An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3

Jay H. Lawrimore,[1] Matthew J. Menne,[1] Byron E. Gleason,[1] Claude N. Williams,[1] David B. Wuertz,[1] Russell S. Vose,[1] and Jared Rennie[2]

Received 30 April 2011; revised 11 July 2011; accepted 15 July 2011; published 15 October 2011.

[1] Since the early 1990s the Global Historical Climatology Network-Monthly (GHCN-M) data set has been an internationally recognized source of data for the study of observed variability and change in land surface temperature. It provides monthly mean temperature data for 7280 stations from 226 countries and territories, ongoing monthly updates of more than 2000 stations to support monitoring of current and evolving climate conditions, and homogeneity adjustments to remove non-climatic influences that can bias the observed temperature record. The release of version 3 monthly mean temperature data marks the first major revision to this data set in over ten years. It introduces a number of improvements and changes that include consolidating "duplicate" series, updating records from recent decades, and the use of new approaches to homogenization and quality assurance. Although the underlying structure of the data set is significantly different than version 2, conclusions regarding the rate of warming in global land surface temperature are largely unchanged.

Citation: Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie (2011), An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3, *J. Geophys. Res.*, *116*, D19121, doi:10.1029/2011JD016187.

1. Introduction

[2] The first version of the Global Historical Climatology Network-Monthly (GHCN-M) data set was released in 1992. GHCN-M was built upon earlier data collection efforts including the decadal volumes of *World Weather Records* [*Clayton*, 1927] and the National Center for Atmospheric Research (NCAR) World Monthly Surface Station Climatology. Compiled from these sources plus 13 other data sets, version 1 included monthly mean temperature summaries from approximately 6000 land surface stations worldwide [*Vose et al.*, 1992]. At the time of its release, GHCN-M was the most comprehensive monthly land surface database publicly available.

[3] Efforts to acquire additional data sources through personal contacts and bilateral agreements continued through the mid-1990s and led to the release of version 2 in 1997 [*Peterson and Vose*, 1997]. Version 2 of GHCN-M, compiled from 31 source data sets, substantially improved coverage in data-sparse regions of the world relative to version 1, and expanded the total number of stations with mean monthly temperatures to 7280. Relative to version 1, version 2 also included a number of data processing enhancements. These enhancements included: (1) routine data updates for over 2000 stations; (2) additional station metadata such as population, vegetation, and topography; (3) a more comprehensive set of quality assurance checks; and (4) an assessment of and adjustments to account for inhomogeneity in mean temperature series (i.e., automated checks for systematic changes in bias caused by station moves, instrument changes, and other artifacts not related to climate).

[4] Given the comprehensive nature of the data set and the automatic mechanisms for updates, GHCN-M version 2 quickly became the source of land surface temperature records in systems developed at the National Aeronautics and Space Administration (NASA) [e.g., *Hansen et al.*, 1999, 2010] and the National Oceanic and Atmospheric Administration (NOAA) [*Quayle et al.*, 1999; *Smith et al.*, 2008] for monitoring changes in global surface temperatures worldwide. It also became an essential part of national and international efforts for assessing climate change [e.g., *Intergovernmental Panel on Climate Change* (*IPCC*), 2001, 2007; *Karl et al.*, 2009]. In addition to monthly mean temperature, GHCN-M version 2 included 4966 stations in two data sets of monthly mean maximum and minimum temperature as well as more than 20,000 stations with monthly total precipitation.

[5] While GHCN-M version 2 continued to be the largest monthly land surface temperature database available throughout the first decade of the 21st century, many aspects of the monthly temperature data processing at NOAA's National Climatic Data Center (NCDC) have remained unchanged since its inception. Here we describe some recent changes to the GHCN-M temperature processing system that simplify and/or improve upon various aspects of the construction of

[1]National Climatic Data Center, NOAA, Asheville, North Carolina, USA.

[2]Cooperative Institute for Climate and Satellites, National Climatic Data Center, NOAA, Asheville, North Carolina, USA.

0148-0227/11/2011JD016187

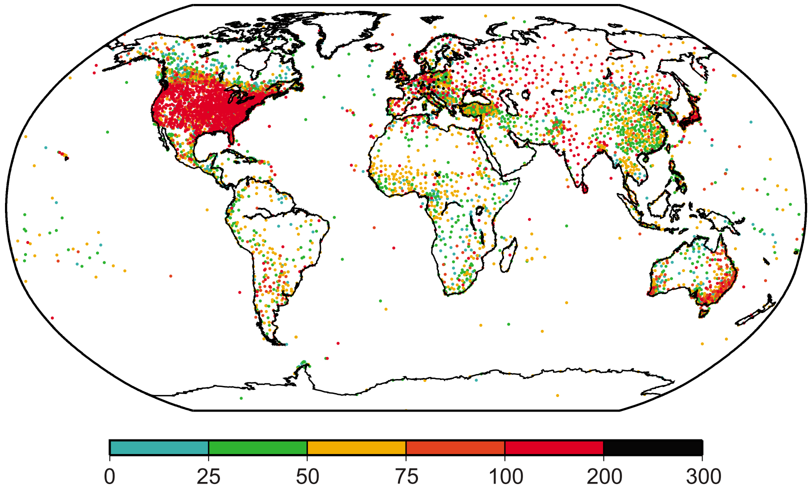

Figure 1. Location of the 7280 stations in the GHCN-M inventory. The color corresponds to the number of years of data available for each station.

the monthly mean temperature data set. Although this effort focuses on monthly mean temperature, monthly mean maximum and minimum temperature have benefited from the application of bias correction and quality control practices as described for monthly mean temperature, but improvements including updates and additions to the historical record of maximum and minimum temperature data will be the focus of follow-on efforts.

[6] The changes to monthly mean temperature include consolidating "duplicate" series, updating records from recent decades, and implementing new approaches to quality assurance and homogenization. They are described in section 3, and a brief overview of the GHCN-M version 3 processing system is described in section 4. In section 5, comparisons between land surface temperature trends calculated from versions 2 and 3 and from the raw and bias-corrected versions of the data sets are presented. Some concluding remarks are given in section 6.

2. Description of GHCN-M Version 2

2.1. Temporal and Spatial Coverage

[7] The 7280 stations that comprise GHCN-M version 2 monthly mean temperature provide a historical and modern-day record of instrumental observations of surface air temperature on every continent (Figure 1). Although station coverage varies greatly both spatially and temporally throughout the period of record, the data set provides decadal and century-scale climate perspectives at local, regional, and global scales. For this reason, the version 2 data set is retained as the foundation for version 3.

[8] Monthly mean temperatures can be calculated as an average of daily observations (e.g., the average of daily maximum and minimum temperature) or from daily averages of observations collected at various times during a day. These can include averages of observations taken every hour or an average of observations made at various times during a day (e.g., every three or six hours or at three fixed hours during the day). There are many other ways of calculating monthly mean temperature, and in many cases more than one method was used for calculating mean temperature for the same station (see section 2.2).

[9] For some stations and periods of time, especially before 1950, only monthly mean temperature data are available. This is also the case for many current observations which are transmitted via the Global Telecommunications System (GTS). Approximately half of the monthly mean temperature data transmitted as CLIMAT reports over the GTS are not available as daily observations. In GHCN-M version 3, CLIMAT reports are used as the source of current observations even if daily observations are also available (see sections 3.2 and 4).

[10] In some cases historical observations remain in paper records in the archives of National Meteorological and Hydrological Services. Because of data modernization efforts such as NCDC's Climate Database Modernization Program (CDMP; http://www.ncdc.noaa.gov/oa/climate/cdmp/cdmp.html) millions of paper forms with monthly, daily, and sub-daily observations (e.g., hourly, three-hourly, or synoptic) have been imaged over the past 10 years. However, while the forms have been preserved in electronic form, the data on the vast majority of the imaged forms have not been digitized, meaning the observations are not machine readable and therefore unavailable for data set development and analysis.

[11] Individual monthly mean temperature station records in GHCN-M version 2 begin as early as 1701, and observations from 69 countries and territories are available by 1880. The number of stations increased worldwide during the first half of the 20th century in association with a growing need for weather and climate observations to support aviation, agriculture, and other aspects of developing societies. This is reflected in a steady increase in the number of GHCN-M stations that peaks near 6000 stations in the 1960s and 1970s (Figure 2). The decline in the number of GHCN-M stations since the 1970s is due in part to station closures. For example in Russia, the number of reporting stations has decreased by at least 20% since 1990 (P. Groisman, personal communication, 2009). However, the decline primarily reflects the need to strengthen international data exchange efforts to capitalize to the greatest extent possible on the full set of observations

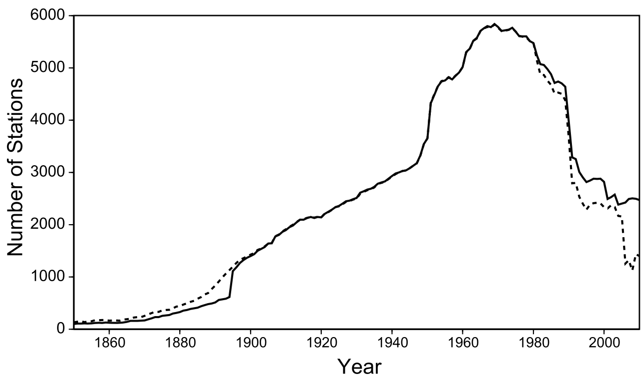

Figure 2. Number of Stations in GHCN-M version 2 (dashed line) and version 3 (solid line) with at least nine months of data each year from 1850 through 2010. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

collected by National Meteorological and Hydrological Services as well as other public and private sector entities worldwide [*Thorne et al.*, 2011]. Formal international data exchange is made possible by programs such as the World Meteorological Organization's CLIMAT program [*World Meteorological Organization* (*WMO*), 2009] which supports the worldwide transmission of monthly climate summaries from more than 2000 stations each month, most of which are included in the GHCN-M data set. Observations from other public and private networks are collected worldwide but are not currently distributed internationally on a regular basis.

2.2. Multiple Temperature Series Attributed to a Single Station

[12] A unique feature of the version 2 data set is the inclusion of duplicate station records for approximately one-third of its stations. The data set contains 2706 mean temperature stations that have two or more separate sets of observations. Although informally referred to as "duplicates," these are similar but not exact copies of each other.

[13] 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. As there is not one standard method for calculating monthly mean temperature, there is a high likelihood that monthly mean temperatures provided by one organization will be slightly different than those calculated by another for the same station. The similar but not necessarily identical duplicate station records made reconciling the multiple data sources into a single GHCN-M station record problematic [*Peterson and Vose*, 1997].

[14] In GHCN-M version 2, reconciliation of multiple source records involved automated processing to identify overlapping station records that shared the same value (to tenths of a degree Celsius) at least 90% of the time and when metadata indicated that the source records should be attributed to the same station. Data sources meeting these criteria were merged into a single station series. However, because monthly mean temperature can be calculated in so many different ways, it was possible to only identify about 10,000 out of 30,000+ time series as station duplicates using automated methods [*Peterson and Vose*, 1997]. The remaining time series were evaluated manually to identify probable duplicates.

