Environmental Sciences	5328
1872-station temperature data base for global land areas	P. D. Jones et al. University of East Anglia	1872
World Monthly Surface Station Climatology	W. Spangler and R. Jenne National Center for Atmospheric Research	3806
World Weather Records (1961-1970 & 1971-1980)	National Climatic Data Center	2411
6775-station temperature/precipitation data base for global land areas	Climate Analysis Center	6775

[a]Denotes the maximum number of stations; each data set can have a different number of stations for each variable.

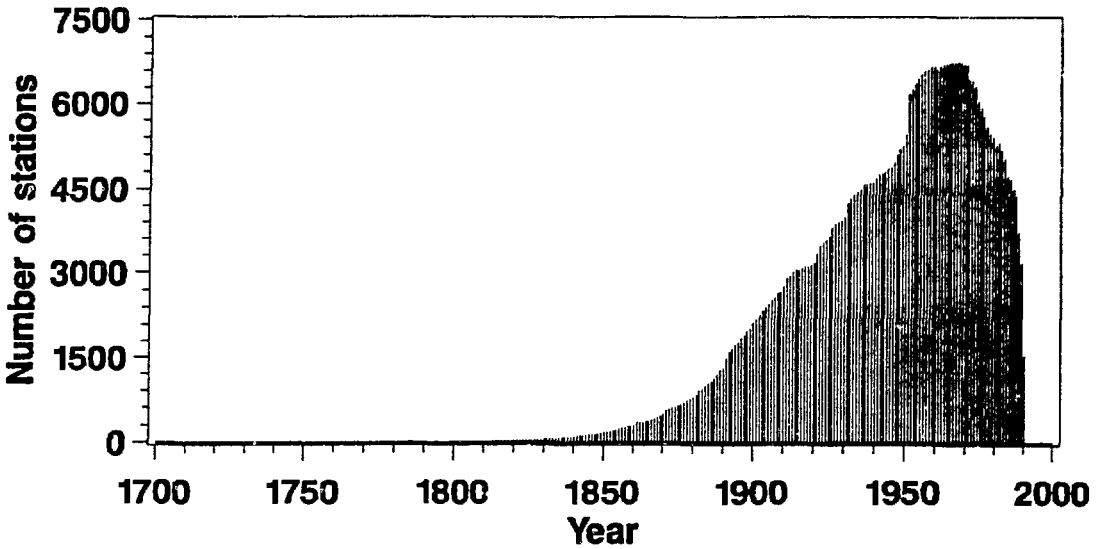

Figure 1. Plot of the number of precipitation stations by year from 1700-1990.

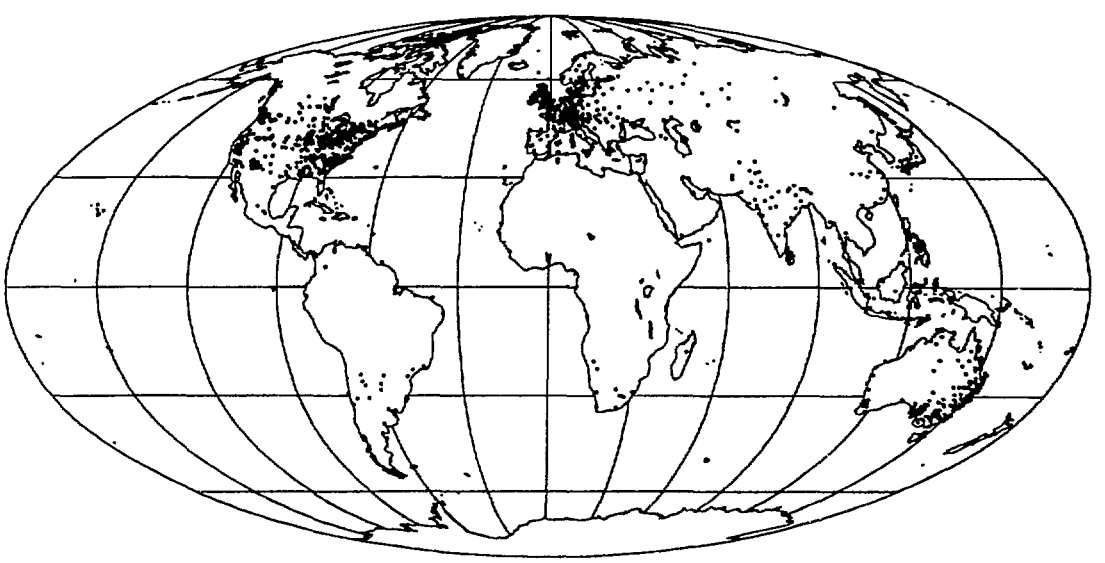

Figure 2. Locations of precipitation stations in 1885.

