Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010

by

P.D. Jones (1, 3), D.H. Lister (1), T.J. Osborn (1), C. Harpham (1), M. Salmon (1) and C. P. Morice (2)

(1) Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK
(2) Met Office Hadley Centre, Exeter, UK
(3) Center of Excellence for Climate Change Research / Dept of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

January 2012 (revised version – each revision highlighted)

Abstract

This study is an extensive revision of the Climatic Research Unit (CRU) land station temperature database that has been used to produce a grid-box dataset of 5º latitude by 5º longitude temperature anomalies. The new database (CRUTEM4) comprises 5583 station records of which 4842 have enough data for the 1961-90 period to calculate or estimate the average temperatures for this period. Many station records have had their data replaced by newly homogenised series that have been produced by a number of studies particularly from National Meteorological Services (NMSs).

Hemispheric temperature averages for land areas developed with the new CRUTEM4 dataset differ slightly from their CRUTEM3 equivalent. The inclusion of much additional data from the Arctic (particularly the Russian Arctic) has led to estimates for the Northern Hemisphere (NH) being warmer by about 0.1°C for years since 2001. The NH/SH warms by 1.12/0.84°C over the period 1901-2010. The robustness of the hemispheric averages is assessed by producing five different analyses each including a different subset of 20% of the station time series and by omitting some large countries.

CRUTEM4 is also compared with hemispheric averages produced by reanalyses undertaken by the European Centre for Medium-Range Weather Forecasts (ECMWF) - ERA-40 (1958-2001) and ERA-Interim (1979-2010) datasets. For the NH, agreement is good back to 1958 and excellent from 1979 at monthly, annual and decadal timescales. For the SH agreement is poorer, but if the area is restricted to the SH north of 60°S the agreement is dramatically improved from the mid-1970s.

1. Introduction

The purpose of this paper is to revise, improve and update the gridded land-based Climatic Research Unit (CRU) temperature database (CRUTEM4), last documented by Brohan *et al*. (2006, CRUTEM3). There are two principal reasons for such an analysis at the present time. First, some years have passed since it was last undertaken and significant changes and improvements have been made to the availability of monthly average temperature data in real time. The second reason is that several national and other initiatives (co-ordinated by National Meteorological Services, NMSs) have also dramatically improved the quantity and quality of monthly-mean temperature data available. Some countries have extensively homogenised significant parts of their entire national holdings, releasing the results for all to use. Both these developments should improve the coverage of available data.

Despite these improvements to the quantity and quality of data available, it is not expected that major changes will occur in the hemispheric-average series, as at these scales the existing averages are highly robust. The principal reason for expecting only small changes is that time series of the many thousands of station records are not statistically independent of each other. The number of statistically-independent locations (at timescales above annual) over the Earth's surface has been estimated by several authors to be about 100 or less (see discussion in Jones *et al*., 1997). The improvements to data quality and quantity in the present study, though, should impact individual grid-box series and analyses of spatial patterns.

The paper is organised in the following way. Section 2 extensively discusses the sources of additional data used in CRUTEM4 and the challenges of merging, replacing and updating the existing station-based records. Section 3 discusses the gridding technique used to develop the improved grid-box datasets. Section 4 presents extensive comparisons of the new analyses with those already available, illustrating the improvements in coverage. Section 5 concludes.

2. Data

The station data sources incorporated into previous versions of the CRUTEM database have been extensively discussed in Jones *et al*. (1985, 1986), Jones (1994), Jones and Moberg (2003) and Brohan *et al*. (2006). The station data used in the CRUTEM3 dataset had assigned codes to each station giving the principal source for each series (see above references). These have been augmented here and a full list of source codes is given in Table 1. Although the ultimate sources of all the station data are the NMSs, much of these data have made their way to users via a number of World Meteorological Organization (WMO) and Global Climatological Observation System (GCOS) initiatives, as well as NMS websites and scientific publications. We have replaced station data in CRUTEM3 with improved data from NMSs for stations with the same locations as these were deemed to be of better quality. In some cases, the improvement could simply have been a more complete series with fewer missing monthly values.

