INTERNATIONAL JOURNAL OF CLIMATOLOGY
*Int. J. Climatol.* **25**: 693–712 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/joc.1181

# AN IMPROVED METHOD OF CONSTRUCTING A DATABASE OF MONTHLY CLIMATE OBSERVATIONS AND ASSOCIATED HIGH-RESOLUTION GRIDS

TIMOTHY D. MITCHELL[a] and PHILIP D. JONES[b,*]
[a] *Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK*

*Received 3 March 2004*
*Revised 19 January 2005*
*Accepted 24 January 2005*

ABSTRACT

A database of monthly climate observations from meteorological stations is constructed. The database includes six climate elements and extends over the global land surface. The database is checked for inhomogeneities in the station records using an automated method that refines previous methods by using incomplete and partially overlapping records and by detecting inhomogeneities with opposite signs in different seasons. The method includes the development of reference series using neighbouring stations. Information from different sources about a single station may be combined, even without an overlapping period, using a reference series. Thus, a longer station record may be obtained and fragmentation of records reduced. The reference series also enables 1961–90 normals to be calculated for a larger proportion of stations.

The station anomalies are interpolated onto a 0.5° grid covering the global land surface (excluding Antarctica) and combined with a published normal from 1961–90. Thus, climate grids are constructed for nine climate variables (temperature, diurnal temperature range, daily minimum and maximum temperatures, precipitation, wet-day frequency, frost-day frequency, vapour pressure, and cloud cover) for the period 1901–2002. This dataset is known as CRU TS 2.1 and is publicly available (http://www.cru.uea.ac.uk/). 

KEY WORDS: climate; observations; grids; homogeneity; temperature; precipitation; vapour; cloud

## 1. INTRODUCTION

Climate variability affects many natural and human systems. A major constraint on research is the need to obtain suitable information that is ordinarily held within a variety of different disciplines. There are never sufficient resources for climatologists to customize climate information to provide a product to meet every need. However, a large proportion of these needs may be met through providing a standard set of 'climate grids', defined here as monthly variations over a century-long time scale on a regular high-resolution (0.5°) latitude–longitude grid. Such grids may be inappropriate for small study regions, but for larger areas they may be more useful than a set of individual stations: through a mathematical construct the coverage of a few stations may be expanded to cover a wide area. A prior set of 0.5° grids for 1901–95 (CRU TS 1.0: New *et al*., 2000) has been used to examine the transmission of malaria (Kuhn *et al*., 2003), Canadian carbon sinks (Chen *et al*., 2003), and the demography of the holly-leaf miner (Brewer and Gaston, 2003); this list is not exhaustive. These grids were subsequently updated and extended to 2000 (CRU TS 2.0: Mitchell *et al*., 2004). Other workers have provided shorter records for individual variables; examples include precipitation since 1979 (Xie and Arkin, 1997) or 1986 (Huffman *et al*., 1997).

The construction and routine updating of climate grids depend on information from the global network of meteorological observing stations. Stations are preferred to satellites for these tasks for two reasons: satellite

* Correspondence to: Philip D. Jones, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK; e-mail: p.jones@uea.ac.uk

information only becomes available after 1970, and satellites measure conditions through the depth of the atmosphere rather than at the surface (e.g. Susskind *et al*., 1997). The latter factor also applies to blended products, in which satellite information is used to expand the coverage from stations, a number of which are compared by Casey and Cornillon (1999). However, it is not trivial to build a suitable station database; notable sustained attempts include:

- the Global Historical Climatology Network (GHCN; Vose *et al*., 1992; Peterson and Vose, 1997);
- the Jones temperature database (Jones, 1994; Jones and Moberg, 2003);
- the Hulme precipitation database (Eischeid *et al*., 1991; Hulme *et al*., 1998).

New *et al*. (2000) incorporated this prior work into the database underlying CRU TS 1.0, and wherever possible added information from other sources to extend both the number of climate variables included and the spatio-temporal coverage. This database may also now be augmented with near-real-time information, such as that from the Global Climate Observing System (GCOS) surface network (GSN; Peterson *et al*., 1997). As the number of sources has multiplied, and as additional information is routinely added, it seems necessary to take additional steps to maintain the quality of the database.

1. New station records must be checked to ensure that they present a homogeneous record in which variations are caused only by variations in climate.
2. Information from additional sources must be checked against the existing database, to guard against unnecessary duplication.
3. Where new information is available for an existing station, it must be ensured that the different sources provide consistent records.
4. The number of stations useful for constructing grids must be maximized.

This article describes how the existing database has been expanded, improved, and used to construct a set of climate grids (CRU TS 2.1). A method is developed that addresses the criteria given above (Section 2), the new database and grids are described (Section 3), and the usefulness of the new method is evaluated (Section 4).

## 2. DATA AND METHOD

The sources and assimilation of station records are described first (Section 2.1). The approach to homogenization (Section 2.2) takes the form of an iterative procedure (Section 2.3) in which reference series (Section 2.4) are used to correct any inhomogeneities in a station record (Section 2.5) and the corrected data are merged with the existing database (Section 2.6). The data are converted into anomalies (Section 2.7) and used to construct climate grids (Section 2.8).

### *2.1. Data sources*

Station records were obtained from seven sources (Table I). Jones and Moberg (2003) and Hulme (personal communication) were the primary sources for temperature and precipitation respectively. Both have much in common with Peterson *et al*. (1998c), who were also the primary source for diurnal temperature range (DTR). These three sources have all been extensively checked by their authors. New *et al*. (2000) included these sources but augmented them for some variables. Hahn and Warren (1999) provided a high-quality cloud record (1971–96), accompanied by unchecked information for other variables. There were alternative versions of the CLIMAT messages on the GSN (Peterson *et al*., 1997); the DTR data were derived by Mitchell *et al*. (2004). Sunshine duration data were obtained to augment sparse cloud cover measurements in recent years.

Taking each variable in turn, each source was absorbed into the database in the order indicated in Table I; for cloud, Hahn took priority. Thus, it was ensured that if there were two sources for the same station, precedence was given to the source likely to be more reliable. The station records are held electronically

 

Table I. The sources of station records from which the database was constructed. The climate variables to which the sources contribute are temperature (tmp), DTR (dtr), precipitation (pre), vapour pressure (vap), cloud cover (cld), sunshine duration (spc), and wet days (wet). The dtr includes information from individual records of daily temperature minima (tmn) and maxima (tmx). These labels are used in subsequent tables and figures

| Label | Reference | Information | Period |
|---|---|---|---|
| Jones | Jones and Moberg (2003) | tmp | 1701–2002 |
| Hulme | Mike Hulme, personal communication | pre | 1697–2001 |
| GHCN v2 | Peterson *et al*. (1998c) | tmp, dtr, pre | 1702–2001 |
| Mark New | New *et al*. (2000) | tmp, dtr, vap, cld, spc | 1701–1999 |
| Hahn | Hahn and Warren (1999) | tmp, vap, cld | 1971–96 |
| MCDW | William Angel, personal communication | tmp, pre, vap, spc, wet | 1990–2002 |
| CLIMAT | UK Met Office, personal communication | tmp, dtr, pre, vap, spc, wet | 1994–2002 |

in space-delimited fixed-format ASCII files, which limits the metadata that can be retained, and fixes the units and precision of the data. The latitude and longitude attached to a station record were critical when homogenizing it, so each stated location was compared with a central location and radius for the stated country of origin, to ensure that the location was plausible.

### *2.2. Approach to homogenization*

The potential sources of inhomogeneities in station climate and methods of correction were reviewed by Peterson *et al*. (1998a). The GHCN method of homogenization is well documented, is designed for the automatic treatment of large datasets with global coverage, and has already been applied to a well-established dataset (Peterson and Easterling, 1994; Easterling and Peterson, 1995). The method uses neighbouring stations to construct a reference series against which a candidate series may be compared. Neighbouring stations are selected by a correlation method. If the correlation is performed on absolute values, then a candidate station with a discontinuity may be better correlated with an inhomogeneous neighbour than with one without the discontinuity. Therefore, series of first differences are correlated, to limit the effect of any discontinuity to a single value.

The GHCN method identifies potential discontinuities by correlating subsections of the candidate and reference series; if correlation is significantly improved by using subsections rather than the entire series, then a potential discontinuity is identified. The GHCN method is targeted at abrupt discontinuities, but gradual inhomogeneities will also be detected unless they are widespread. However, it is not critical (or perhaps desirable) to eliminate widespread gradual changes in the station environment, such as large-scale urbanization. The database and the grids subsequently constructed from it are designed to depict the month-to-month variations in climate experienced at the Earth's surface, rather than to detect changes in climate resulting from greenhouse gas emissions.

The GHCN method requires modification for two reasons.

1. The GHCN method is designed for datasets with complete station records for a given period of time. As will be discussed in Section 2.7, the method must be adapted for datasets with incomplete station records and neighbouring stations that only partly overlap in time. This adaptation requires a corresponding change in the use of first differences to build reference series (Section 2.4).
2. Monthly series must be used to detect inhomogeneities, rather than annual series, since some inhomogeneities may have opposite effects in different seasons and so be undetectable in the annual mean. (The GHCN method uses annual series for detection, but Peterson *et al*. (1998a: section 4.2.2) report that inhomogeneities are corrected using a seasonal filter.)