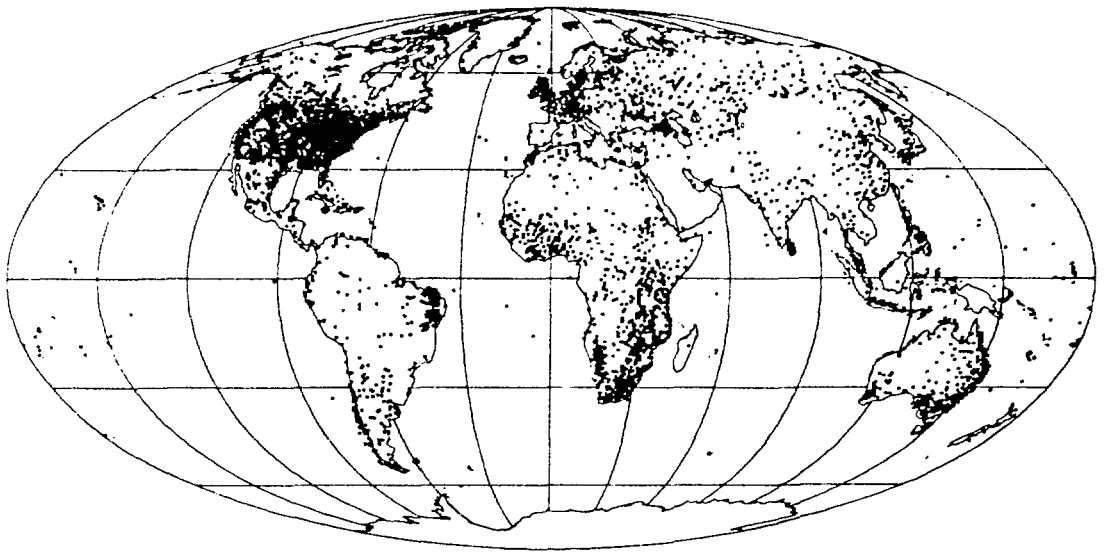

Figure 3. Locations of precipitation stations in 1935.

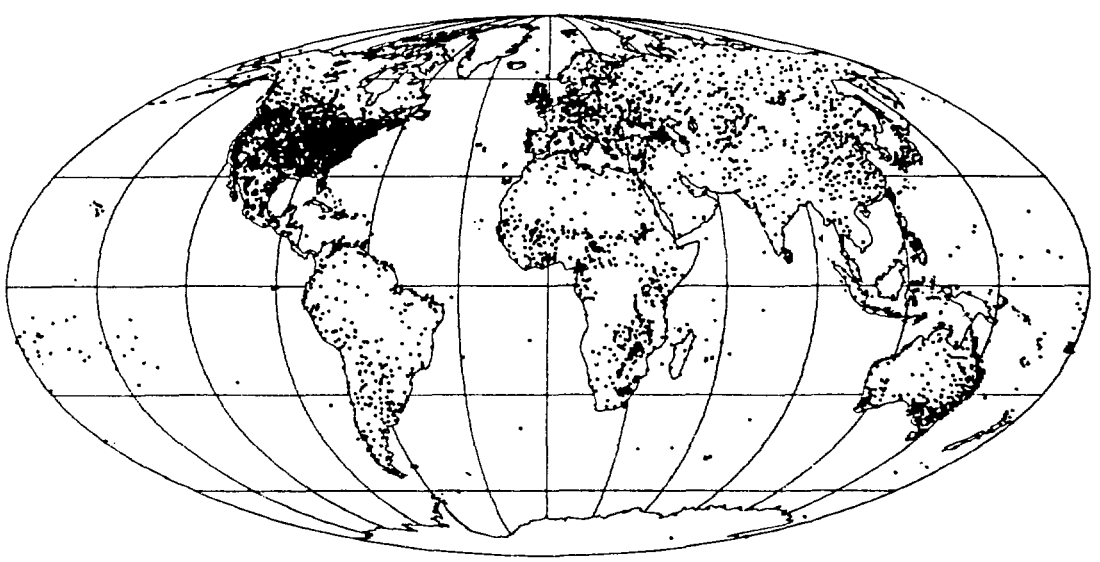

Figure 4. Locations of precipitation stations in 1985.

DISCLAIMER