[15] In cases where the data were considered likely but not unquestionably from the same station, the duplicate sources were assigned to the same station number instead of being merged, but with the addition of a 12th digit to the 11-digit GHCN-M station identifier to distinguish the station duplicates from each other [*Peterson and Vose*, 1997]. Of the 2706 mean temperature stations containing one or more duplicates, the mean difference in temperature during periods of overlap was less than 1°C for approximately 98.5% of the observations as shown in Figure 3. To simplify temperature analysis and open the data set to a wider group of users, version 2 duplicates were combined into single station series in version 3 as described in section 3.1.

2.3. Version 2 Quality Control and Bias Correction

[16] GHCN-M version 2 consists of a data set of quality controlled raw observations and a second data set corrected to remove inhomogenities caused by factors such as station relocations, changes in instrumentation, and environmental changes. Quality control of the version 2 data set took place initially during construction of the data set and then as part of routine operational processing. Three stages of quality control were used during data set development. The first was aimed at identifying and removing sources of low quality; for example bias corrected observations with no original raw

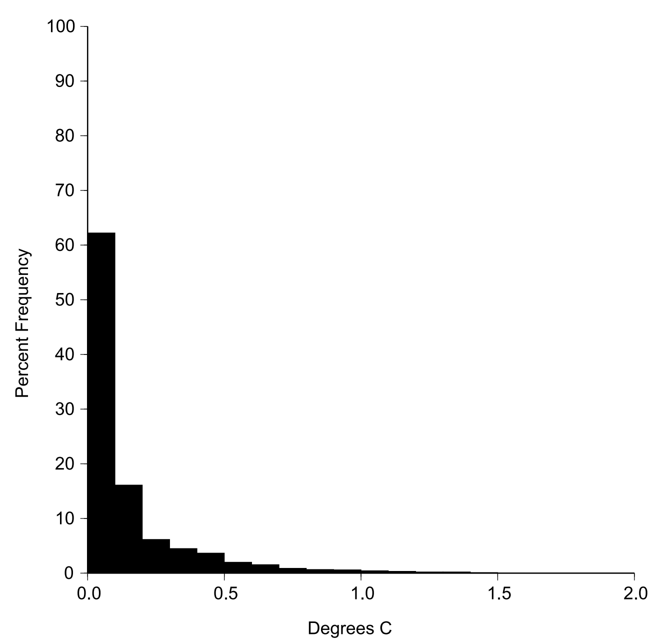

Figure 3. Percent frequency of mean differences (degrees Celsius) between duplicate time series and their associated overlap periods. Maximum difference of 3.3°C is not shown. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

observations, monthly means derived from incomplete daily reports, and other factors that indicated the source data set was unreliable. A second set of tests also implemented on a one-time basis were used to identify problems such as digitization errors, mislocated stations, and repeated reports of the same value [*Peterson et al.*, 1998a]. The third stage of quality control was conducted during data set development and also as part of routine operational processing during monthly update cycles. This involved evaluating individual data points to determine if they were outliers in time or space. Station values were compared to period of record observations and to the values at neighboring stations to assess data quality. This method of routine operational quality control is retained in version 3, but is enhanced with new quality control tests as described in section 3.3.

[17] The version 2 homogeneity adjustment process was based on the creation of a composite reference series for each station, which is subsequently used to identify and remove inhomogeneities in the target station [*Peterson and Easterling*, 1994; *Easterling and Peterson*, 1995]. This method did not rely on station history metadata for identifying or removing inhomogeneities. Adjustments were applied only to stations that had at least 20 years of data, and some stations were not adjusted if they were remote and did not have nearby neighbors sufficiently correlated for creating a reference series. The homogeneity adjusted version 2 data set consists of the 4771 stations which were either deemed to be homogenous or for which homogeneity adjustments could be applied. This data set also includes the 1221 stations from the U.S. Historical Climatology Network (USHCN) version 1 network which were adjusted using the USHCN metadata approach [*Easterling et al.*, 1996].

3. Changes Incorporated in Version 3

[18] Development of GHCN-M Version 3 focused on four areas: (a) consolidating duplicate station records; (b) improving station coverage, especially during the 1990s and 2000s; (c) enhancing quality control, and; (d) applying a new bias correction methodology that does not require use of a composite reference series. These changes establish a new baseline from which further improvements will be made in the future.

3.1. Consolidation of Multiple Temperature Series Attributed to a Single Station

[19] Although retaining multiple series for a given location preserves more of the source data obtained for a station within the integrated data set, the presence of duplicates became a widely misunderstood characteristic of GHCN-M version 2, and made analysis and maintenance of the merged mean temperature data more complicated than necessary. For this reason, the 2706 stations in version 2 having one or more duplicates were processed to combine the duplicates into single station series in version 3. (Version 2 data remain available on the GHCN-M website for users requiring the original duplicates.)

[20] The duplicates were merged based on a process whereby the longer duplicate time series were given higher preference. Thus if a station had three candidate duplicate time series in version 2, and they were, for example, of lengths 56, 44, and 32 years respectively, the final resultant time series in version 3 would consist of all the data (missing and non-missing data) within the 56-year series, any data in the 44-year series that were outside the period covered by the 56-year series, and finally any data in the 32-year series not contained within any of the other two series. There were instances where a shorter station series, which was removed in the de-duplication process, was of higher quality than the data retained. These short high quality station series were part of the World Weather Records data set. As part of the version 3 update process they were retained in the final version 3 data set because they existed as a separate source data set with a higher priority than the version 2 data (see section 4 and Table 3).

[21] Because both missing and non-missing values are retained for the entire segment length covered by each duplicate selected, it is possible that some non-missing values in shorter duplicate series are replaced by missing values in a longer duplicate series in the final merged records. Missing values within a duplicate are not replaced with observations from the non-selected duplicate because of the potential for introducing undetectable inhomogeneities in the station series.

[22] Comparisons between version 2 and version 3 data indicate that this type of replacement reduced the number of non-missing station months in version 3 by less than 0.5% compared to version 2. It is important to note, however, that each of the original data sources merged to create GHCN-M version 3 are also available to data users as part of an international global surface temperature databank [*Thorne et al.*, 2011].

3.2. Changes and Updates to the GHCN-M Database

[23] In 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 for the World (MCDW). WWR consists of monthly mean observations collected from WMO members at the conclusion of each decade, and a new decadal volume was compiled since GHCN-M version 2 was released in 1997. The MCDW data set is essentially a compendium of all CLIMAT reports received by NCDC. Inclusion of observations from WWR and MCDW made it possible to increase (by over 400) the number of existing GHCN-M stations having at least 9 months of data each year during the 1990s (Figure 2).

[24] Since the release of GHCN-M version 2, there have been three principal sources used to update the GHCN-M climate record on an ongoing basis: 1) WMO CLIMAT bulletins disseminated in near-real time each month via the Global Telecommunication System; 2) NCDC's Monthly Climatic Data for the World (MCDW) which is comprised of CLIMAT data from various sources including data disseminated and processed with a delay of three to four months; and 3) the U.S.-based Cooperative Observer Program (COOP) *Summary of the Day* data set. These sources provide data for more than 2500 stations worldwide, and they remain the primary sources for updates to version 3.

[25] CLIMAT bulletins transmitted via the Global Telecommunication System (GTS) provide data each month for approximately 1400 GHCN-M stations in more than 125 countries and territories. Most of these data, which are

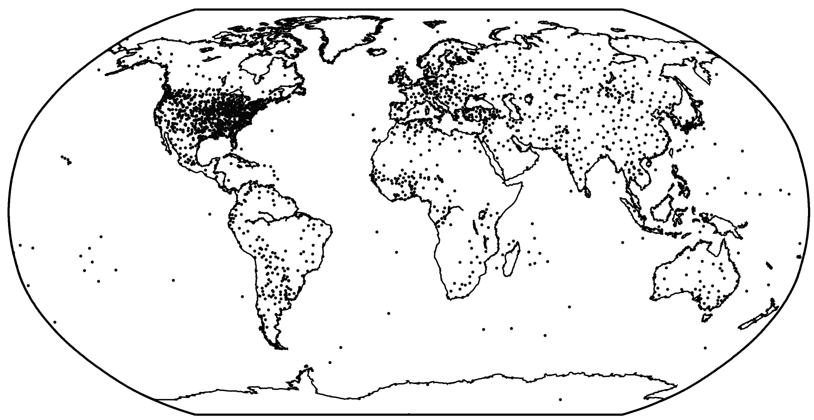

Figure 4. Locations of the approximately 2300 GHCN-M stations for which data are routinely available. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

routed through the Washington DC. Regional Telecommunication Hub, typically arrive within three to eight days of the end of each data month. Other CLIMAT bulletins that do not reach NCDC through the Washington, D. C., GTS pathway are provided by the UK Met Office via ftp (http://hadobs.metoffice.com/crutem3/data/station_updates/) approximately 20 days after the end of the data month. This is an auxiliary source that helps fill in gaps that can result from GTS telecommunications and routing problems. An extra layer of quality control provided by the UK Met Office contributes to the additional 10 to 12 day delay.

[26] Although there are some gaps in South America and Africa where data are not provided in near real-time, the collection of near real-time observations from CLIMAT messages are well distributed worldwide. The number of available stations far exceeds the number required for calculating an accurate global land surface temperature [*Jones*, 1994] (Figure 4), but the greater density provided by GHCN-M is necessary for regional and sub-regional analyses, and it improves the efficiency of homogenization and quality control. An additional 100 to 150 stations are provided to NCDC via e-mail and parcel post. These are processed in delayed mode as part of NCDC's updates to the MCDW data set and are added to version 3, typically three to four months after the data month.

[27] In the transition to GHCN-M version 3, data from the U.S. Historical Climatology Network (USHCN) version 1 data set which ended in 2006 were replaced with the 1218 stations in USHCN version 2 [*Menne et al.*, 2009]. This replacement had little impact on spatial coverage for land areas as a whole. However, because the USHCN version 2 development process used only monthly mean maximum and minimum temperatures in order to produce monthly mean temperature, USHCN version 2 (and thus GHCN-M version 3) contains 136 fewer stations with mean temperature data before 1895 across the contiguous U.S. (Figure 2).

3.3. Changes to Quality Control Algorithms

[28] Version 3 monthly temperatures undergo routine automated quality control procedures (QC) that are objective, reproducible, traceable, and applied consistently throughout the data set. Based on design principles established at NCDC [*Durre et al.*, 2008], the QC process continues the tradition of automated quality control that supports rapid updates to the data set while building upon the QC procedures used in version 2 [*Peterson et al.*, 1998a]. The quality control algorithms are a combination of algorithms applied in version 2 with others adapted from those used to QC the GHCN-Daily data set [*Durre et al.*, 2010] and to produce the USHCN-Monthly version 2 data [*Menne et al.*, 2009].