The next sections introduce much of this additional material, but only the major source codes in Table 1 are discussed. Apart from NMS source material there are three additional sources that incorporate station data across the world’s land areas: CLIMAT (WMO co-ordinated transmission of many meteorological parameters including monthly average temperatures), Monthly Climatic Data for the World (MCDW), and the decadal World Weather Records (WWR) volumes (from the 1950s onwards up to the 1990s). CLIMAT and MCDW are sources that are available in real time and near-real time respectively, and contain data for approximately 2000-2500 stations, though the number of stations available varies from month to month, particularly so for some developing countries. MCDW is available slightly later (3-4 months) than CLIMAT and tends to contain the same stations (though with fewer missing values), but considerably more for the contiguous United States (US). We do not use all the station data that report in CLIMAT and MCDW, but restrict ourselves to stations that have enough data to calculate 1961-90 averages (see section 3.1).

The WWR volumes are released every ten years after the completion of each decade. WWR is an important source of data for South America, Africa, Asia and many island groups. The availability of WWR data only every decade is part of the reason why the coverage of data in near-real time appears to reduce since the last decade of WWR was released for the 1990s. Part of this reduction is due to incomplete availability rather than the non-existence of data and should not be interpreted as evidence that the network of stations across the world’s land area is reducing. WWR sources can additionally be important in other parts of the world for infilling missing monthly values that occasionally occur in CLIMAT and MCDW sources.

The numbers of stations from each source are included in Table 1. Although each station is allocated a source code, most station series do not come from a single source (see also Jones and Moberg, 2003 and Brohan *et al*. 2006). Real-time monthly updating has to be based on CLIMAT and MCDW data, and most NMSs do not fully assess the quality of these data in real time. The CLIMAT and MCDW data are quality controlled by Meteorological Office staff. Within a few years we would expect to replace the recent data for some series with data from direct NMS sources or from the 2001-2010 WWR volume when it becomes available. Further details about updating are given in Section 2. A station series is, therefore, often based on a combination of multiple sources: the source code given in Table 1 for each station indicates only the dominant source code. The ordering of the updating affects (to some extent) the exact number of sites added from each source.

Another potential, extensive source of additional data is daily and hourly Synoptic Reports (SYNOP). SYNOP data also include many other weather variables and are one of the principal sources of input data for operational weather forecasts. We have never used data from the SYNOP source in our earlier versions of CRUTEM, and continue to exclude it from the new database. There are a number of reasons for this. First, SYNOP data are operational in nature, so are not always extensively quality controlled by NMSs. Second, their coverage tends to be denser in regions where we already have many series. Finally, monthly averages derived from SYNOP data are often found to be biased compared to CLIMAT and MCDW data for several reasons (see discussion in van den Besselaar *et al*., 2011). These reasons include: incomplete numbers of days in each month and the daily maximum and minimum temperatures not necessarily being the true values in mid and high latitude regions of the world. Additionally, many countries do not calculate monthly averages from monthly mean maximum and minimum temperatures averages, so potential biases will be introduced into series updated with SYNOP data.

United States

Previous versions of CRUTEM incorporated more stations for this region than any other land area. Our earlier work used almost all the station series available from CLIMAT and MCDW. The only series we excluded were those that we had deemed to have non-correctable inhomogeneities which we documented in Jones *et al*. (1985, 1986). For CRUTEM2 (Jones and Moberg, 2003) this was supplemented by an additional 1023 series for the contiguous US, but these all ended in 1996. We never sought to update these data for CRUTEM3, as the number reporting from CLIMAT and WWR for this region was already denser than any other region of the world (see discussion in Jones and Moberg, 2003). With CRUTEM4 we have replaced the 1023 series with 892 series from the current US Historical Climatic Network (USHCN, which contains 1218 stations for the contiguous United States, see code 44 in Table 1) described by Menne *et al*. (2009). The version we have used includes adjustments for time of observation bias and site relocations (see details in Menne *et al*., 2009). As many of the additional USHCN series (i.e. the 1220 minus the 892) report through CLIMAT or MCDW, we have replaced our original series for these locations with USHCN data. With both additions we had to ensure that no data series appeared twice. Additionally, the earliest year in all the USHCN series is 1895, so in order not to lose any useful 19th century data from the series we replaced, we compared USHCN series with those from the replaced set during the 1895-1900 period and kept any pre-1895 data where there was no step jump in 1895. Of the 892 USHCN stations incorporated into CRUTEM4, 525 stations had additional years added before 1895. The USHCN data we use will be periodically updated from the above source. Later we will show that the contiguous US has only a negligible impact on average NH temperatures, by removing all station data from this region.