A common problem with homogenization methods is the prior need for a set of stations, known to be homogeneous, against which candidate stations may be safely compared. How can such a set be obtained without testing their homogeneity? This chicken-and-egg problem is addressed here through an iterative procedure (Section 2.3) with three components, one of which itself includes another iterative procedure (Section 2.4.4).

### 2.3. Iterative checking

The first pass through the dataset was an attempt merely to identify (not correct) all potential inhomogeneities. All stations were allowed to contribute to the construction of reference series (Section 2.4). The priority in constructing a reference series was to match the length of the candidate as far as possible, even if this was at the expense of some loss of correlation in the reference series. (This trade-off will be explained in Section 2.4.3.) The reference series were used to identify suspected discontinuities (Section 2.5), but no corrections were made and the stations did not yet enter the final database.

On each subsequent pass through the dataset only those stations where one of the following conditions was met was 'trusted' to contribute to the reference series for any other station:

- it had already been corrected (where necessary) and added to the final database;
- no discontinuities were suspected on the initial iteration;
- it could be split into independent sections using any suspected discontinuities as the boundaries.

Using the trusted stations, a reference series was constructed for as many candidate stations as possible (Section 2.4). Each reference series was used to identify any discontinuities in the candidate and correct them (Section 2.5); then the candidate gained trusted status and was merged into the final database (Section 2.6). The additional trusted stations then allowed reference series to be constructed for further stations.

When no more reference series could be obtained, the omissions criterion was relaxed. The omissions criterion $\lambda$ was the number of years in the candidate that might be without corresponding values in the reference series. The omissions criterion was initialized to zero to ensure that the full record was checked for as many stations as possible, but subsequently it was relaxed, 5 years at a time, so that more stations might have most of their record checked.

The iterative procedure ended for each variable when the level set in the omissions criterion exceeded the length of the longest unchecked station. Then, all the unchecked stations were added to the final database; this was justified for two reasons:

1. The near-real-time sources (notably the CLIMAT messages and the MCDW reports) were not archived prior to 1990. The method of checking for inhomogeneities requires longer records to be effective, so the stations from these sources were added to the database without any checks.
2. Most of the unchecked data were from areas and periods when density is low. Therefore, omitting the unchecked data would have had a disproportionately large effect on the number of grid boxes for which a genuine record of climate variations may be calculated. An unhomogenized station is likely to provide a better record of climate variations than will an assumption of zero anomalies.

### 2.4. Creating a reference series

In order to check the homogeneity of the data, reference series were created from adjacent stations, broadly following the GHCN method (Peterson and Easterling, 1994). A reference series was required for each calendar month, to permit more inhomogeneities to be identified (Section 2.2). Building a reference series from a single station, or a single set of overlapping station sections, relies too much on a single record that may have unusual features or even undetected inhomogeneities. Therefore, it is better to construct a number of such records ('parallels') and combine them, following the GHCN method. There are two key differences from GHCN at this point:

1. The GHCN method uses five parallels; here, five was an ideal maximum and two was the acceptable minimum, since it was better to check using a suboptimal number of parallels than not to check at all. The number of parallels was allowed to vary from one calendar month to another.

 

2. The GHCN method was tested on a simulated dataset in which all stations covered the same time period. Here, it was necessary to merge stations that only partially overlap into a single parallel. Since merging the first-difference series (used by GHCN) in this way would create an inhomogeneity, each parallel was constructed using absolute values.

When a reference series for a candidate station was to be constructed, the initial steps were to fill in any gaps in adjacent station records (Section 2.4.1) and identify suitable neighbours (Section 2.4.2). An iterative procedure was used to select the neighbours to use (Section 2.4.3). Once the selection was made, the neighbours were formed into parallels and the parallels combined into a reference series (Section 2.4.4).

*2.4.1. Completion of station records.* An incomplete station record could not be allowed to contribute to a reference series, because the missing values introduce inhomogeneities to the first-difference series (Section 2.2). The loss from excluding all incomplete station records would be prohibitive, so instead the missing data were replaced with estimates for the limited purpose of constructing a reference series.

The replacement was not done indiscriminately, because the reference series should be largely based on genuine data. Instead, an incomplete station record was subdivided into 'sections' of at least 5 years (10 years for precipitation) with relatively few missing data; periods with few valid data did not contribute to any reference series. Each section was individually correlated with its closest neighbours using least-squares regression. If a correlation was sufficiently high (0.2), then the relationship was used to replace the missing value, else it was replaced with the section mean. The correlation threshold was relatively low, since more-distant neighbours were less likely to be related, and a weakly correlated neighbour was likely to provide a better estimate than the section mean.

The method for precipitation was augmented because variations between two neighbouring stations are often related non-linearly. Prior to correlating, the neighbour was adjusted to make the relationship linear and, therefore, amenable to least-squares regression. The method of adjustment is described in Section 2.4.4.

*2.4.2. Correlation of neighbours.* Each reference series was built from neighbours where the first-difference series were highly correlated (at least 0.4). For precipitation the first-ratio series was used, any months without rainfall having been temporarily adjusted to 0.1 mm to avoid divisions by zero.

A separate set of neighbours was identified for each calendar month, because the strength of the relationship between one station and another may vary over the seasonal cycle. To limit the computational demands of the search, only the 100 closest stations within a reasonable distance from the candidate were considered. (The reasonable distance was the correlation decay distance, which will be given in Table II.) The initial weight was the square of the correlation coefficient; sections with a weight less than 0.16 (0.04 on the first pass) were discarded.

*2.4.3. Selection of neighbours.* The selection of a set of neighbours from which to form a reference series is not a trivial problem. The best choice depends on a number of decisions, including the proportion of the candidate record that must be matched, the trust placed in weakly correlated neighbours, and the benefits of a larger number of parallels.

An iterative procedure was developed to find an acceptable solution wherever possible. Figure 1 details the method of determining the part of the candidate record that may be matched by a reference series. Within this procedure it was necessary to attempt a match for a given period and calendar month; this was achieved by the sub-procedure in Figure 2. Here, the problem was restricted to identifying sections from neighbours that could be combined into parallels that extended the full length of the given period. Initially, an attempt was made to construct five parallels, but if this failed then a minimum of two parallels could be accepted.

When a solution was found it was given a score $z$. On the initial pass through the data (Section 2.3) the priority was to obtain the longest possible reference series, so in this special case $z = n_y$, and the omissions

 

Table II. The information on which the CRU TS 2.1 climate grids were based. The primary variables were based solely on station observations. For the secondary variables, the station data were augmented with synthetic estimates from the primary grids in regions where there were no stations within the correlation decay distance. The variables derived were obtained directly from the primary variables. Both the distances and the method of obtaining synthetic estimates were obtained from New *et al.* (2000)

| Type | Var. | Stations | Secondary | Distance (km) |
|---|---|---|---|---|
| Primary | tmp | tmp | — | 1200 |
| | dtr | dtr | — | 750 |
| | pre | pre | — | 450 |
| Secondary | vap | vap | from tmp and dtr | 1000 |
| | wet | wet | from pre | 450 |
| | cld | cld (1901–95), spc (1996–2002) | from dtr (1901–95) | 600 |
| | frs | — | from tmp and dtr | 750 |
| Dervied | tmn | — | from tmp and dtr | — |
| | tmx | — | from tmp and dtr | — |

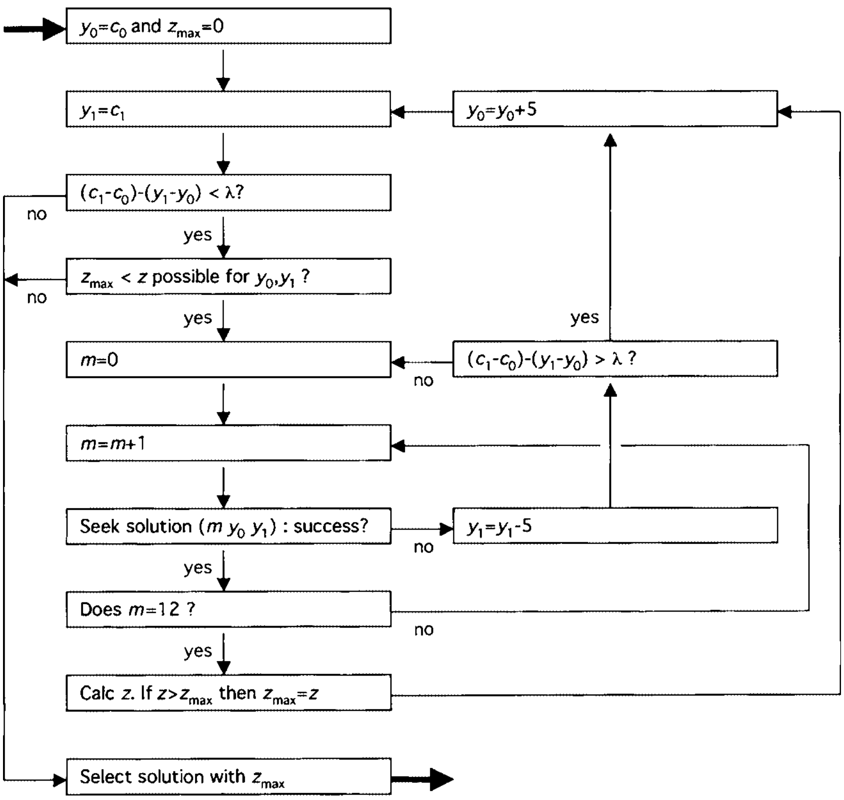


Figure 1. This diagram details the selection of a set of stations to form a reference series for a given candidate station. The period to be covered by the reference series ($y_0$, $y_1$) depends partly on the period covered by the candidate ($c_0$, $c_1$). The procedure considers each month $m$ individually and evaluates different alternatives using a score $z$. A limit λ may be placed on the number of years that may be present in the candidate, but not in the reference series. The 'seek solution' step is amplified in Figure 2

criteria λ was not set. On subsequent iterations (Section 2.3) the score was based on the number of parallels $p$ provided for each calendar month $m$, their length, and the weight $w$ attached to the section from which a