[29] The quality control checks are listed in Table 1 in the order in which they are performed. They can be grouped into three general categories: basic integrity, outlier, and spatial consistency [*Durre et al.*, 2010]. The quality control process for version 3 mean temperature begins with three basic integrity checks followed by one outlier and one spatial consistency check. Once an observation fails a quality control check, the value is excluded from subsequent checks during that processing cycle. The quality control flags listed in Table 1 are included in the version 3 data set for any datum 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 version 3 data set. It is appended to each observation along with a measurement flag and a source flag, as discussed in section 4.

[30] An observation with no quality control flag indicates that the datum passed all checks applied. But given that some checks have minimum data record requirements, not all monthly values are necessarily subjected to the full suite of tests due to insufficient data or a lack of neighbors.

Table 1. Quality Control Checks for Uncorrected GHCN-M Version 3 Data

Quality Control Check	Quality Control Flag	Flag Rate as Percent of All Non-missing Data	Flag Rate as Percent of All Quality Controlled Flags
Month over Month Duplicate	W	0.029	13.31
Yearly Duplicates	D	0.007	3.28
Isolated Values	L	0.010	4.65
Climatological Outliers	O	0.070	32.31
Spatial Check	S	0.100	46.45
Expert Assessment[a]	Z	0.00	0.00

[a]Flags associated with expert assessment and incorporated into NCDC's DATZILLA system will be incorporated as necessary in the future.

[31] It should be noted that if the quality of any observation is subsequently determined to be different than that classified by the automated quality control process, the version 3 update system allows for implementation of exceptions. In the case of false positives (valid observations erroneously flagged by automated algorithms), an override is applied to ensure the value is not flagged in the version 3 data set. For unflagged observations that are shown to be invalid through other corroborating evidence, a quality control flag is appended to the observation as shown in Table 1. The corroborating evidence includes specific verifiable information such as that provided by a local expert who witnessed the extreme event or has other evidence to support the change in quality. All information related to the event in question and the evidence which supports a change to a quality indicator is documented as part of NCDC's Datzilla system [*Shein*, 2008] and a corresponding source flag appended to the observation as discussed in section 4.

3.3.1. Month-Over-Month Duplication

[32] This check identifies errors resulting from a problem that can occur in the transmission of CLIMAT bulletins over the GTS; the retransmission and incorrect labeling of data that results in the mean temperature for the current data month being repeated from the prior month. This specifically targets data problems that typically occur across all stations in a country or region within a country.

[33] Occasionally a country will retransmit the observations from the previous month and misidentify the observations as being for the current data month. Although meteorological conditions can result in a valid recording of the same monthly mean temperature for a station in consecutive months, particularly in the tropics, the occurrence is highly suspicious when several stations from the same country or the same region within a particular country report identical values in consecutive months.

[34] This error can go undetected by other quality control checks because the observations often do not deviate greatly from climatological normals and because other nearby observations reported with the same source of error provide erroneous corroboration. The "month-over-month duplicate" check identifies and flags these errors. The algorithm operates independently on three latitudinal bands, 90°S-30°S, 30°S-30°N, and 30°N-90°N. Temperatures in the tropics (30°S-30°N) vary less from month to month and, therefore, stations have a greater likelihood of having the same mean temperature in two consecutive months than at locations in other regions. In the tropics, for any month in which three or more stations located within the band and from the same country report a value that is an exact duplicate of the previous month's value, the most recent month's values from those stations are considered erroneous. Because this check was designed to identify retransmission of the prior month's data, only the most recent month's datum is flagged as an error. The check operates the same way in the 90°S-30°S and 30°N-90°N bands with the exception that the minimum number of stations from the same country within the band having same-month duplicates is two, rather than three.

[35] The month-over-month duplicate check is applied to data from January 2000 to present, because data since that time are based largely on observations transmitted in CLIMAT bulletins, in which this problem is most prevalent. Through September 2010, approximately 2000 of the mean temperature data values in GHCN-M version 3 had been identified as bad by this algorithm. Locations with values flagged by this check are shown in Figure 5a.

3.3.2. Duplicate Year

[36] Duplication of observations also can occur on annual timescales due to problems with data collection and processing errors. It is unlikely that meteorological conditions will produce exactly the same set of monthly mean temperatures in any two years. As such, the duplicate year check was designed to identify and flag these occurrences by comparing on a per station basis data for every month in a year to every other year for that station. If any two years are identical, all 12 months for both years are flagged. By retaining the value with the duplicate flag, users can decide whether to eliminate both years from an analysis or query, or keep at least one of the year's data. This condition exists in the record of less than 20 stations as shown in Figure 5b.

3.3.3. Isolated Value

[37] The final basic integrity check identifies "isolated values"; a monthly value or cluster of values that are isolated in time and have no immediate non-missing values within 18 months of either side of the value or the cluster. Experience has shown that a datum, or a small collection of data, is likely invalid when found to be isolated in time from the main collection of a station's data. In order to identify these situations, any station having up to three consecutive observations separated from other data by at least 18 months or more of missing observations, before or after the time period containing these data, are flagged as "isolated." Stations with values flagged by this check are shown in Figure 5c.

3.3.4. Climatological Outlier

[38] The version 3 quality control process seeks to identify outliers using robust and resistant statistical techniques [*Mosteller and Tukey*, 1977], specifically the biweight mean and biweight standard deviation, which is more fully described for climatological purposes in *Lanzante* [1996]. The period of record biweight mean and standard deviation are used to normalize station data through the calculation of a z-score for each month and year of data, in the same way as applied in the version 2 QC process. Any observation equal to or greater than 5 biweight standard deviations above or below its period of record biweight mean for the month in question is flagged as an outlier. Figure 6 shows two outliers detected within the station record for *Centreville, Alabama, U.S.* and displayed within their annual cycle. This test operates under the assumption that the data are normally distributed and requires that a station record have at least 10 years of data for any month. Normality is generally valid for monthly mean temperature data, but skewness in the distribution can result in over-flagging. The absence of any consistent spatial preference (Figure 5d) supports the adequacy of this test.

3.3.5. Spatial Inconsistency

[39] For observations that are less than 5 sigma, but more than 2.5 sigma from the station's biweight mean temperature, a comparison with neighbors is used to assess its validity. Proven to be effective at verifying the validity of observations in the 15 years since it was first applied, the spatial consistency check developed for version 2 [*Peterson et al.*, 1998a] also has been carried over to the quality control process in version 3. This check is implemented while recognizing 1)

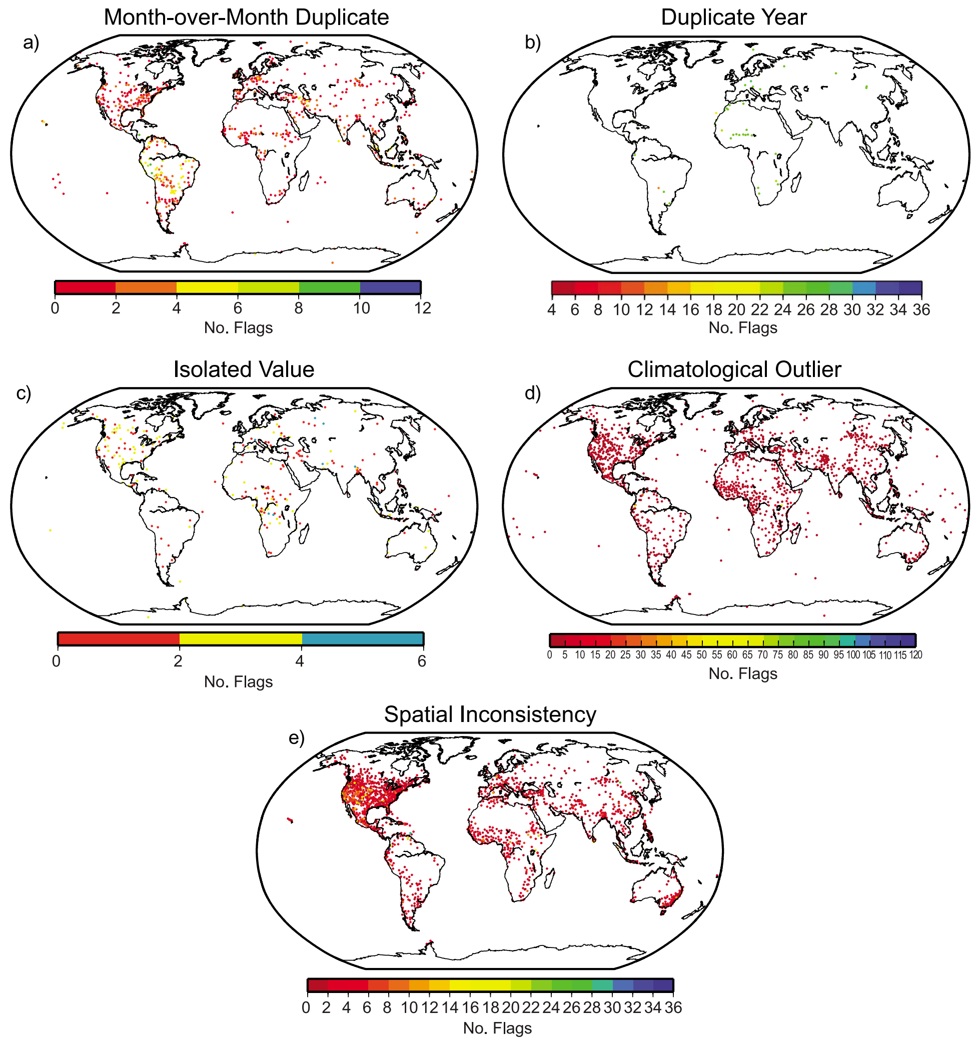

Figure 5. Stations with at least one or more quality control flags during the period of record for (a) month-over-month duplicate check, (b) duplicate year check, (c) isolated value check, (d) climatological outlier check, and (e) spatial inconsistency check. (GHCNM version: ghcnm.v3.0.0-beta1.20110331).

reliance on an implicit assumption that neighboring stations share the characteristics of the target station, and 2) for non-uniformly spaced data fields, regions with sparse data may not provide any representative neighboring data.

[40] The check is based on a *z*-score comparison with the five nearest neighbors to identify occurrences of extreme temperature at the target station also observed at neighboring stations. *Peterson et al.* [1998a] identified the point at which errors could be detected as 2.5 biweight standard deviations from the mean. Selection of the five neighbors for comparison is based only on proximity to the target station (i.e., those closest to the target). Correlation with neighbors is not considered because the GHCN-M periods of record vary greatly and a neighbor may only have a few overlapping years of data making the calculation of correlation impractical.

[41] The validity of suspect observations is based on the magnitude of the normalized value of its neighbors. At least one neighbor having a z-score as shown in Table 2 and of the same sign as the target station provides confirmation of a valid observation of the target station. If the validity of an observation is not verified by comparison with its neighbors, the corresponding flag is appended to the observation (Table 1). The requirement that only one of the five neighbors provide corroboration was determined through evaluation of test results [*Peterson et al.*, 1998a]. However, it is possible that all five neighbors could be separated from the target by great distances or topographic features (e.g., mountain ranges or bodies of water), lessening the likelihood that the extreme value would be corroborated.