Russian Federation

Monthly temperature time series for 475 stations were obtained from the All Russian Research Institute of Hydrometeorological Information - World Data Center (RIHMI-WDC, see Code 43 in Table 1). We compared these data series with those we already held and identified three groups of stations: those in common to both datasets (131), those only in the CRUTEM database and those only in the RIHMI-WDC dataset (344). The latter group were incorporated into CRUTEM4, and those stations unique to CRU were retained. For the 131 stations in common, comparison revealed differences for some of the series. The differences were of two kinds: (1) systematic offsets between the data series (consistently differing for different months of the year) very suggestive of homogeneity adjustments having been applied to RIHMI-WDC data and (2) apparently random differences. We are confident that the systematic offsets were applied to the data obtained from RIHMI-WDC rather than to our CRUTEM data, since the latter come from earlier World Weather Record (WWR) sources and we applied few adjustments to former Soviet Union (fUSSR) data in the 1980s (see details in Jones *et al*., 1985, 1986).

The apparently random differences were also assessed and while the Russian source mostly seemed to be a more reliable value (compared with neighbouring stations) this was not always the case. We contacted the Russian NMS and sought to find any documentation about the systematic and random differences. We were not successful in finding any information for the systematic differences, but received considerable help with the random ones. At the end of the exercise, the number of sites in CRUTEM4 was increased by 344 (i.e. the number in the third category above). For some other sites, the majority of the series came from this source, so these are also classified as source 43 (see Table 1).

Former Soviet Union

For countries entirely within the former Soviet Union (fUSSR) we updated data from daily data from 223 locations in the fUSSR, also downloaded from RIHMI-WDC (Code 51 in Table 1). We downloaded series from 1990 onwards (for series already in the CRUTEM database) which offered useful updates, recalculating monthly averages from the daily data in the archive. Most of these series are within Russia, but there were series for other fUSSR countries.

Additionally for central Asian countries within the fUSSR, we added in additional data from the National Snow and Ice Center (NSIDC) in Boulder, CO, choosing only stations for which we already had some temperature data (Williams and Konoyalov, 2008; see Code 50 in Table 1). The records for 61 series within Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan were extended and/or improved (fewer missing values).

Canada, Australia and New Zealand

In both our previous two versions, CRUTEM2 (Jones and Moberg, 2003) and CRUTEM3 (Brohan *et al*., 2006) we have incorporated Canadian station temperature series which have been tested for homogeneity and adjusted for discontinuities due to site relocations and changes in observing procedures (Vincent and Gullett, 1999 and Vincent *et al*., 2002). The convention followed by CRU in the 1980s (e.g. Jones *et al.*, 1985) was that all necessary homogeneity adjustments were applied to the earlier part of a station timeseries, so that ongoing updates could be appended to the modern end of the series without the need for them to be adjusted. The adjustments applied by Vincent and Gullett (1999) and Vincent *et al*. (2002) have not followed this convention. Some minor further adjustments have been applied to the data since its last update in 2008 (see Code 42 in Table 1) to address the change in observing time at airport stations in the eastern regions of the country (discussed briefly in Brohan *et al*., 2006). We apply these adjustments, therefore, to real-time CLIMAT updates for sites in this region prior to appending them to the modern end of a series, so that they are homogeneous with the past data.

The station data we are using for Australia and New Zealand were discussed briefly in Brohan *et al*. (2006). Source details (web sites or literature references) for these and other groups in this section are given in Table 1 (codes 40 and 41). For CRUTEM4, we downloaded these homogenized data again and checked against what we had, incorporating all the changes made in Australia and New Zealand.

