Summary of Recent Changes in the GHCN-M Temperature Dataset and Merged Land-Ocean Surface Temperature Analyses

2 May 2011

Global Historical Climatology Network-Monthly Dataset:
The Global Historical Climatology Network-Monthly (GHCN-M) has been the official land surface mean temperature dataset since its release and has been widely used in several international climate assessments, as well as NCDC's climate monitoring activities. Effective May 2, 2011, the GHCN-M version 3 dataset of monthly mean temperature replaced GHCN-M version 2 monthly mean temperature dataset. Beginning with the April 2011 State of the Climate Report, GHCN-M version 3 will be used for NCDC climate monitoring activities, including calculation of global land surface temperature anomalies and trends. It will also be merged with the Extended Reconstruction Sea Surface Temperature (ERSST) version 3b dataset to form the merged land and ocean surface temperature dataset, which is used to calculate the global average temperature from 1880 to present.

The GHCN-M version 3 monthly mean temperature dataset introduces a number of changes that include: consolidating "duplicate" series, updating records from recent decades, and the use of new approaches to quality assurance and homogenization (the process of removing the impact of non-climatic changes in climate time series). These improvements have enhanced the overall quality of the dataset; nonetheless, conclusions regarding global land surface temperature change are little affected by this release.

Description of Major Changes in Version 3

Removal of Station Duplicates: A unique feature of the GHCN-M version 2 dataset is the presence of duplicate station records for approximately one-third of its stations. The dataset contains 2,706 stations that have two or more separate sets of observations informally referred to as "duplicates." The term notwithstanding, the two or more duplicate mean temperature series attributed to a single station are, in fact, similar *but not exact* copies of each other. Duplicates occur because there are often multiple sources of temperature data for any given observing station. For some stations included in GHCN-M version 2, data attributed to a single station were provided in ten or more different databases. These various sources of data often overlap in time, and while the values between sources are generally similar, they are often not identical. The differences most commonly result from the many different ways in which monthly mean temperature can be calculated. In GHCN-M version 3, duplicates are combined into single station series based on a process whereby the longer duplicate time series were given higher preference.

Data Additions to the GHCN-M database: In GHCN-M version 3, additions to the historical record were made to fill in data gaps during the 1990s and first decade of the 21st century by incorporating the most recently available data from World Weather Records (WWR) as well as additional data from NCDC's Monthly Climatic Data of the World (MCDW). Inclusion of observations from WWR and MCDW made it possible to increase by nearly 500, the number of existing GHCN-M stations having at least 9 months of data each year during the 1990s.

Changes to the Quality Control Process: The GHCN-M version 3 quality control checks can be grouped into three general categories: basic integrity, outlier, and spatial consistency. Once an observation fails a quality control check, the value is excluded from subsequent checks during that processing cycle. The quality control flags are included in the version 3 dataset for any value identified to be in error, providing information on the type of error associated with a value. The quality control flag is one of three types of metadata information included in the GHCN-M version 3 dataset. It is appended to each observation along with a measurement flag and a source flag. Details on the quality control, measurement, and source flags are available in the version 3 README file.

Bias Corrections: Surface weather stations are frequently subject to minor relocations throughout their history of operation. They may undergo changes in instrumentation, observing practices may vary through time, and the land use/land cover in the vicinity of an observing site can be altered by either natural or man-made causes. Any of these kinds of modifications to the circumstances behind temperature measurements have the potential to alter a thermometer's microclimate exposure characteristics or otherwise change the bias of measurements relative to those taken under previous circumstances. This can result in an abrupt shift in the mean level of temperature readings that is unrelated to true climate variations and trends. The removal of the impact of these non-climatic changes in climate series is called homogenization. The process of homogenization of the GHCN-M version 3 data is conducted through use of the Pairwise Homogeneity Adjustment algorithm. This was initially applied to the U.S. Historical Climatology Network Version 2 dataset and is described in Menne and Williams, 2009 and Menne et al. 2009.

Merging of GHCN-M Version 3 Land Surface Temperatures with ERSST Version 3b:
The GHCN-M version 3 station land surface temperature anomalies are averaged within a 5˚ by 5˚ grid box to obtain gridded anomalies, which are then merged with the ERSST version 3b 5˚ by 5˚ gridded sea surface temperature anomalies (see next paragraph) to get a more complete picture of global temperature variability and trends.

Even during the periods of greatest station density, there are many areas where land surface observations are unavailable. Observations can be estimated in such areas using a variety of interpolation techniques. A method developed by van den Dool et al. (2000) and applied to the development of a merged land and ocean surface temperature dataset (Smith et al. 2008) is used to estimate temperature anomalies in areas with little-to-no data. This method uses spatial pattern recognition (Empirical Orthogonal Teleconnections) to fill in the areas with little-to-no data, and it forms the basis for global surface temperature calculations used in NCDC's climate monitoring activities.

In high latitude areas, the method of spatial pattern recognition is less effective at filling in areas with sparsely reported climate observations. Prior to GHCN-M version 3, the global merged land and ocean surface temperature gridded dataset would set these areas to missing, regardless of whether land-only data was available for the grid point. With the release of GHCN-M version 3, the land surface temperature observations will be included in the merged dataset in these high-latitude areas.

REFERENCES:

Menne, M.J., and C.N. Williams, Jr. (2009), Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700-1717.

Menne, M.J., C.N. Williams, and R.S. Vose (2009), The United States Historical Climatology Network Monthly Temperature Data - Version 2. *Bulletin of the American Meteorological Society*, **90**, 993-1107.

Smith, T. M., R.W. Reynolds, T.C. Peterson, and J.H. Lawrimore (2008), Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006). *J. Climate*, **21**, 2283-2293. doi: 10.1175/2007JCLI2100.1

van den Dool, H. M., S. Saha, Å Johansson (2000), Empirical Orthogonal Teleconnections. *J. Climate*, **13**, 1421–1435.