[42] This check has the highest flag rate (0.10%) and is most effective in areas of high spatial density where the greater number of neighbors provides more opportunities for corroboration (Figure 5e).

3.3.6. Summary of Quality Control Practices

[43] The quality control processes applied to the version 3 mean temperature data were designed to preserve high frequency variability as well as the long-term mean. Each QC check has a low false positive rate, which reduces the likelihood that valid observations will be erroneously flagged. While this approach to quality control increases the chance that erroneous observations will not be detected by the

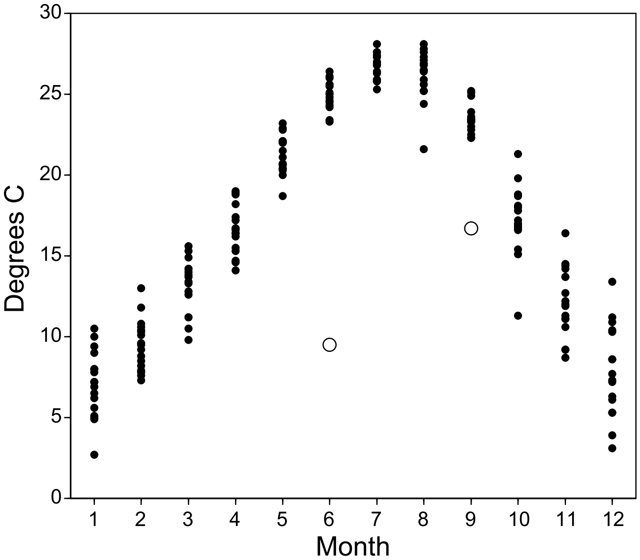

Figure 6. Monthly mean temperature data (solid black circles) for GHCN-M station 42572229000 (Centreville, Alabama, U.S.; period of record 1983–2000), and two outliers (unfilled circles) in June (1994) and September (1996). (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

automated algorithms, the ability to override quality indicators based on expert input and the use of the Datzilla system for recording the history of such changes mitigates this weakness.

[44] The climatological outlier and spatial check together combine for more than 75% of all flagged values; 32% and 46% respectively (Table 1). Although the spatial check can validate outliers between 2.5 and 5.0 sigma, the relatively large number of outlier flags within this range is in part a consequence of a lack of neighbors. Another 13% of the flagged values are identified by the month-over-month duplicate check, and the isolated values and yearly duplicates checks together comprise less than 8% of the flagged values. Values flagged by expert assessment and incorporated into the Datzilla system will be incorporated as necessary in the future.

3.4. Homogeneity Testing and Correction

[45] Surface weather stations are frequently subject to minor relocations throughout their history of operation. Observing stations may also undergo changes in instrumentation as measurement technology evolves. Furthermore, 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. The manifestation of such changes is often an abrupt shift in the mean level of temperature readings that is unrelated to true climate variations and trends. Ultimately, these artifacts (also known as inhomogeneities) confound attempts to quantify climate variability and change because the magnitude of the artifact can be as large as or larger than the true background climate signal. The process of removing the impact of non-climatic changes in climate series is called homogenization, an essential but sometimes overlooked component of climate analysis.

[46] Artificial shifts in a climate series are perhaps most efficiently detected as changes relative to surrounding, highly correlated series from neighboring stations. In essence, homogenization involves identifying and correcting for abrupt shifts in a particular station series when these shifts appear to be unique to that series. The assumption in tests for relative homogeneity is that geographically isolated shifts in temperature series that endure with time are artificial, or, at least, are likely to have originated from causes other than background variations in weather and climate. This assumption can sometimes be verified when a shift in temperature values from a target location relative to other nearby (correlated) values coincides with a known change in observation practice at the target site such as a small station move [*Karl and Williams*, 1987]. Unfortunately, because station history records are generally incomplete if available at all, undocumented shifts may be present throughout the periods of record in a data set such as GHCN-M. While the impacts of these changes are often random, their collective impact can nevertheless systematically bias regional and global temperature trends [*Menne et al.*, 2009].

[47] In version 3 of the GHCN-M temperature data, the apparent impacts of documented and undocumented inhomogeneities are detected, and corrected for, through automated pairwise comparisons of mean monthly temperature series as detailed by *Menne and Williams* [2009]. In contrast to the methodology applied in version 2 [*Peterson and Easterling*, 1994; *Easterling and Peterson*, 1995], the version 3 methodology avoids problems inherent in constructing a reference series of unknown quality to test the relative homogeneity of any particular series [*Menne and Williams*, 2005]. In brief, the creation of a homogenous composite reference series from neighboring stations cannot be ensured, which may cause inhomogeneities in the reference series to be erroneously attributed to the target series [*Menne and Williams*, 2005].

[48] The pairwise algorithm [*Menne and Williams*, 2009] does not assume reference series homogeneity because comparisons are made between numerous combinations of temperature series in a region to identify cases in which there is an abrupt shift in one station series relative to many others. The algorithm starts by forming a large number of pairwise difference series between serial monthly temperature values from a region. Each difference series is then statistically evaluated for abrupt shifts, and the station series responsible for a particular break is identified automatically. Neighbors used in creating the difference series are those which are best correlated with the target. There is no limit to the physical distance between the target and its neighbors. In

Table 2. Threshold Values for Spatial Quality Control Check[a]

Target Station Normalized Temperature (σ)	Threshold for 1 of the 5 Nearest Neighbors (σ)
4.0–5.0	1.9
3.0–4.0	1.8
2.75–3.0	1.7
2.5–2.75	1.6

[a]*Peterson et al.* [1998a].

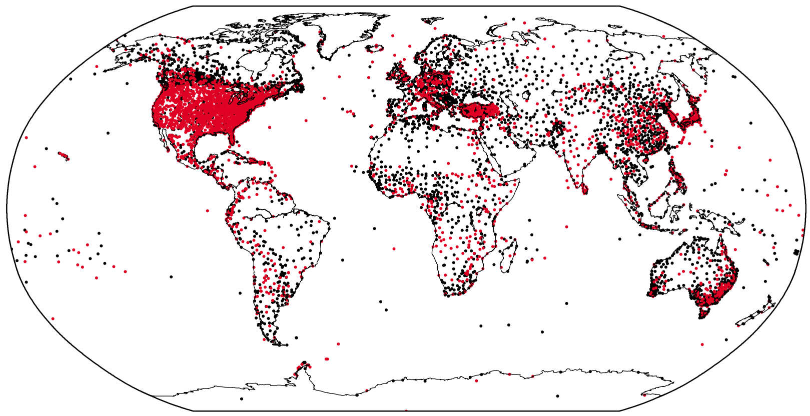

Figure 7. Location of the 7280 monthly mean temperature stations in GHCN-M (black), and the location of every station for which one or more bias corrections were applied (red).

at least once case (St Helena Island), neighbors more than 1000 km away were sufficiently correlated to identify and correct for an inhomogeneity that occurred in 1976. Others have also identified this inhomogeneity [*Hansen et al.*, 2010].

[49] After all of the shifts that are detectable by the algorithm are attributed to the appropriate station within the network, an adjustment is made for each target shift. Adjustments are determined by estimating the magnitude of change in pairwise difference series between the target series and highly correlated neighboring series that have no apparent shifts at the same time as the target. Adjustments are not applied for statistically insignificant changes.

[50] All GHCN-M stations having data from 1801 through present were subjected to the homogeneity assessment process; this included all but one station (64502627001; Lund, Sweden; period of record 1753–1773). One or more bias corrections were applied to 3297 of the 7279 stations (Figure 7). The magnitude of corrections necessary for removing inhomogeneities from station records were applied equally to all months preceding the inhomogeneity, and corrections generally ranged from +/−0.2°C to 2.0°C. Less than 5% of all corrections exceeded +/−2.0°C as shown in Figure 8.

[51] The efficiency of pairwise relative homogeneity testing is, in part, a function of station density. Higher densities

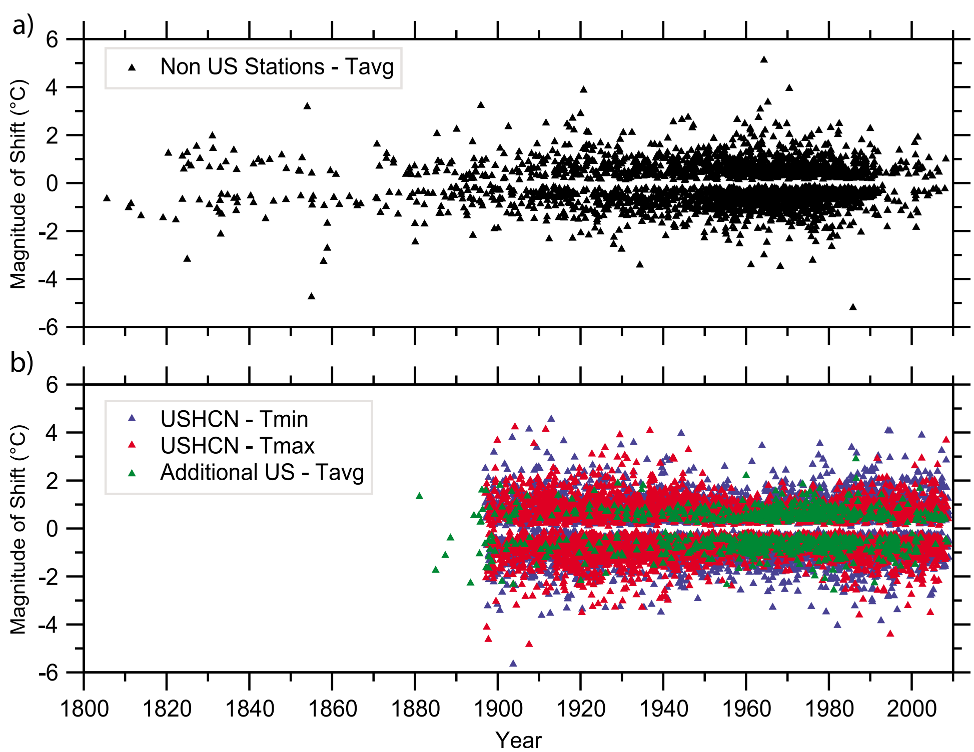

Figure 8. Magnitude and timing of shifts identified by the pairwise homogenization algorithm for (a) non-U.S. stations and (b) stations in the U.S. Bias corrections to USHCN stations are applied to maximum and minimum monthly temperature with subsequent averaging to produce monthly mean temperature.