Arctic

Bekryaev *et al*. (2010) analyzed recent trends in Arctic temperatures. In order to improve coverage across the Eurasian and North America parts of the Arctic, they have gained access to more series (with respect to the series already in CRUTEM3) from the region. This dataset was compared with the CRUTEM database (after the inclusion of the additional Russian and Canadian data discussed earlier). From this source 125 stations were new to CRUTEM (coming mainly from Alaska, Canada and Russia, but Greenland is considered separately later). Additionally many of the other records extended some series and/or made some series more complete, so were added where there was good overlap agreement. It may seem somewhat surprising that there are more data than analysed or available from the Russian and Canadian NMS. This just illustrates that personal contact has the potential to elicit additional sources beyond those that an NMS makes available over WMO systems such as CLIMAT or via its web site. Also, many NMSs often have sites classified as being first- or second-order stations or being climatological or agri-climatological stations, so some series may not be available in near-real time to the NMS. Such sites may be considered not available to be transmitted over CLIMAT, so are not made available for other sorts of international exchange.

Greater Alpine Region

This is a network of 141 stations for the Greater Alpine Region (GAR) developed during a number of projects led by the Austrian Meteorological Service. Many of the data series extend back to the 18th century and cover Austria, Switzerland, Slovenia, Croatia and parts of France, Italy, Germany, the Czech Republic, Slovakia and Hungary. These data have been extensively assessed for long-term homogeneity (see

Auer *et al*., 2001) and have additionally been adjusted for biases in the period before the introduction of ‘screened’ thermometer housing (Böhm *et al*., 2010). The issue of the introduction of thermometer screens will be discussed further in section 4. Temperature data for 107 stations were added. The additional 34 are either precipitation-only measuring stations or their data were of insufficient length for inclusion. The HISTALP source (Code 49 in Table 1) does not include the Swiss stations. These were added from a different source (Code 52 in Table 1).

Greenland, Faroes and Denmark

Long mean temperature station series for Denmark (5), Faroes (1) and Greenland (7) were updated or added using data from recently completed Danish Meteorological Institute (DMI) reports (Cappeln *et al*., 2010, 2011) and (Vinther *et al*. 2006). The two DMI reports are given separate source codes (47 and 48 in Table 1).

WWR Decadal Volumes

When the 1991-2000 WWR decade was received (~2006) we were able to infill significant numbers of missing values in our CRUTEM3 monthly series. The 1961-1990 volumes had been assessed for additional series in the course of the development of CRUTEM2 and CRUTEM3 (Jones and Moberg, 2003 and Brohan *et al*., 2006), but the additional data from the 1991-2000 volumes were not always included. During the development of the CRUTEM4 database, it was realized that some of the series we had which came from Global Historical Climatology Network version 2 (GHCNv2, see Jones and Moberg, 2003) did not always include data from earlier WWR decades (for 1961-70, 1971-80 and 1981-90).

GHCNv2 kept all sources of data separately for each station by the use of version numbers. The problem of deciding which might be the best source for a given year has been partly resolved within GHCNv3 (http://www.ncdc.noaa.gov/ghcnm/v3.php) as a single series has been developed for each location. We say partly, as there has to be some automation in any decision and without manually checking each it is unlikely to be the best source in every case. As we noticed that some of our station series did not include the WWR data (which we almost always deemed to be of better quality than that received over CLIMAT and/or MCDW) we checked the data series we had for the three decades (1961-90) against the WWR data. For a few stations we added in the WWR source (Code 37 in Table 1) mainly for series from South America, Africa, southern and eastern Asia, parts of Europe and for many island groups around the world.

How many CRU homogenized series remain in the CRUTEM4 database?

Inhomogeneities may be introduced into a station series by a variety of effects, such as changes in instrument location, local environment, exposure or recording practices (issues discussed in Trewin, 2010). An early major effort by CRU in the 1980s identified, and attempted to correct where possible, significant inhomogeneities by inspection of data series and, particularly, by comparison with multiple neighbours. The results were fully documented in Jones *et al*. (1985, 1986). One conclusion from this exercise was that the large-scale (hemispheric and global series) were little affected by the application of the adjustments to remove inhomogeneities, partly because positive and negative adjustments tend to cancel each other out (Figure 4 of Brohan et al., 2006). The adjustments did make improvements to the temperature series for individual grid boxes, but no further inhomogeneity adjustments were applied by CRU following those reported in the 1980s. Instead, we recommended (e.g. Jones and Moberg, 2003) that homogeneity assessments, and the development of adjusted series, should instead be undertaken by NMSs because they would in most cases have access to additional meta-data and additional measurement series that would allow more accurate results to be achieved. WMO has a number of documents detailing the need for homogeneity adjustments (Aguilar *et al.*, 2003 and WMO, 2011).