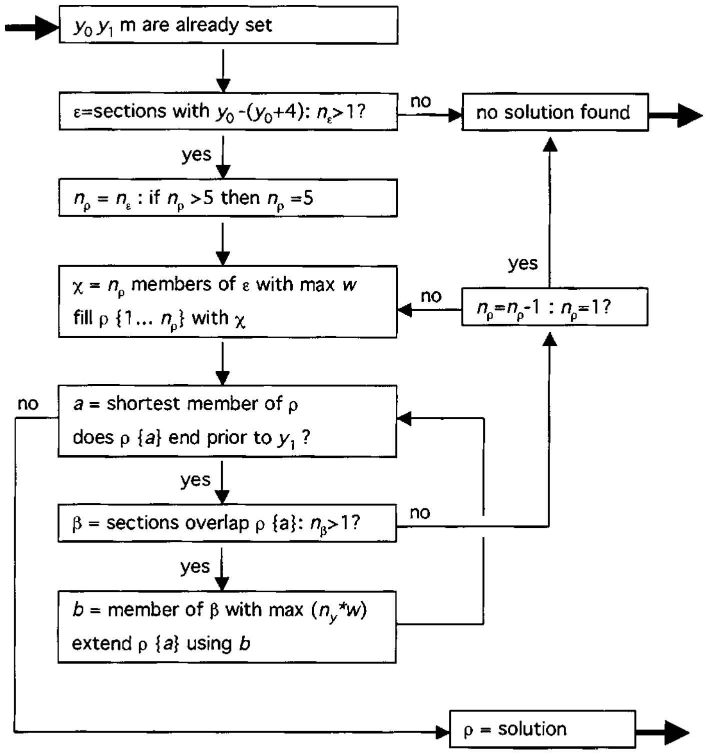


Figure 2. This diagram details the selection of a set of stations $\rho$ (a 'solution') for a given period ($y_0$, $y_1$), calendar month $m$, and using pre-identified neighbours (Section 2.4.2), each of which has a weight $w$ attached. The solution comprises two to five parallels $n_\rho$ that must each extend over the full given period. The parallels are initialized using the most highly weighted ($\chi$) among the sections ($\epsilon$) from those pre-identified neighbours that include the first 5 years (10 years for precipitation) of the given period. The parallels are then extended to cover the given period by identifying additional sections $\beta$ that overlap with the existing parallels, and selecting the section $b$ that can make the greatest contribution to the shortest parallel $a$

value was assigned to a particular year $y$ in a particular parallel.

$$z = \sum_{m=1}^{n_m} \frac{\sum_{p=1}^{n_p} \sum_{y=1}^{n_y} (w_{pym})}{\sqrt{n_{pm}}} \tag{1}$$

*2.4.4. Combination of neighbours.* An overlap of at least 5 years (10 years for precipitation) was required to merge sections from two stations into a single parallel; the overlap was used to adjust the later section to match the earlier section. If the overlap exceeded 10 years (20 years for precipitation), then the adjustment was based on the final 10 years of the overlap to reduce the probability of including any undetected inhomogeneities in the adjustment factor. For most variables the adjustment assumed a linear relationship between sections; precipitation was assumed to follow a gamma distribution, so sections might be related non-linearly.

For variables other than precipitation, the adjustment used the mean $\overline{x}$ and standard deviation $\sigma$ of the earlier (0) and later (1) sections. The original values $x$ in the later section were transformed as follows to give final values $y$:

$$y_1 = \overline{x}_0 + \frac{\sigma_0}{\sigma_1}(x_1 - \overline{x}_1) \tag{2}$$

 

For precipitation the adjustment used the scale ($\beta = \sigma^2/\overline{x}$) and shape ($\gamma = \overline{x}^2/\sigma^2$) parameters of the gamma distribution. Precipitation was adjusted thus:

$$y_1 = a x_1^b \tag{3}$$

The constant $b$ is the power to which the values of the later section had to be raised such that $\gamma_0 = \gamma_1$, and was obtained by iteration. The constant $a$ was obtained after raising the set of $x_1$ to the power $b$ and was given by $\overline{x}_0/\overline{x}_1$.

The two to five parallels for each calendar month were merged into a reference series matching the candidate station. Each parallel was adjusted to match the statistical characteristics of the candidate to avoid any implicit weighting, and was then explicitly weighted by the square of its correlation coefficient with the candidate. The weighted mean of the parallels was adjusted to match the statistical characteristics of the candidate, thus forming the reference series.

### 2.5. *Correction of inhomogeneities*

The detection of inhomogeneities employed the residual sum of squares (RSS) statistics from the GHCN method (Easterling and Peterson, 1995: 371), but applied them at the monthly time scale. Therefore, 12 series of the differences between the candidate and reference series were required. However, it was still assumed that any discontinuity would be introduced instantaneously, so any evaluation of discontinuities could not be treated independently from one calendar month to the next. A two-stage process was adopted:

1. $RSS_1$ and $RSS_2$ (see Easterling and Peterson (1995)) were calculated independently for each calendar month, and $RSS_2$ was made comparable across months by dividing it by $RSS_1$.
2. A single statistic for each year was obtained by averaging this ratio across all 12 months; the most suspicious year was given by the minimum of this time series.

The most suspicious year was evaluated by applying the $F$-test and $t$-test (after GHCN) to each of the 12 difference series. If either test yielded at least 3 months with significances of 95%, it was regarded as a potential break. If consecutive months in the difference series were statistically independent, then this condition would be met by chance on fewer than 2% of occasions; yet the condition is sufficiently relaxed to allow the detection of weak inhomogeneities that are strong in just one season. A non-parametric test was subsequently applied (after GHCN) with the same criterion of 3 months with significances of 95%.

If an inhomogeneity was confirmed, a correction value was obtained to apply to each calendar month. Since the samples on which the correction was based were often small, the correction values themselves were prone to inaccuracies, potentially causing misleading changes in the seasonal cycle. This risk was ameliorated by smoothing the set of 12 correction values using a Gaussian filter and adjusting to preserve the original mean and standard deviation.

Which part of the station record should be corrected? The decision depends on the eventual use of the record. Section 2.7 will describe how some methods interpolate between stations using absolute values, in which case it would be appropriate to correct all stations relative to their 'normal' value from a common baseline period (perhaps 1961–90). New *et al.* (2000) interpolated using anomalies, but calculated them using a supplementary source of normals; in this case it would be essential to correct all stations to match the baseline of the normals (again 1961–90). Therefore, neither of these methods can subsequently append any recent observations unless they too are corrected (see also Jones and Moberg, (2003)). The method adopted in Section 2.7 allows the station records to match any period, so they were corrected in such a way that the final values remain unchanged. Therefore, recent observations may be appended without difficulty.

### 2.6. *Merging*

Once a station had been checked and any inhomogeneities corrected, it was merged into the final database. This was achieved through the WMO code attached to the station. However, not all sources attach WMO codes to their data, and not all stations have been assigned WMO codes, so additional information was used: location, name and country. Each additional station was compared with the stations already in the database, both to avoid unnecessary duplication and to ensure that each station record is as complete as possible.

If an additional station was already present in the database, then the two records were compared. (Information from two or more sources may have been corrected differently for inhomogeneities, or may have been adjusted by others prior to acquisition.) The comparison was based on any available overlap between the records; if none was available, then an attempt was made to construct a reference series that overlapped both records (as in Section 2.4). If an overlap was found, then it was used to alter the statistical characteristics of the additional station to match those of the existing record, using the method in Section 2.4.4; the two records were then merged. If no overlap was found, then the records were assumed to be for different stations, because of the possibility of the two records having different normals.

Where the sources were very recent (CLIMAT and MCDW) the additional station was assumed to be the same without the above data check. This was justified because the normals from these sources were likely to be the same as the post-adjustment normals from other sources. This assumption was necessary for some climate variables (notably wet days) for which overlaps with stations from other sources were very rare; without it the normals could be calculated for very few recent data.

### *2.7. Converting to anomalies*

To obtain a climate grid of normals, the absolute values from all available stations might be used (e.g. New *et al.*, 1999). It is possible to construct a gridded time series similarly, by using all the absolute values available at each moment in time. However, this method is highly vulnerable to fluctuations in spatial coverage. For example, if there is a gap in the record at a mountain station, then the local value may be estimated by interpolating between adjacent valley stations. This vulnerability is so important that the interpolation must be restricted to the period for which there is an adequate set of stations with a complete record.