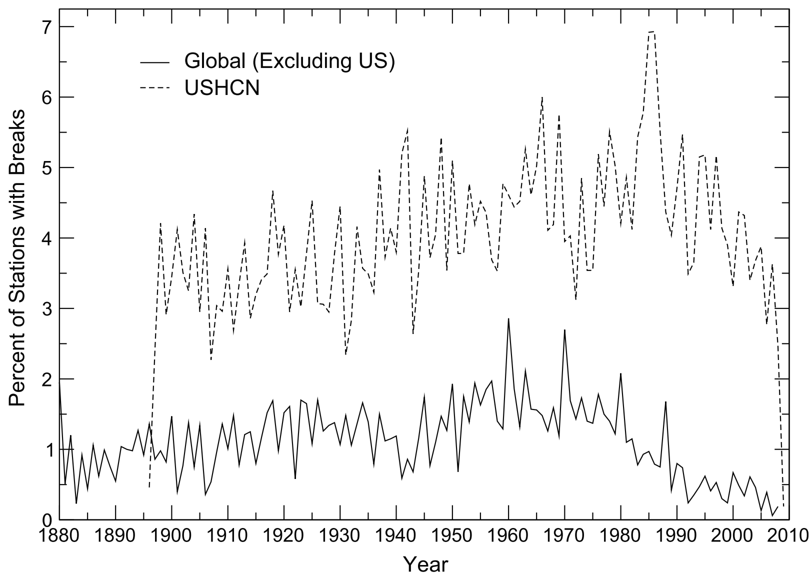

Figure 9. Annual percentage of stations for which bias corrections were applied (1880–2010); USHCN (dashed line) and all other GHCN-M stations (solid line).

generally increase the covariance between stations and improve the signal-to-noise ratio between shifts in systematic bias and the random differences between stations. The effect of station density on the efficiency of data homogenization can be seen in Figure 9, which shows the percentage of stations for which bias corrections were applied on an annual basis. The network of USHCN data is shown separately because identification of inhomogeneities is aided by the use of station history metadata, and because its data density is consistently much higher than elsewhere in the world. The application of bias corrections for USHCN is highest generally during the 1960s through 1980s. This was a period when the number of stations was greatest and also when there were widespread changes in time of observation and instrumentation. The peak during the 1980s reflects the impact of a transition from Cotton Region Shelters to the Maximum-Minimum Temperature System (MMTS) at many USHCN stations. Although approximately 60% of the COOP network was converted to MMTS instrumentation during a five-year period, the number of unaffected stations was sufficient to support relative homogeneity testing and bias correction even during this period when a majority of the COOP network was affected by a change in instrumentation. Outside the U.S. the highest proportion of stations receiving bias corrections coincides with the 1950s through 1970s peak in the number of stations in the GHCN-M data set.

[52] In the case of spatially isolated series, relative homogeneity testing is less likely to reveal the impact of artificial station changes. Conversely, the relatively dense station network behind the U.S. contribution to GHCN-M, as well as the more uniform record length of USHCN stations, allows for more efficient relative changepoint detection and bias correction.

[53] The impact of systematic shifts in temperature measurement bias in the U.S. climate network is described in detail by *Menne et al.* [2009]. In short, changes in temperature measurement in the U.S. have been shown to have a systematic impact on the magnitude of regionally averaged temperature trends that are not related to true climate change. These impacts can be revealed by asymmetries in the sign of the shifts detected via relative homogeneity testing. As shown in Figure 10, such an asymmetry in changes in bias over time appears also to have occurred for the GHCN-M stations outside of the U.S. based on the results of the pairwise homogenization algorithm, which identified about 220 more negative shifts (cold step changes) than positive shifts. (The shifts in the U.S. are of a similar magnitude as those shown in Figure 10 for areas outside the U.S.) Because there are more cold step changes than warm in the historical record, the bias correction process produces global trends that are slightly higher than those based solely on raw observations. The reason for the larger number of cold step changes is unclear but may be due in part to systematic changes in

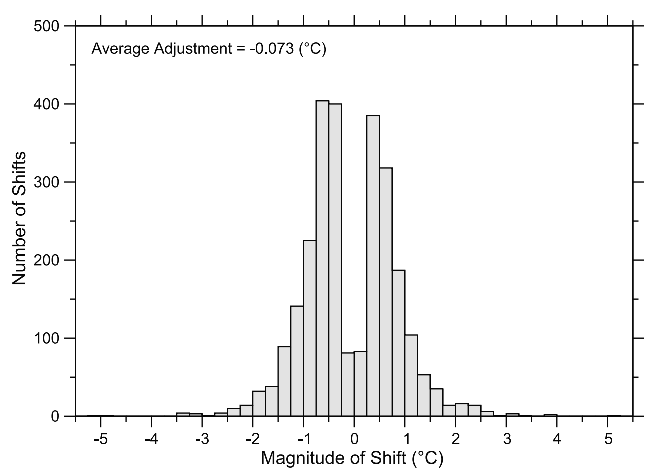

Figure 10. Histogram of the shifts identified in mean monthly temperature series outside of the U.S.

Table 3. Source Data Sets From Which GHCN-M Version 3 is Constructed and Maintained

Priority	Source Data Set	Source Flag
1	Datzilla (Manual/Expert Assessment)	Z
2	USHCN-M Version 2	U
3	World Weather Records	W
4	KNMI Netherlands (DeBilt only)	N
5	Colonial Era Archive	J
6	MCDW (DSI 3500)	M
7	MCDW quality controlled but not yet published	C
8	UK Met Office CLIMAT	K
9	CLIMAT bulletin	P
10	GHCN-M Version 2	G[a]

[a]For any station incorporated from GHCN-M version 2 that had multiple time series ("duplicates") for mean temperature, the 'G' flag is replaced by a number from 0 to 9 that corresponds to the particular duplicate in version 2 from which it originated. This number is the 12th digit in the version 2 station identifier.

station locations from city centers to cooler airport locations (section 5.3).

4. GHCN-M Version 3 Processing

[54] A new paradigm for updates and maintenance of GHCN-M version 3 has been established to address limitations in the previous version. The GHCN-M version 2 update process was structured to quickly and efficiently add and quality control recent observations as they were made available to NCDC via the pathways discussed in section 3.2. Once the version 2 data set was initially constructed from its many sources the historical record remained fixed unless specific errors were identified. This process worked well to incorporate recent observations, but experience has shown that changes in the historical record of source data sets often occur. This can be manifested through the addition of data from new or higher quality sources, and also through the influence that additional data can have on the overall effectiveness of spatial and temporal quality control checks. To account for this possibility, the update process was designed to quickly and efficiently incorporate such changes, document the source of each observation, and verify the quality and homogeneity of the data set as a whole by subjecting the full period of record to the quality control and homogeneity process during each update cycle.

[55] The GHCN-M version 3 processing system is based on an approach first used in the production of GHCN-Daily (M. J. Menne et al., An overview of the Global Historical Climatology Network Daily Database, submitted to *Journal of Atmospheric and Oceanic Technology*, 2011). In this approach, each of the monthly mean temperature sources listed in Table 3 is maintained separately and the entire version 3 data set is reconstructed from these sources on a continual basis. Reconstruction of the database occurs every day at which time newly available data are added. The incorporation of new data is followed by quality control, bias correction, and finally output of the data files and associated summary statistics and graphical products as outlined in Figure 11.

[56] The data integration phase begins by assembling and merging the various source level data sets. Although a single datum may be provided by more than one source, only one value is added to version 3 for any particular month. The datum is selected based on availability and a hierarchical process involving priority levels based on the reliability and quality of the source. Data from sources considered to be of higher quality and reliability are used preferentially over other sources. Table 3 lists the sources, and their order of assemblage (highest priority listed first). For example, if a non-missing datum is present for the same date/location from data source M (MCDW) and data source P (CLIMAT bulletin), the datum from data source M will be placed in the data set. The source from which each datum originated is indicated in the version 3 data set by a source flag as shown in the table. Daily reconstruction of the data set using this method ensures that any changes made in the source data sets get incorporated into GHCN-M while also allowing for the reproduction of the version 3 data set by other institutions or entities.

[57] In addition to the source flag and the quality control flag for each observation, version 3 also includes a measurement flag. This provides information on the number of daily observations missing in the calculation of monthly mean temperature. This information is currently available only for the 1218 stations from the USHCN version 2 source.

[58] As one of the source data sets integrated into version 3, the U.S. Historical Climatology Network version 2 data set is also reassembled using essentially the same phases (i.e., retrieving updates, reassembling the database, applying the quality assurance and pairwise homogenization algorithm). In this case, the source of updates is daily data for U.S. stations that are fully integrated into the GHCN-Daily data set (Menne et al., submitted manuscript, 2011). Quality control and homogenization occurs using the full set of U.S. data available in GHCN-Daily but not currently incorporated into GHCN-M. Also, in the case of the USHCN, station history changes are available and used by the pairwise homogenization algorithm, which improves changepoint detection. Such histories are not currently available for GHCN-M stations outside the U.S. After reprocessing of the USHCN data, these station records are merged with the rest of GHCN-M.

[59] Data merging and integration is followed by quality control checks which are applied to the full version 3 period of record. The GHCN-M version 2 process applied QC checks to only recently received data and never reassessed the quality of any observation previously identified as valid. By applying quality control to all observations in the version 3 data set during each reprocessing, data quality assessments are provided consistently across the period of record. Additionally, the ability to process the entire period of record makes it possible to apply quality control retrospectively as new methods are developed and to do so in a consistent manner throughout the life of the data set.

[60] Although bias corrections are applied so that the historical record is homogenous with current observations, as the record length increases, the potential for new inhomogeneities increases due to new station moves, instrument changes, and other factors unrelated to climate. This necessitates continual assessment and application of bias corrections. As part of the version 3 update process, the pairwise homogeneity algorithm is applied during each update cycle and bias corrections made as necessary to better ensure the homogeneity of the complete GHCN-M record.

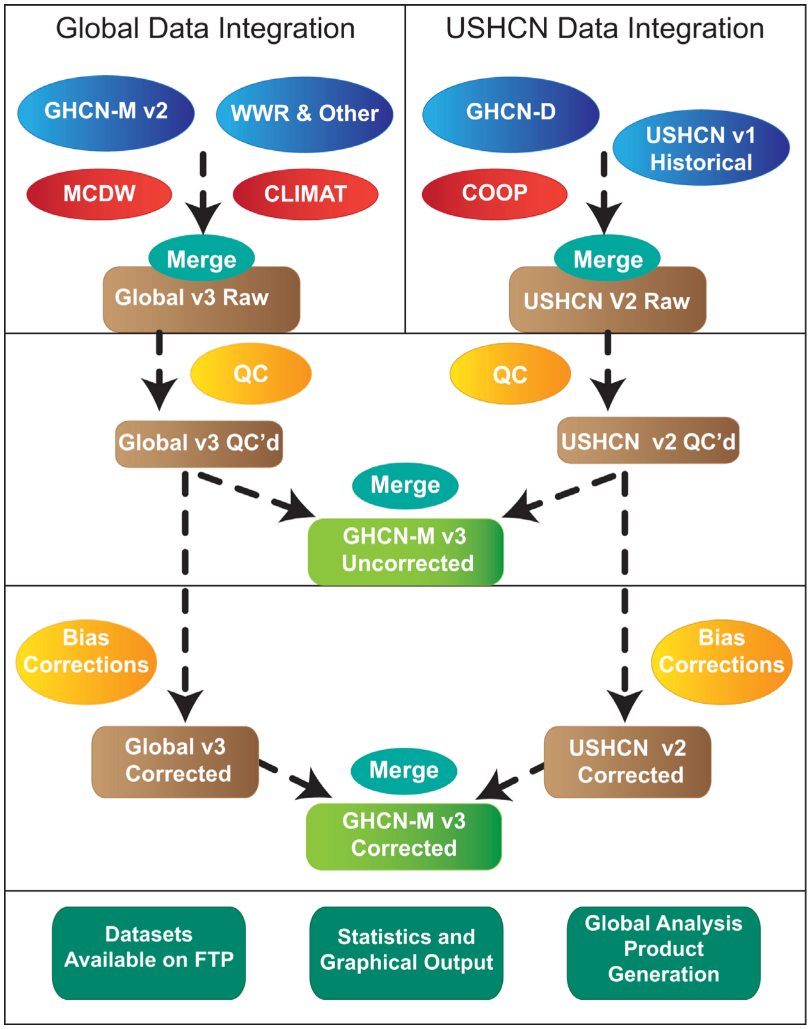

Figure 11. Four-phase processing system for GHCN-M version 3; Data integration, Quality Control, Bias Correction, Output and Product generation.

[61] Following the application of bias corrections, two data files, one each for the uncorrected and corrected GHCN-M mean temperature data, are produced along with a series of output statistics in text and graphical form. These are archived at NCDC as part of each update cycle and made freely available on the GHCN-M website (http://www.ncdc.noaa.gov/ghcnm/).

[62] The corrected and uncorrected data are version controlled using a three-digit number (x.y.z) and a date-time stamp appended to each output data file as part of the file naming convention. The use of a date-time stamp allows for tracking changes to the data set that occur during routine updates and processing while the three-digit versioning tracks changes resulting from minor bug fixes up through major structural enhancements. By providing GHCN-M data with explicit reference to the version number, users are able and encouraged to specifically cite the version used in any analysis or for any purpose.

[63] The file naming structure is ghcnm.v*X.Y.Z.YYYYMMDD* where

[64] 1. *X* is incremented when there is a major change to the data set such as implementation of a new bias correction algorithm or new quality control system. These changes are made through the peer review process and documented within a journal article.

[65] 2. *Y* is incremented when there are one or more significant changes to the data set such as the implementation of a single new quality control algorithm or the addition, correction, or removal of a large number of stations. These changes are included in a technical review document.

[66] 3. *Z* is incremented when any minor change is made. These can include minor bug fixes, correction of minor data errors, minor changes to bias correction or quality control processes, and small additions of new station data. Users are notified of these changes through an online status file that accompanies the data files.

[67] 4. *YYYYMMDD* is the year, month, and day the data set was updated, quality controlled, and bias corrected.

[68] GHCN-M data are an integral part of NCDC climate monitoring activities, and associated products and analyses are included in monthly and annual State of the Climate reports (available from NOAA/NCDC at http://www.ncdc.noaa.gov/climate-monitoring/index.php) to provide perspectives on global and regional temperature anomalies

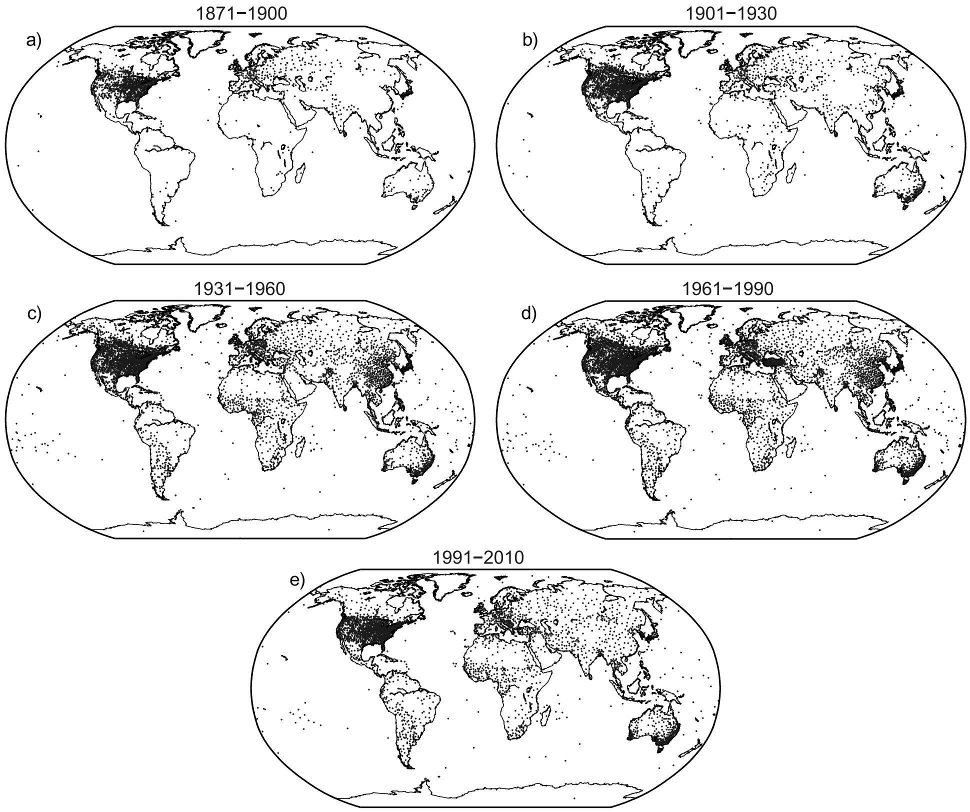

Figure 12. Stations with at least one month of data during the periods (a) 1870–1900, (b) 1901–1930, (c) 1931–1960, (d) 1961–1990, (e), 1991–2010. (GHCN-M version: ghcnm.v3.0.0-beta1.20101111)

and trends. This includes blending with NCDC's Extended Reconstructed Sea Surface Temperature data set [*Smith et al.*, 2008].

5. Global Analysis: Comparison of Version 2 and Version 3

[69] As in version 2, there are more than 1500 version 3 stations with data in 1900 and observations at many of these stations continue into the 21st century. Approximately 4300 stations have at least 50 years of data and an additional 1500 stations have at least 30 years of data. There are 150 stations with as few as 10 years of data.

[70] Variations in the number of GHCN-M stations through time are reflected in changing patterns of spatial coverage from the late 1800s through the 20th century. Figure 12 shows multidecadal station coverage for version 3 from the late 1800s through 2010. Although station density is greatest during 1961–1990, station coverage is widespread from 1900 to present.

[71] Differences in annually averaged global temperature between versions 2 and 3 are generally greatest in the first half of the instrumental record when the density of the network is lowest and any changes to data completeness or homogeneity have the greatest influence on the globally averaged temperature. These differences are evident in global temperature trends; the greatest differences between version 2 and version 3 occur on the century scale, while global trends calculated with either version are virtually identical over recent decades (sections 5.2 and 5.3).

5.1. Analysis Method

[72] The fundamental aspect of any global temperature analysis is the calculation of global temperature anomalies. One of the most common ways temperature anomalies are calculated is through the use of the Climate Anomaly Method (CAM) [*Ropelewski et al.*, 1984; *Jones and Moberg*, 2003]. This method involves the calculation of temperature anomalies (departures from the climatological average) for every month and year of a station's period of record. The climatological average (also referred to as the station's "normal" or "base period" temperature) is typically calculated over a 30-year period, such as 1961–1990 or 1971–2000. To include a station in such an analysis requires that it have some minimum amount of data during the 30-year base period, or that its base period temperature can be reliably estimated. A major drawback to this requirement is the exclusion of some stations from an analysis if the station did not operate during the 30-year base period.

[73] In our comparison of GHCN-M version 2 and version 3, we use the First Difference Method (FDM) [*Peterson et al.*, 1998b]. This method is not dependent on the presence of data during a pre-defined base period and as such is better suited to the many short duplicates in version 2. With this method any station can be used in an analysis whether or not a base period temperature can be calculated or estimated. It relies only on the calculation of a difference in temperature from one year to the next. If a station has two or more consecutive years of data it can be included in an analysis. Calculation of global trends using both the CAM

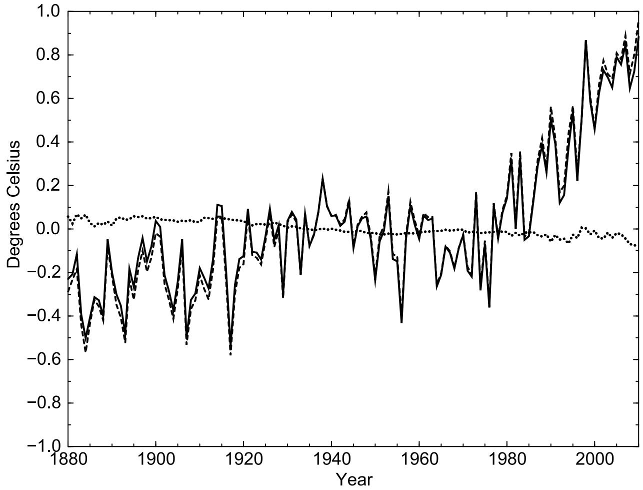

Figure 13. Annual global temperature anomalies from 1880 through 2010 using uncorrected version 2 (dashed line) and uncorrected version 3 (solid line) data. Anomalies are expressed with respect to the 20th Century average temperature. A difference time series (v3 uncorrected minus v2 uncorrected) is shown as a dotted line. Linear trends for various time periods are included in Table 4. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

and the FDM would produce the same result if every station in the data set had 30 years of data during the climatological base period and no missing data throughout its period of record. This is almost never the case.

[74] Using the FDM, the difference in temperature between successive years is calculated for each monthly temperature at each station. These station-based "first differences" are then averaged into 5° by 5° latitude-longitude grid boxes for each year-month from 1880 to 2010. Global first differences are computed by area-weighting each grid box by the cosine of the central latitude and averaging all of the weighted grid box first difference values in the given year-month. The global first difference series is then converted to an anomaly time series by sequential summing (adding the first difference values in series from the first year to last) followed by adjustment to the desired base period; in our analysis 1901–2000.

[75] In assessing how global anomalies and trends are affected by the changes from version 2 to version 3 in sections 5.2 and 5.3, the FDM is applied to both versions, and global temperature averages are calculated from raw uncorrected observations as well as the bias corrected data. These comparisons conclude in section 5.4 with an analysis of global average anomalies and trends computed from the full GHCN-M data set (baseline network) compared against anomalies and trends based on a subset of stations for which observations were available on an ongoing operational basis in 2009 (a real-time network). This is used to assess the impact that fluctuations in the number of stations through time have on global land surface temperature trends. Included in this final section are trends based on the FDM as well as the CAM to illustrate how results can be affected by the use of different analysis methods.

5.2. Global Trends and Anomalies Using Version 2 and 3 Uncorrected Data

[76] Differences in global anomalies and trends between version 2 and version 3 uncorrected data are due to the removal of station duplicates (section 3.1), changes in quality control algorithms (section 3.3), the addition of data during the 1990s and first decade of the 2000s (section 3.2), and the reduction in the amount of U.S. data before 1895 (section 3.2). The combination of all of these changes results in small differences in global trends and variability between version 2 and version 3 uncorrected data.