Although the normal may vary considerably over a small area, for most aspects of climate the variations from year to year take place on much larger spatial scales. This permits a great improvement in the method of constructing a gridded time series: anomalies are interpolated, rather than absolute values. Under the anomaly method (Jones, 1994; New *et al.*, 2000) the station time series may be expressed as anomalies relative to a chosen baseline period (1961–90), interpolated onto a grid, then combined with an equivalent grid of normals for the same baseline period. Stations with missing values may be included, unlike the 'first-difference method' (Peterson *et al.*, 1998b), since anomalies may be estimated from adjacent stations when it is not safe to estimate absolute values. (Section 2.8 will explain how unwarranted extrapolation is guarded against.) This method also uses all the spatial information that is available, unlike the 'reference station method' (Hansen and Lebedeff, 1987).

Therefore, the final database was converted into anomalies relative to the 1961–90 normal. Difference anomalies were used for all variables except precipitation and wet-day frequency, for which relative anomalies were used. For many stations the normal could be calculated from the existing series. However, since the normals influence every value from a station, it was important to ensure their accuracy. Therefore, any extreme values were omitted and counted as missing; extreme values were defined as those more than three (four for precipitation) standard deviations from the mean (Jones and Moberg, 2003: 213). A large number of missing values would also make the estimate of the normal inaccurate; so, if more than 25% of the values from 1961–90 were missing for any single calendar month, then the normal was not calculated.

One weakness of the anomaly method is that it excludes any station without the appropriate normal. New *et al.* (2000) alleviated this weakness by using a supplementary source of normals (WMO, 1996) to reduce the number of stations excluded through having too many missing values in the period 1961–90. However, this alleviation is necessarily restricted to stations that were taking measurements during the baseline period and, therefore, reporting to the WMO. There are no WMO normals for stations that ceased recording prior to 1961, or which began subsequent to 1990.

This weakness prompted a modification to the anomaly method. The number of stations with normals was not expanded using a supplementary source, but by estimating normals using neighbours. An attempt was made to create a reference series (including 1961–90) from adjacent stations, as described in Section 2.4. If successful, then the mean of the reference series during 1961–90 was taken as the normal for the candidate. Thus, normals were constructed not only for stations with missing values in the baseline period, but also for stations that did not even exist then.

The calculated anomalies were subjected to two further checks prior to interpolation. First, the three standard deviation limit was reimposed to exclude extreme values from the time series, not just the normals. Then, any stations within 8 km of each other were merged; this was partly to avoid introducing duplicate records into the interpolation, and partly to ensure that the interpolated surface varied at coarser spatial scales.

### *2.8. Gridding*

The station anomalies were interpolated onto a continuous surface from which a regular grid of boxes of 0.5° latitude and longitude was derived. To ensure that the interpolated surface did not extrapolate station information to unwarranted distances, 'dummy' stations with zero anomalies were inserted in regions where there were no stations or synthetic estimates within the correlation decay distance (Table II); thus, the gridded anomalies were 'relaxed' to zero. For primary variables, only the stations for those variables contributed to the interpolation; the secondary variables were augmented with additional ('synthetic') data derived from the primary variables. Details of the interpolation were given by New *et al*. (1999, 2000).

Since there were no station observations of cloud cover available after 1996, cloud anomalies were used for 1901–95 and sunshine duration anomalies used thereafter. Because of the short length of most sunshine records, the sunshine anomalies were calculated relative to 1994–2000 and corrected to be relative to 1961–90 using the cloud grids from CRU TS 2.0 (New *et al*., 1999), following Mitchell *et al*. (2004). The cloud and sunshine anomalies were merged under the assumption that they are of equal magnitude but opposite sign.

The anomaly grids were adjusted so that the 1961–90 mean was zero for every box and calendar month. The adjustment was an absolute value (a ratio for precipitation and wet-day frequency) and was applied throughout the series, with the exception of zero anomalies. The exception was to ensure that gridded anomalies relaxed to zero would take the value of the normal at the end of the process and, therefore, be identifiable by users.

The anomaly grids were combined with the 1961–90 normals (CRU CL 1.0; New *et al*., 1999) to obtain absolute values. Any impossible values were converted to the nearest possible value, and a fresh adjustment (using a ratio) made to ensure that the 1961–90 mean corresponded to the normal. In addition, the wet-day frequency normal and time series were not permitted to take a larger value (in days) than was recorded for precipitation (in millimetres) for that grid box. The final grids constitute CRU TS 2.1.

## 3. RESULTS

### *3.1. Station quality*

The homogenization of station records may be illustrated using two stations. The DTR record at Yozgat provided by GHCN shows a shift in 1973–74 in all seasons (Figure 3). The reference series shows no such change. The shift could be due to a station relocation; the station is at a high altitude (1298 m) in mountainous territory, so any station movement is likely to result in a change in altitude, and thus in the mean DTR. The shift is detected as an inhomogeneity and corrected using a fixed reduction (in degrees Celsius) that varies between calendar months.

The precipitation record at Zametcino (Figure 4) is notable for low totals and low variability in winter (November–March) during the period 1928–64. Since this feature is absent from the reference series, it may arise from a long-term undercatch of solid precipitation (e.g. Adam and Lettenmaier, 2003). The restriction of this feature to only part of the record may be due to instrument changes or to corrections previously applied to other parts of the record. This feature ought to be corrected to avoid spurious long-term changes in the station and subsequent grids. Making this particular correction does not imply that gauge undercatch is generally corrected in the grids, since this is dependent on the normal from New *et al*. (1999).

 

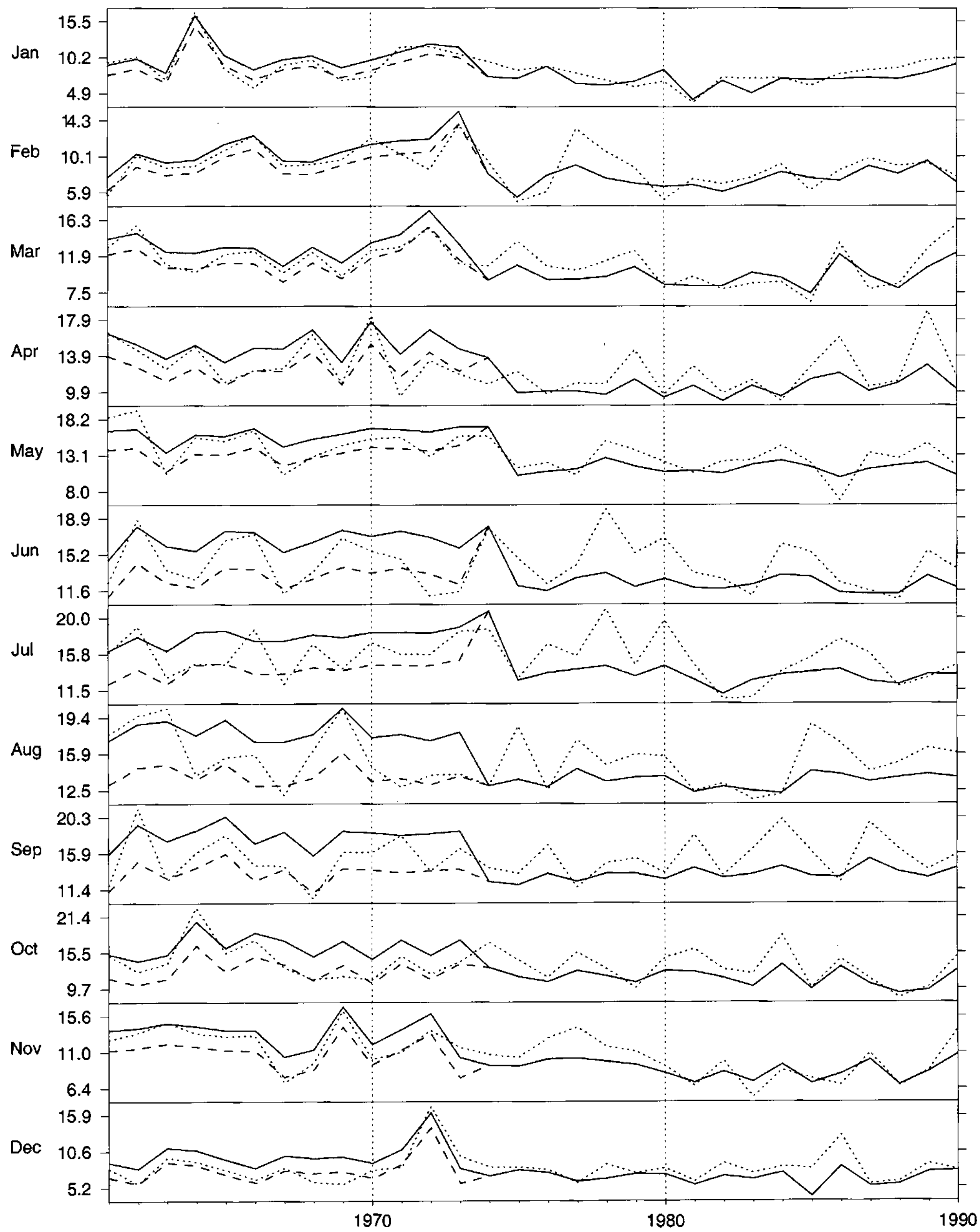


Figure 3. The DTR record for Yozgat (Turkey, 171 400, 39°49′N, 34°48′E) for each calendar month (in degrees Celsius). The solid line is the full record (1961–90) obtained from GHCN; the dotted line is the reference series; the dashed line is the final record after correcting the data prior to 1974

Three inhomogeneities were detected in the Zametcino precipitation record. The inconsistency of 1928–64 was successfully detected despite the inhomogeneities at the beginning and end applying only to the winter months. The series was corrected using a reduction by a fixed ratio, largest (2.33) in January during 1928–64. The detection at 1988 was probably erroneous, but the only substantial corrections were applied in March and July, resulting in inflated precipitation records in both months throughout almost all the record. The inflated records did not greatly affect the grids, because it is anomalies that are interpolated, not absolute values.

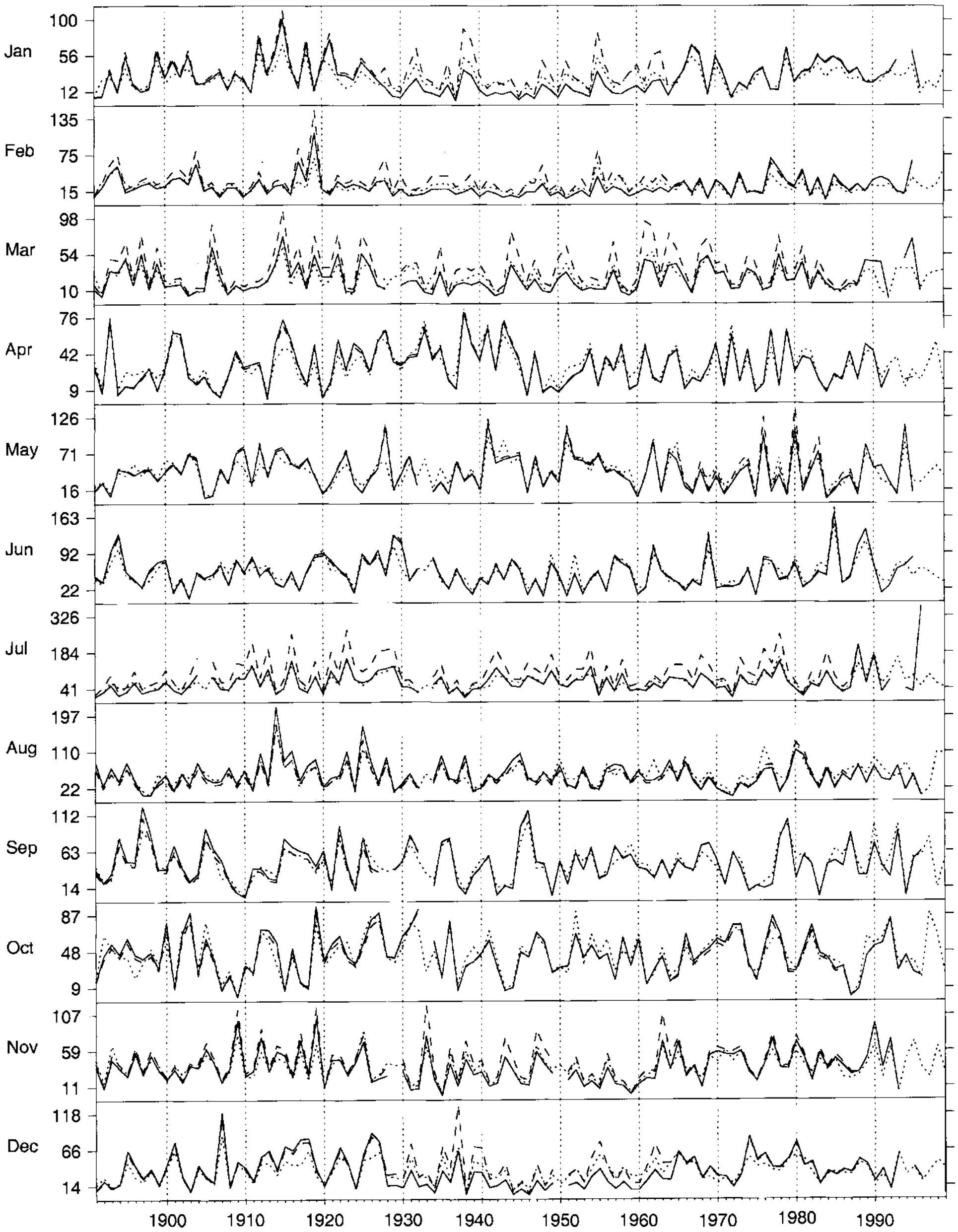


Figure 4. The precipitation record for Zametcino (Russia, 278 570, 53°30′N, 42°37′E) for each calendar month (in millimetres). The solid line is the full record (1891–1999) obtained from GHCN; the dotted line is the reference series; the dashed line is the final record after correcting at 1928, 1965 and 1988

### 3.2. Station totals

The total information acquired is indicated in Figure 5, which identifies the contribution from each source by variable and year. The sources with longer series all show a steady increase in the number of stations available during the 20th century, a peak around 1980, and a rapid decline to the present.

Jones provides carefully homogenized temperatures originally intended to monitor climate change and subsequently used in the detection of anthropogenically induced climate change. This source may be augmented with stations for which the long-term changes are not sufficiently accurate for detection, but

 

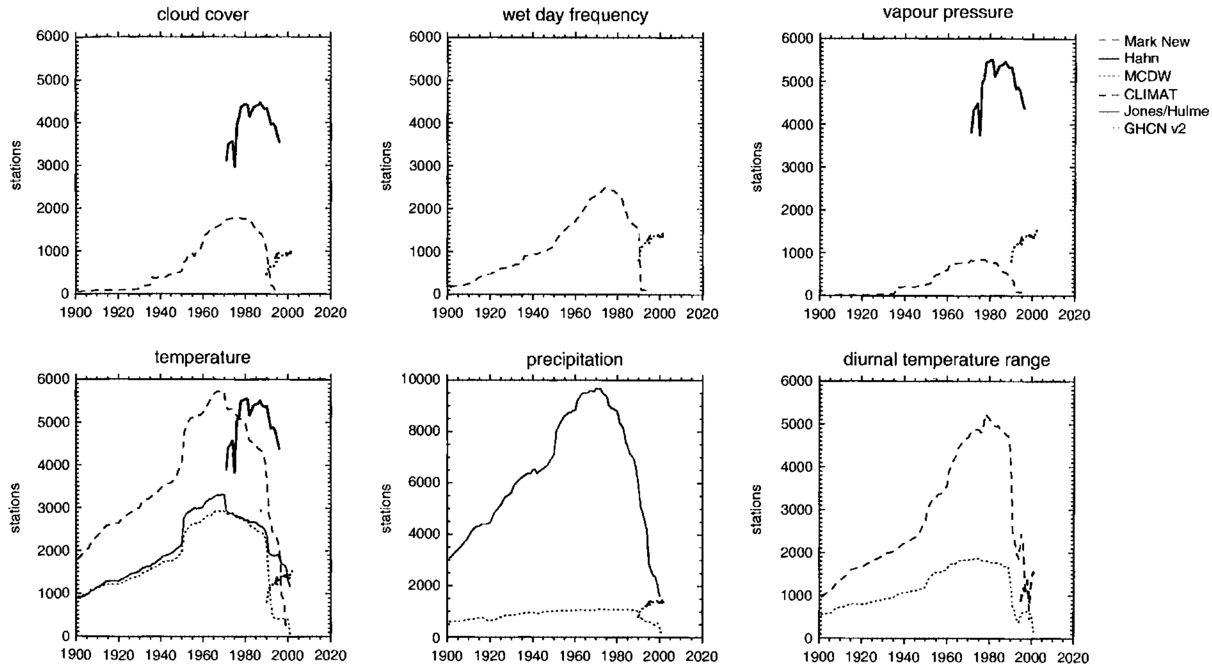


Figure 5. The amount of relevant data acquired, identified by climate variable, source and year. The cloud cover information includes cloud coverages (Hahn), sunshine durations (CLIMAT and MCDW), or both (Mark New); see Table I


Figure 6. The continental-scale regions used in summarizing the results. The regions were chosen on the basis of the classification of meteorological stations adopted by the WMO, with some further subdivisions

which are nonetheless a good record of year-to-year temperature variations. Precipitation is dominated by the Hulme source, but is extended in recent years by MCDW and CLIMAT. For a relatively short period (1971–96), Hahn increases by a factor of 3–5 the amount of cloud cover and vapour pressure data available.