[77] Global average temperature anomalies using both GHCN-M versions of uncorrected data are shown in Figure 13. There are very small differences in annual average temperature even during the decades when the number of stations and spatial coverage between version 2 and 3 differed the most; prior to 1895 and after 1990. Replacing missing values with newly collected data over the period 1991 to 2010 increased coverage in some areas of the world where data were absent in version 2. This is evident in Figure 14, which shows the number of 5° by 5° grid boxes containing at least one annual observation from 1880 through 2010. During the 1990s there are approximately 600 grid boxes in version 2 compared to approximately 650 in version 3. Conversely, before 1895 there are fewer grid boxes with data in version 3. This is due to reductions in coverage in the United States resulting from changes in the composition of the USHCN version 2 data set (section 3.2).

[78] Although there are small differences on a year-to-year basis, the trends over the most recent 30- and 60-year periods based on version 2 and 3 uncorrected data are nearly identical. The century-scale trends differ more (0.08°C/Century, as shown in Table 4). This reflects the greater impact that changes in station data have in the early decades when data coverage is sparsest, while changes that affect station records since 1950 when the network is densest have less impact on the global trend.

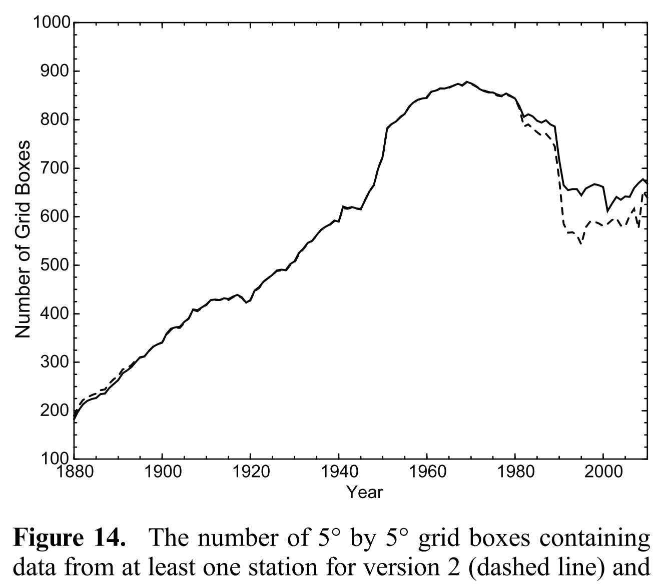

Figure 14. The number of 5° by 5° grid boxes containing data from at least one station for version 2 (dashed line) and version 3 (solid line) based on the First Difference Method. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

Table 4. Trends in Annual Global Land Surface Temperature (Least Squares Regression) for GHCN-M Version 2 and Version 3 Raw Uncorrected and Bias-Corrected Data Over the Periods 1880–2010, 1901–2010, 1951–2010 and 1981–2010[a]

	v2 Uncorr	v3 Uncorr	v2 Corr	v3 Corr
1880–2010	0.69°C/Century	0.61°C/Century	0.76°C/Century	0.79°C/Century
1901–2010	0.78°C/Century	0.70°C/Century	0.88°C/Century	0.91°C/Century
1951–2010	0.17°C/Decade	0.16°C/Decade	0.18°C/Decade	0.18°C/Decade
1981–2010	0.28°C/Decade	0.27°C/Decade	0.28°C/Decade	0.27°C/Decade

[a]The First Difference Method is used.

5.3. Global Trends and Anomalies Using Version 2 and 3 Bias Corrected Data

[79] Trends in average global temperature based on bias corrected version 2 and version 3 data are nearly identical at all time scales (Figure 15 and Table 4). Implementation of the Pairwise algorithm (section 3.4) has improved inhomogeneity detection and bias adjustments at the local and regional levels and changed global means in some years and decades. But short- and long-term version 3 temperature trends, when compared to those computed using bias corrected GHCN-M version 2, provide the same results; the global land surface temperature has increased approximately 0.8°C per century since 1880 and approximately 0.2°C per decade since 1951.

[80] Although global trends from version 2 and 3 bias corrected data are nearly identical, the application of bias corrections results in greater rates of warming since 1880 than found in the analyses based on uncorrected data alone. The difference in the 1880–2010 global trends between the corrected and uncorrected version 3 data approaches 0.2°C per century (Figure 16). But there is very little difference between trends in the corrected and uncorrected data in the periods since 1950 as shown in Table 4.

[81] Given the lack of available station history metadata for stations outside the U.S., determining the cause of higher trends in the bias corrected data is difficult. Possible causes include the need to remove artificial cooling in the climate record that occurred as a result of station moves from downtown locations to more rural airport locations beginning in the 1930s. However, further research and improvements in metadata holdings will be required before definitive conclusions can be made.

5.4. Assessing the Impact of the Decline in Stations Since the 1970s

[82] The global analyses described above are based on stations whose coverage varies both spatially and temporally as discussed in previous sections. Not only are there fewer stations in the early part of the observational record, issues associated with international exchange and station closures have resulted in a decline in the number of GHCN-M stations in recent decades. The large drop-off in recent years is also reflective of the data archeology and collection projects undertaken when version 2 was developed. These projects produced lengthy historical records for many stations and an overall large inventory of stations with coverage that peaked in the 1960s and 1970s [see, e.g., *Peterson and Griffiths*, 1997]. Since that time the number of GHCN-M stations with available data in 2010 has fallen to less than 40% of the total inventory.

[83] To assess the impact of this decline in station coverage on global temperature averages, we conducted an analysis based on the creation of a "baseline" series and a "real-time"

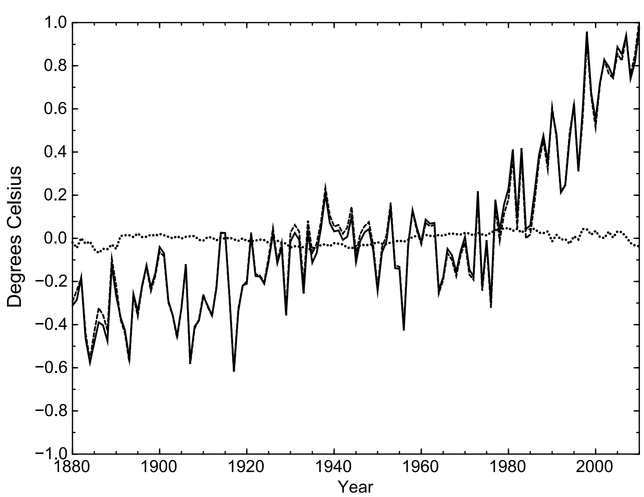

Figure 15. Annual global temperature anomalies from 1880 through 2010 using version 2 bias corrected data (dashed line) and version 3 bias corrected data (solid line). A difference time series (v3 bias corrected minus v2 bias corrected) is shown as a dotted line. Linear trends for various time periods are included in Table 4. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

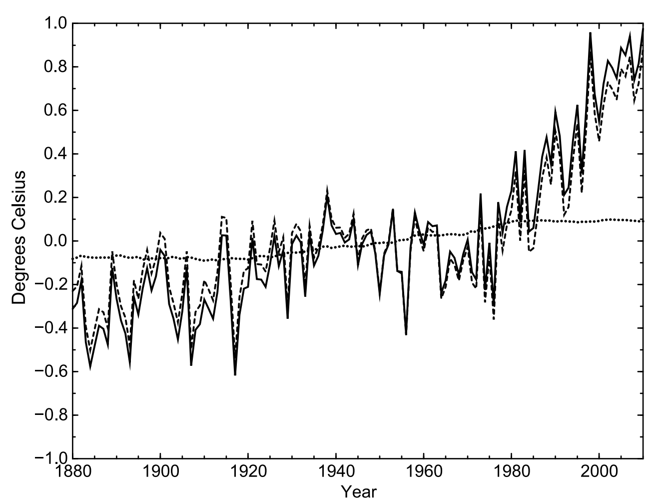

Figure 16. Annual global temperature anomalies from 1880 through 2010 using uncorrected version 3 (dashed line) and bias corrected version 3 data (solid line). A difference time series (v3 bias corrected minus v3 uncorrected) is shown as a dotted line. Linear trends for various time periods are included in Table 4. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

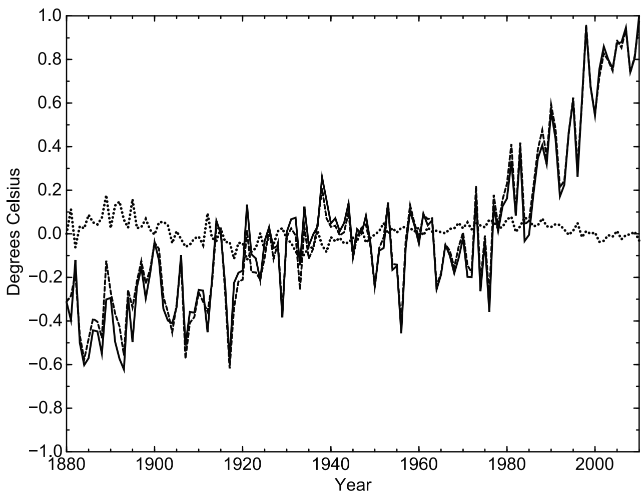

Figure 17. Time series of annual global land surface temperature using a real-time network (solid line) and baseline network (dashed line) analyzed based on GHCN-M version 3 bias corrected data set. The difference series (baseline minus real-time network) is included (dotted line). Linear trends for various time periods are included in Table 5. (GHCN-M version: ghcnm.v3.0.0-beta1.20110331)

series for the global land surface temperature for the period 1880–2010. The real-time network consists of 2749 stations which had data in 2009, while the baseline network consists of the 7279 stations in the full bias corrected GHCN-M inventory. The two networks were analyzed using the FDM and the CAM. The baseline network was reduced to a network of approximately 4800 stations in the analysis using the CAM because of the requirement that stations have at least 20 years of data during the 1961–1990 base period.

[84] Figure 17 shows the annual time series for the real-time and baseline networks from 1880 to 2010 (FDM). Regardless of the method used, the two series are virtually identical, with a mean absolute difference of 0.040°C based on the FDM and 0.025°C with the CAM. The two series have the same or very similar trends over the century-scale and multidecadal periods as shown in Table 5. With the FDM, the 1880–2010 trends are 0.79°C/Century for both the baseline and real-time networks. The 1951–2010 trends are 0.18°C/Decade for the baseline and 0.19°C/Decade for the real-time network over the 1951–2010 period. Using the CAM the 1880–2010 trends are 0.83°C/Century and 0.88°C/Century for the baseline and real-time network, respectively, and 0.18°C/Decade for both networks from 1951 to 2010. The comparable results from both networks indicate that global land surface anomalies and trends are not adversely affected by the decline in the number of stations over the past two to three decades. In fact trend differences between the FDM and CAM can be greater than the differences based on network configuration alone.

5.5. Uncertainties in the Global Land Surface Record

[85] Version 3 enhancements to temporal and spatial data coverage, quality control, and bias correction have improved the overall quality of the GHCN-M monthly mean temperature data set. However, uncertainties in the observational record remain. Sources of uncertainty include random or systematic errors that are undetected by quality control processes, and there are errors associated with changes in spatial sampling through time. Uncertainty is also associated with residual biases in homogeneity corrected data that arise either from discontinuities that remain uncorrected, are poorly adjusted, or falsely corrected.