The database constructed from these sources is summarized for a set of nine continental-scale regions (Figure 6) in Figure 7. The relatively abundant precipitation data was beneficial when interpolating, since precipitation has the greatest spatial variability. There are some source-related variations (notably from Hahn), but the network changes over the 20th century are remarkably consistent between regions. There are greater

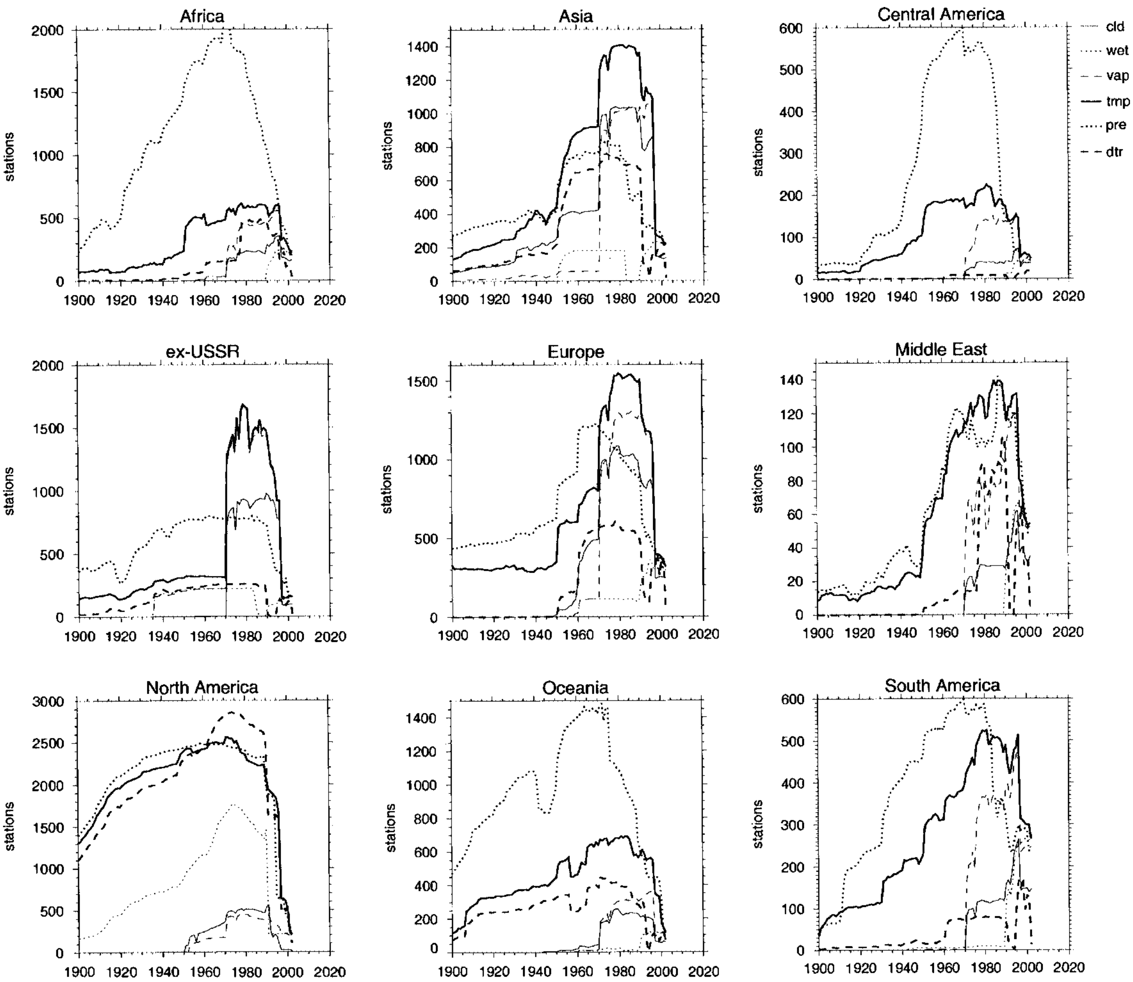


Figure 7. The size of the final station database for each climate variable, broken down by continent. All the data described in Figure 5 are included

regional variations in the average density of observation; the contrast between Europe and South America is particularly acute.

The temporal and spatial density of observations may be due to the limitations of this particular database, of data exchange and storage, or of the observing network.

- The evident improvement obtained through the Hahn source suggests that data storage is an issue. Hahn and Warren (1999) were able to build their database by gathering and editing surface synoptic weather reports. This task is resource intensive.
- The density of precipitation records in poorly observed regions reflects a long-term effort to obtain (through private contacts) information that is not publicly available (Mike Hulme, personal communication); evidently, data exchange is an important constraint.
- Although the shrinkage of the reporting network in recent decades is reflected in the early peaks around 1980, for some variables the recent decline is reduced or even reversed. This is largely due to the improved exchange of information through the CLIMAT messages and GCOS initiatives.
- The multi-variable databases that might otherwise be used for comparison have been incorporated as sources. Some single-variable databases match the density here. Xie and Arkin (1997) used 6700 precipitation gauges

 

for a short period (1979–96), but the spatial coverage was poor: half the 2.5° land grid-boxes were empty. Adler *et al*. (2003) achieved a similar density.

The database (Figure 7) includes all the available information, both checked and unchecked. It was possible to check a higher proportion of the data for regions and periods when the observed density is greater, and for variables (such as temperature) that vary on larger spatial scales (Figure 8).

When normals had been estimated for as many stations as possible, the absolute values in the databases were converted into anomalies (Figure 9). The proportion converted depended on two factors:

1. The number of stations with records of 1961–90 was critical. For example, cloud cover records began in 1950 in North America, but in 1971 in South America (Figure 7); therefore, anomalies could be calculated in North America, but not South America (Figure 9).
2. The spatial scales of interannual variability were also important. In Africa, despite the greater density of precipitation observations, a far higher proportion of temperature stations could be converted into anomalies.

The same two factors are also reflected in Figure 10, which displays the proportion of the database used in gridding. For the variables particularly dependent on the Hahn source (cloud cover and vapour pressure),

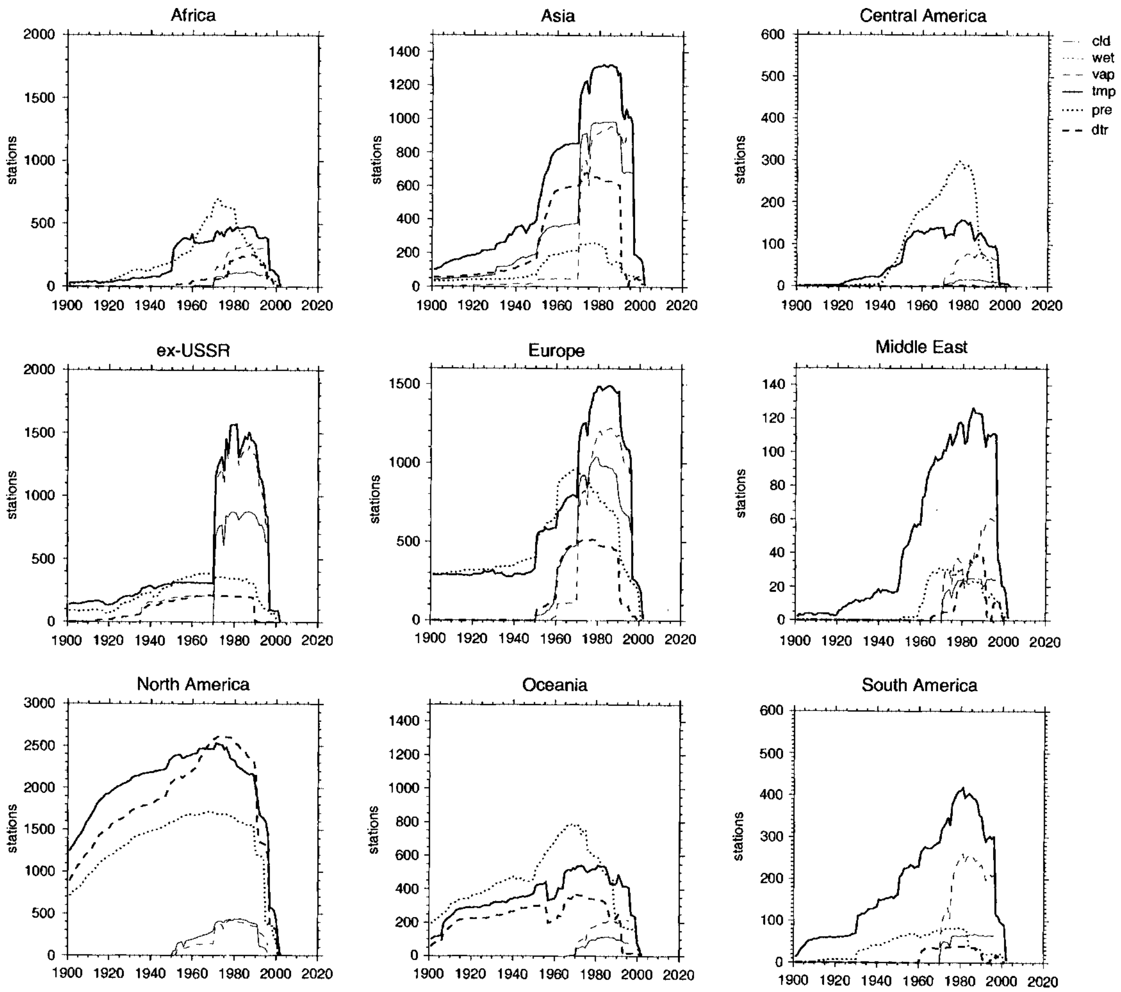


Figure 8. The subset of the final station database (Figure 7) for which it was possible to check for inhomogeneities. No wet-day frequencies were checked

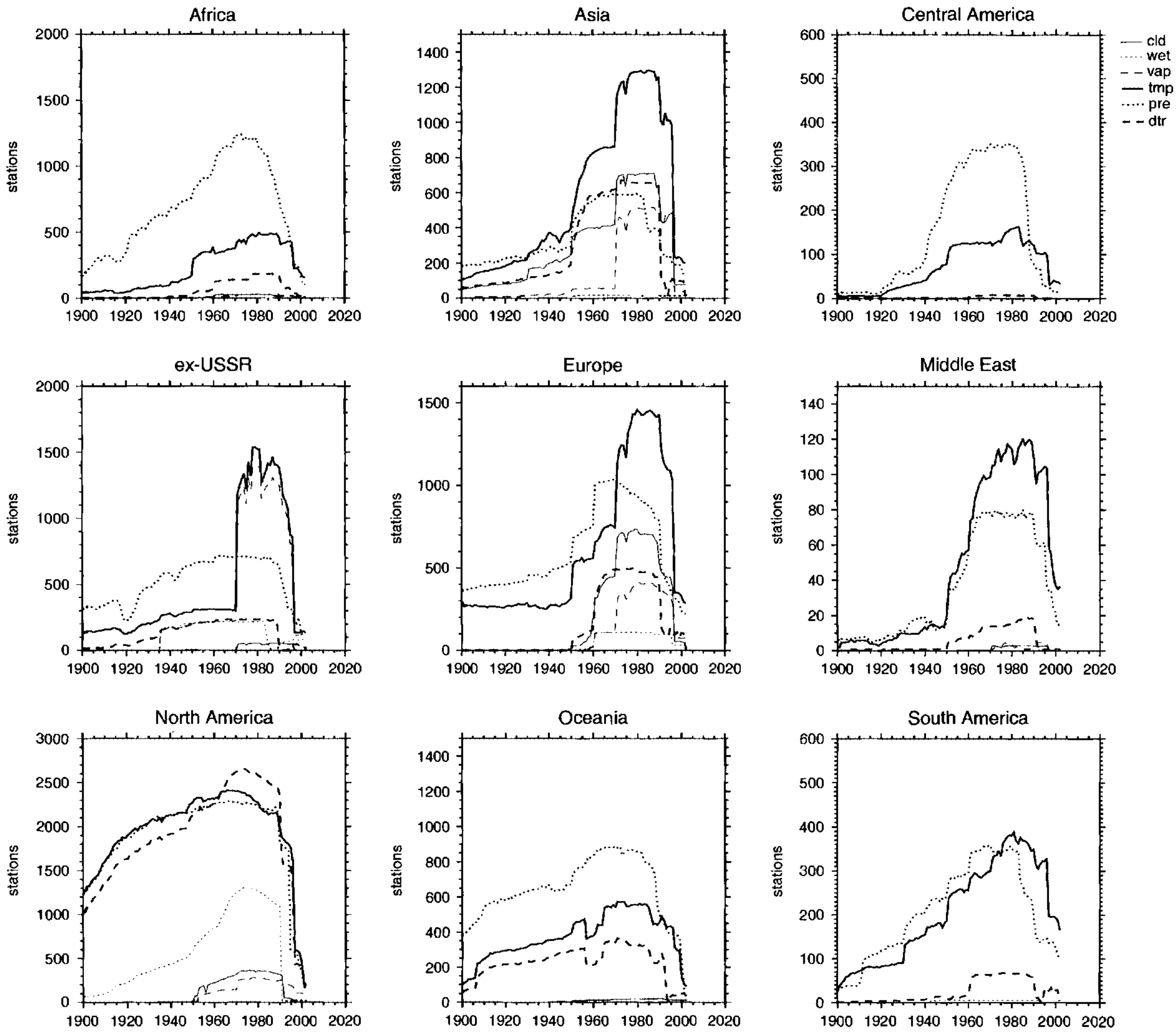


Figure 9. The subset of the final station databases (Figure 7) for which it was possible to convert the absolute values into anomalies

half the available data could not be used through lack of a normal. Therefore, a strategic investment in this database might aim to extend the work done by Hahn and Warren (1999) from 1971 back to 1961. The wet-day frequencies are dominated by the CLIMAT bulletins (Figure 5), which began in 1990; therefore, no normal could be calculated for a third of the data, and a further tenth represents overlaps between the CLIMAT and MCDW sources. Since precipitation is so spatially variable, a large proportion of those stations without the 1961–90 period were also without sufficiently well-correlated neighbours for the normal to be estimated.

### 3.3. *Climate grids*

The station anomalies were interpolated onto a 0.5° grid. Figure 11 shows the area for which non-zero anomalies were calculated. This provides an approximate measure of the area for which a genuine estimate could be made, instead of imposing a zero anomaly through a lack of observations. The estimate is slightly biased, since some genuine estimates are included among the zero anomalies. The bias is likely to be greatest for DTR and smallest for precipitation. Nonetheless, the proportion of the land surface with estimates is much higher for temperature and precipitation than for DTR. The relatively poor coverage of DTR is particularly

 

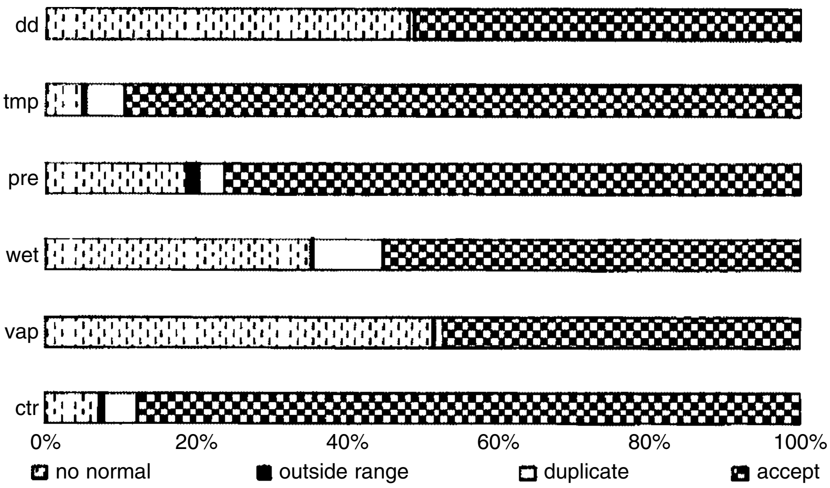


Figure 10. The proportion of the final station databases that were accepted for use in gridding (accept) and the proportions rejected because no normal could be calculated (no normal), because the calculated value lay outside the acceptable range of values (outside range), or because the station was within 8 km of another station with an equivalent value (duplicate). All data are included, not just 1901–2002

damaging, because five of the secondary variables were at least partly derived from it (Table II). The poor coverage arose from:

- a lack of observations (Figure 7); for example, that the area covered in Central America always exceeded 60% must be largely due to interpolation from stations in North America;
- the relatively low correlation decay distance (Table II).

Outside Europe, Asia and North America, there were very few cloud cover observations available (Figure 9). The DTR observations were interpolated to provide synthetic estimates of cloud, and the final cloud grids were interpolated from the synthetic and direct observations. Thus, large areas of the final cloud grids may be based on a very small number of DTR observations. This explains why cloud cover (and the other secondary variables) could be estimated over such large areas, but it also exposes the weakness with which these grids are likely to represent actual cloud variations. However, there are substantial problems with direct cloud observations prior to the 1950s (Moberg *et al*., 2003). The double interpolation explains how cloud cover could have better coverage than DTR.

## 4. CONCLUSIONS

A database of stations of monthly variations in climate has been constructed from various sources following New *et al*. (2000) and Mitchell *et al*. (2004). A large proportion of the data were checked for inhomogeneities using an automated method, developed from the GHCN method (Peterson and Easterling, 1994; Easterling and Peterson, 1995). Since any inhomogeneities were corrected so as to make the record consistent with its final values, near-real-time observations may be appended without introducing inhomogeneities. The method developed offers a number of improvements:

1. It is an iterative method, in which a subsection of a candidate may be checked if the full record cannot be checked, but in which the amount of unchecked data is minimized.
2. Incomplete station records are used in constructing reference series where the temporal data density warrants it. The gaps are filled by correlating with neighbouring stations.

 