[86] There have been efforts to quantify the uncertainty in land and ocean surface temperature records associated with sampling and bias error [*Smith and Reynolds*, 2005]. This method is used in NCDC State of the Climate reports (see, e.g., http://www.ncdc.noaa.gov/sotc/global/2011/5) to provide a measure of the uncertainty in the global land surface temperature, which is on the order of 0.1°C annually and 0.1°C to 0.3°C on monthly timescales.

[87] More recently there have been efforts to better quantify the residual bias (red noise error) in homogeneity adjusted data (C. N. Williams et al., Benchmarking the performance of pairwise homogenization of surface temperatures in the United States, submitted to *Journal of Geophysical Research*, 2011). This study finds that the pairwise homogeneity adjustment algorithm is grossly adequate and that uncertainties in contiguous U.S. temperature do not include zero or negative trends over the last 30, 50, or 100 years. Rather it points to a greater likelihood that the homogeneity corrected data underestimate the true trend in contiguous U.S. monthly mean temperature. This method provides a basis for further improving the quantification of uncertainties associated with the bias correction process in the future.

6. Concluding Remarks

[88] For more than 15 years the GHCN-M mean temperature data set has been a cornerstone of efforts to understand how the Earth's temperature has varied and changed since the late 1800s. It has served as NOAA's official source of land surface temperature data for climate monitoring, and ongoing efforts to update the data set each month have provided continuing perspectives on how temperatures are being affected by natural and man-made influences. The release of version 3 is part of broader efforts to provide the

Table 5. Trends in Annual Global Land Surface Temperature (Least Squares Regression) for GHCN-M Version 3 Using a "Real-Time" Network of Stations and the "Baseline" Network of Stations Based on Bias Corrected Data[a]

	v3 Baseline (FDM)	v3 Real-Time (FDM)	v3 Baseline (CAM)	v3 Real-Time (CAM)
1880–2010	0.79°C/Century	0.79°C/Century	0.83°C/Century	0.88°C/Century
1901–2010	0.91°C/Century	0.86°C/Century	0.91°C/Century	0.94°C/Century
1951–2010	0.18°C/Decade	0.19°C/Decade	0.18°C/Decade	0.18°C/Decade
1981–2010	0.27°C/Decade	0.29°C/Decade	0.27°C/Decade	0.28°C/Decade

[a]One analysis using the First Difference Method (FDM) and the second using the Climate Anomaly Method (CAM).

highest quality climate data that is easily accessible to scientists as well as the general public.

[89] Although conclusions regarding land surface temperature variability and change are little affected by this release, the improvements that are part of version 3 have enhanced the overall quality of the data set. The removal of station duplicates has greatly simplified the use of the data set by removing a feature that created confusion and kept some users from gaining full use of the data set. The addition of data from updated sources including World Weather Records and Monthly Climatic Data for the World has enhanced the spatial and temporal completeness of the data set. Introduction of a new bias correction methodology has improved the detection and removal of inhomogeneities, and along with new quality control procedures, better ensures the overall quality of the climate record. Last the implementation of new principals of data set construction and configuration management improve the ability to track data from its point of origin and through each step of processing up to and including archive and distribution.

[90] While these greatly enhance the quality of the GHCN-M monthly mean temperature data, more improvements are already in development and plans include introducing revisions on an annual to biannual basis. Primary among these is an effort to increase the number of stations in the GHCN-M data set. A significant step toward this involves the addition of stations that are part of the GHCN-Daily data set (Menne et al., submitted manuscript, 2011). This data set contains more than 25,000 stations with daily observations of maximum and minimum temperature. Although more than half of these stations are located in the United States or Canada, this is a source of data for thousands of stations in other countries for which monthly mean temperatures can now be calculated, primarily from the 1950s to present.

[91] A parallel effort is focused on identifying and incorporating new sources of data from other meteorological services around the world. This is expected to benefit from a new initiative co-sponsored by NCDC to create a global surface temperature databank [*Thorne et al.*, 2011]. The databank is part of an international grand challenge to build data sets that will better meet user needs for climate information in the 21st century. This will include new methods for validation and benchmarking, better configuration management practices and data provenance, along with better tools for data access and visualization.

[92] These efforts are part of a larger goal to provide comprehensive surface temperature records for understanding trends and variability in the Earth's climate. Although there will be future enhancements to GHCN-M, the version 3 quality control and homogeneity processes combined with improvements in temporal and spatial coverage provide a significant step forward in stewardship of the land surface temperature record. Users are encouraged to make inquiries and provide feedback on their experience with this data set to aid continued efforts to preserve the global temperature record.

[93] **Acknowledgments.** The authors wish to thank David Easterling, Thomas Peterson, Peter Thorne, Jessica Blunden, and Ahira Sanchez-Lugo, whose helpful comments greatly improved this manuscript.

References

Clayton, H. H. (1927), *World Weather Records*, *Smithsonian Misc. Collect.*, vol. 79, 1196 pp., Smithsonian Press, Washington, D. C.

Durre, I., M. J. Menne, and R. S. Vose (2008), Strategies for evaluating quality assurance procedures, *J. Appl. Meteorol. Climatol.*, *47*, 1785–1791, doi:10.1175/2007JAMC1706.1.

Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose (2010), Comprehensive automated quality assurance of daily surface observations, *J. Appl. Meteorol. Climatol.*, *49*, 1615–1633, doi:10.1175/2010JAMC2375.1.

Easterling, D. R., and T. C. Peterson (1995), A new method of detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, *15*, 369–377, doi:10.1002/joc.3370150403.

Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, and D. P. Bowman (1996), United States Historical Climatology Network (U.S. HCN) monthly temperature and precipitation data, *ORNL/CDIAC-87*, 280 pp., Carbon Dioxide Inf. Anal. Cent., Oak Ridge, Tenn.

Hansen, J., R. Ruedy, J. Glascoe, and M. Sato (1999), GISS analysis of surface temperature change, *J. Geophys. Res.*, *104*, 30,997–31,022, doi:10.1029/1999JD900835.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.

Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J. T. Houghton et al., Cambridge Univ. Press, Cambridge, U. K.

Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., Cambridge, Univ. Press, Cambridge, U. K.

Jones, P. D. (1994), Hemispheric surface air temperature variations: A reanalysis and an update to 1993, *J. Clim.*, *7*, 1794–1802, doi:10.1175/1520-0442(1994)007<1794:HSATVA>2.0.CO;2.

Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Clim.*, *16*, 206–223, doi:10.1175/1520-0442(2003)016<0206:HALSSA>2.0.CO;2.

Karl, T. R., and C. N. Williams Jr. (1987), An approach to adjusting climatological time series for discontinuous inhomogeneities, *J. Clim. Appl. Meteorol.*, *26*, 1744–1763, doi:10.1175/1520-0450(1987)026<1744:AATACT>2.0.CO;2.

Karl, T. R., J. M. Melillo, and T. C. Peterson (Eds.) (2009), *Global Climate Change Impacts in the United States*, Cambridge Univ. Press, Cambridge, U. K.

Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, *16*, 1197–1226, doi:10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L.

Menne, M. J., and C. N. Williams Jr. (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series, *J. Clim.*, *18*, 4271–4286, doi:10.1175/JCLI3524.1.

Menne, M. J., and C. N. Williams Jr. (2009), Homogenization of temperature series via pairwise comparisons, *J. Clim.*, *22*, 1700–1717, doi:10.1175/2008JCLI2263.1.

Menne, M. J., C. N. Williams, and R. S. Vose (2009), The United States Historical Climatology Network Monthly Temperature Data - Version 2, *Bull. Am. Meteorol. Soc.*, *90*, 993–1007, doi:10.1175/2008BAMS2613.1.

Mosteller, F., and J. W. Tukey (1977), *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, Mass.

Peterson, T. C., and D. R. Easterling (1994), Creation of homogeneous composite climatological reference series, *Int. J. Climatol.*, *14*, 671–679, doi:10.1002/joc.3370140606.

Peterson, T. C., and J. F. Griffiths (1997), Historical African data, *Bull. Am. Meteorol. Soc.*, *78*, 2869–2872.

Peterson, T. C., and R. S. Vose (1997), An overview of the Global Historical Climatology Network temperature database, *Bull. Am. Meteorol. Soc.*, *78*, 2837–2849, doi:10.1175/1520-0477(1997)078<2837:AOOTGH>2.0.CO;2.

Peterson, T. C., R. Vose, R. Schmoyer, and V. Razuvaev (1998a), Global Historical Climatology Network (GHCN) quality control of monthly temperature data, *Int. J. Climatol.*, *18*, 1169–1179, doi:10.1002/(SICI)1097-0088(199809)18:11<1169::AID-JOC309>3.0.CO;2-U.

Peterson, T. C., T. R. Karl, P. F. Jamason, R. Knight, and D. R. Easterling (1998b), First difference method: Maximizing station density for the calculation of long-term global temperature change, *J. Geophys. Res.*, *103*(D20), 25,967–25,974, doi:10.1029/98JD01168.

Quayle, R. G., T. C. Peterson, A. N. Basist, and C. S. Godfrey (1999), An operational near-real-time global temperature index, *Geophys. Res. Lett.*, *26*(3), 333–335, doi:10.1029/1998GL900297.
Ropelewski, C. F., J. E. Janowiak, and M. S. Halpert (1984), The Climate Anomaly Monitoring System (CAMS), report, 39 pp., Clim. Anal. Cent., NWS, NOAA, Washington, D. C.
Shein, K. (2008), Interactive quality assurance practices, paper presented at 24th Conference on IIPS, Am. Meteorol. Soc., New Orleans, La. [Available at http://ams.confex.com/ams/88Annual/techprogram/paper_131217.htm.]
Smith, T. M., and R. W. Reynolds (2005), A global merged land-air-sea surface temperature reconstruction based on historical observations (1880–1997), *J. Clim.*, *18*, 2021–2036, doi:10.1175/JCLI3362.1.
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. Clim.*, *21*, 2283–2296, doi:10.1175/2007JCLI2100.1.
Thorne, P. W., et al. (2011), Guiding the creation of a comprehensive surface temperature resource for 21st century climate science, Bull. Am. Meteorol. Soc., in press.
Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. 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*, 325 pp., Carbon Dioxide Inf. Anal. Cent., Oak Ridge, Tenn.
World Meteorological Organization (WMO) (2009), Handbook on CLIMAT and CLIMAT TEMP reporting, *WMO/TD-No.1188*, Geneva, Switzerland.

B. E. Gleason, J. H. Lawrimore, M. J. Menne, R. S. Vose, C. N. Williams, and D. B. Wuertz, National Climatic Data Center, NOAA, 151 Patton Ave., Asheville, NC 28801, USA. (jay.lawrimore@noaa.gov)
J. Rennie, Cooperative Institute for Climate and Satellites, National Climatic Data Center, NOAA, 151 Patton Ave., Asheville, NC 28801, USA.