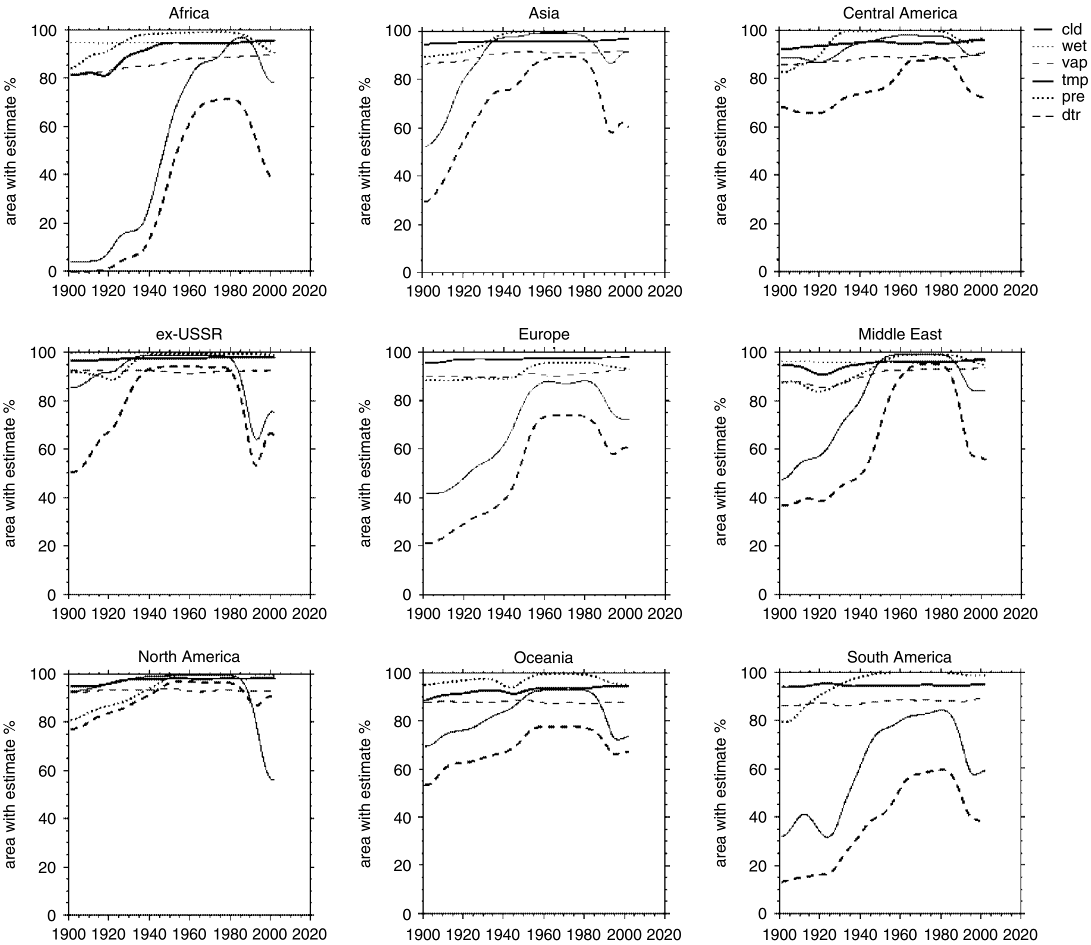


Figure 11. The approximate percentage of the land surface in CRU TS 2.1 with an estimated anomaly (relative to the 1961–90 normal); the remaining area is 'relaxed' to the normal. The percentage is approximate because the remaining area may include some genuine estimates of zero anomalies. The six climate variables represented in the station database are shown. The percentage given is of area rather than grid boxes. The mean of 12 monthly percentages is calculated for each year and the series smoothed with a 30 year Gaussian filter

3. A first-difference series is used to judge the correlation between stations, so that neighbouring stations with similar inhomogeneities are not more highly correlated than with homogeneous neighbours. The development is that anomaly series are used elsewhere, to avoid introducing inhomogeneities into the reference series.
4. Records that only partially overlap with the candidate may be utilized by merging series from two or more neighbours.
5. Stations are selected to form a reference series using a subordinate iterative procedure that balances the objectives of including as much as possible of the period covered by the candidate, using the most highly correlated neighbours and using multiple records.
6. The homogeneity of the candidate is independently checked for each monthly series, and a decision is reached on whether an inhomogeneity has been detected by combining information from each of the 12 sources.

This method of detecting inhomogeneities has its weaknesses. One weakness is that it is designed to detect abrupt rather than gradual inhomogeneities, although gradual inhomogeneities will also be detected unless they are widespread. This method also has none of the advantages of a manual method; an automated method is essential to handle such large quantities of data. However, the method should be sufficient for a database designed to provide best estimates of interannual variations rather than detection of long-term trends. A potential weakness is the adjustment of the absolute values in the 1961–90 period to make them consistent with the final values in a series. This is satisfactory when anomalies are required, but would be a fault if a climatology was being constructed.

Records from different sources were combined into a single database principally through the WMO codes attached to the stations. This process was refined to avoid unnecessary duplication and to combine fragmented records into a longer series, which is more useful. Adjacent station records were checked; any overlap was used to merge the records. If the records did not overlap, then a reference series was constructed to provide an overlap.

The description of the database exposed the sparse coverage of some variables in certain regions and periods, due partly to deficiencies in the observing network, the storage of observations, and their exchange. Converting the database to anomalies resulted in a substantial loss of data, which was reduced by estimating normals using reference series. The loss reached one-half of the cloud cover and vapour pressure records, because of their dependence on the Hahn and Warren (1999) dataset. A strategic investment in the station database to extend that dataset from 1971 back to 1961 could potentially incorporate into the grids triple the number of data involved in the extension, double the number of cloud cover and vapour pressure measurements incorporated into the grids, and eliminate the need for synthetic estimates of cloud cover and vapour pressure after 1960.

The station anomalies were interpolated onto a regular latitude–longitude grid following New *et al.* (2000) and adjusted to correspond to the published normals (New *et al.*, 1999). For temperature and precipitation, estimates were made for 80–100% of the land surface. The sparser coverage for DTR weakened the extent to which the grids of the secondary variables represent interannual variations, since five of the variables depend on estimates from DTR. Therefore, a priority for future work should be to expand the DTR coverage in regions and periods where it remains sparse.

The set of grids extend from 1901 to 2002, cover the global land surface (excluding Antarctica) at a 0.5° resolution, and provide best estimates of month-by-month variations in nine climate variables. This dataset is labelled CRU TS 2.1 and is publicly available (http://www.cru.uea.ac.uk/).

## REFERENCES

Adam JC, Lettenmaier DP. 2003. Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research–Atmospheres* **108**(D9): 4257. DOI: 10.1029/2002JDOO2499.

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie PP, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E. 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology* **4**(6): 1147–1167.

Brewer AM, Gaston KJ. 2003. The geographical range structure of the holly leaf-miner. II. Demographic rates. *Journal of Animal Ecology* **72**(1): 82–93.

Casey KS, Cornillon P. 1999. A comparison of satellite and *in situ*-based sea surface temperature climatologies. *Journal of Climate* **12**(6): 1848–1863.

Chen JM, Ju WM, Cihlar J, Price D, Liu J, Chen WJ, Pan JJ, Black A, Barr A. 2003. Spatial distribution of carbon sources and sinks in Canada's forests. *Tellus, Series B: Chemical and Physical Meteorology* **55**(2): 622–641.

Easterling DR, Peterson TC. 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369–377.

Eischeid JK, Diaz HF, Bradley RS, Jones PD. 1991. A comprehensive precipitation data set for global land areas. DOE/ER-69017T-H1, TR051, United States Department of Energy, Carbon Dioxide Research Program, Washington, DC.

Hahn CJ, Warren SG. 1999. Extended edited synoptic cloud reports from ships and land stations over the globe, 1952–1996. ORNL/CDIAC-123, NDP-026C, CDIAC, ORNL, US DoE, Oak Ridge, TN.

Hansen JE, Lebedeff S. 1987. Global trends of measured surface air temperature. *Journal of Geophysical Research* **92**: 13 345–13 372.

Huffman GJ, Adler RF, Arkin PA, Chang A, Ferraro R, Gruber A, Janowiak J, McNab A, Rudolf B, Schneider U. 1997. The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society* **78**(1): 5–20.

Hulme M, Osborn TJ, Johns TC. 1998. Precipitation sensitivity to global warming: comparison of observations with HadCM2 simulations. *Geophysical Research Letters* **25**: 3379–3382.

Jones PD. 1994. Hemispheric surface air temperature variations: a reanalysis and update to 1993. *Journal of Climate* **7**: 1794–1802.

Jones PD, Moberg A. 2003. Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *Journal of Climate* **16**: 206–223.

 

Kuhn KG, Campbell-Lendrum DH, Armstrong B, Davies CR. 2003. Malaria in Britain: past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America* **100**(17): 9997–10 001.

Mitchell TD, Carter TR, Jones PD, Hulme M, New M. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Tyndall Working Paper 55, Tyndall Centre, UEA, Norwich, UK. http://www.tyndall.ac.uk/ [Last accessed 19 April 2005].

Moberg A, Alexandersson H, Bergstrom H, Jones PD. 2003. Were southern Swedish summer temperatures before 1860 as warm as measured? *International Journal of Climatology* **23**(12): 1495–1521.

New M, Hulme M, Jones PD. 1999. Representing twentieth century space–time climate variability. Part 1: development of a 1961–90 mean monthly terrestrial climatology. *Journal of Climate* **12**: 829–856.

New M, Hulme M, Jones PD. 2000. Representing twentieth century space–time climate variability. Part 2: development of 1901–96 monthly grids of terrestrial surface climate. *Journal of Climate* **13**: 2217–2238.

Peterson TC, Easterling DR. 1994. Creation of homogenous composite climatological reference series. *International Journal of Climatology* **14**: 671–679.

Peterson TC, Vose RS. 1997. An overview of the Global Historical Climatology Network temperature database. *Bulletin of the American Meteorological Society* **78**: 2837–2848.

Peterson T, Daan H, Jones P. 1997. Initial selection of a GCOS surface network. *Bulletin of the American Meteorological Society* **78**: 2145–2152.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Forland E, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998a. Homogeneity adjustments of *in situ* atmospheric climate data: a review. *International Journal of Climatology* **18**: 1493–1517.

Peterson TC, Karl TR, Jamason PF, Knight R, Easterling DR. 1998b. The first difference method: maximizing station density for the calculation of long-term global temperature change. *Journal of Geophysical Research* **103**: 25 967–25 974.

Peterson TC, Vose R, Schmoyer R, Razuvaev V. 1998c. Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *International Journal of Climatology* **18**: 1169–1179.

Susskind J, Piraino P, Ixedell L, Mehta M. 1997. Characteristics of the TOVS pathfinder path A dataset. *Bulletin of the American Meteorological Society* **78**: 1449–1472.

Vose RS, Schmoyer RL, Steurer PM, Peterson TC, Heim R, Karl TR, Eischeid J. 1992. The Global Historical Climatology Network: long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041. (Available from CDIAC, Oak Ridge National Laboratory.)

corpauWMO. 1996. Climatological normals (CLINO) for the period 1961–1990. World Meteorological Organization Document WMO/OMMNo. 847, Geneva, Switzerland.

Xie P, Arkin PA. 1997. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society* **78**: 2539–2558.

 