Following on from our recommendation, we have replaced some of our data series (including some that we had adjusted in the mid-1980s) with the results of a number of international or national NMS-led homogeneity projects (Table 1). The number of CRU-adjusted series in the mid-1980s was 312. With all the additions for this analysis there are now 219 in CRUTEM4. This reduction has come about for the following reasons: 68 series have been replaced with newer series, 15 did not have 1961-90 normals so are not used, and 10 have been removed. This does not mean that there are fewer adjusted series within the database, just that the adjustments have been undertaken by NMSs. The incorporation of the USHCN dataset and the replacement of the co-located series means we already had reduced the number of contiguous US station data in CRUTEM4 that were adjusted by CRU in the 1980s.

For the 219 series that were identified by CRU as in need of adjustment (and which have not subsequently been replaced by alternative data series), we re-visited the neighbour comparisons reported in the mid-1980s (Jones *et al*., 1985, 1986). These comparisons showed that the adjustments for stations in the Southern Hemisphere outside of Africa reported in (Jones *et al*., 1986) had not actually been applied to the station data used in CRUTEM3. These adjustments have now been applied to the station data used in CRUTEM4. For Southern Hemisphere stations in Africa, the comparisons showed that the adjustments had been correctly applied, though the period of adjustment had been reported incorrectly in Jones *et al*. (1986). The adjustment was correctly applied to the data prior to the inhomogeneity, though it was reported that the data after the inhomogeneity were adjusted. The availability of the station series is discussed in a later section, and will allow further inspection of these neighbour comparisons and a comparison between the CRUTEM3 and CRUTEM4 station data.

Updating the series

In the course of all the above work, it became apparent that a number of the WMO Station Identifiers had been changed by the NMSs. Using the latest list of these identifiers (http://www.wmo.int/pages/prog/www/ois/volume-a/vola-home.htm), some of the CRUTEM3 identifiers were changed to the updated numbers. Changes to identifiers seem to be made by some NMSs to indicate that the station is no longer a manned location but has been replaced by an automatic weather station (AWS), but this is not always the case. The WMO list of station identifiers (referred to as Volume

A) is updated at the beginning of every year. It is possible to monitor this and to also flag up any 'new' WMO identifiers that appear in the monthly CLIMAT updates at the end of each year. Updating is much easier using current WMO station identifiers.

CRUTEM4 will continue to be updated in near-real time from CLIMAT and MCDW sources. These sources provide data for far fewer than the total number of series now in CRUTEM4, which is over 5500 (including the additional 892 from USHCN). Updating will, therefore, lead to a significant drop in stations beyond 2010 (between them the two sources have a maximum of about 3000 stations) if only these sources are used. All the series discussed above should be updated on web sites, but with different schedules. We intend to periodically check the web sites and update the series every two years for those that are not updated in a more routine fashion. For Australia and Greenland it is likely these can be incorporated at the same stage as MCDW (i.e. 3-4 months behind the real-time update using CLIMAT). USHCN data will be updated at the end of each year. For the other regions/countries discussed updating should be possible annually or every 2-3 years. There are GCOS initiatives to request more countries to release many more of their national series over the CLIMAT system. Several European countries (e.g. Germany and Spain) have already begun to do this. In terms of the global average it would make most difference if Russia and Canada also did this as their areas are large.

Availability of the station series

Given the importance of the CRUTEM land temperature analysis for monitoring climate change (e.g. Trenberth *et al*. 2007), our preference is that the underlying station data, and software to produce the gridded data, be made openly available. This will enhance transparency, and also allow more rapid identification of possible errors or improvements that might be necessary (see e.g. the earlier discussion of homogeneity adjustments in the SH).

Nevertheless, we are reliant on obtaining some data from NMSs and must be careful not to jeopardise our continued access to these data. Apart from data obtained from public sources, some data in our database was obtained without a clear indication of our freedom to make it openly available or perhaps with informal agreements not to do so. In November 2009, the UK Met Office wrote on our behalf to all NMSs to determine if we could release the versions of their monthly temperature series that we held. Of the about 180 letters, we received 62 positive replies, 5 negative replies and the remainder did not reply. For some of the positive replies conditions were imposed, basically of two kinds: (1) please point users to the NMS web site where they might gain access to more or improved station series, or (2) permission to release some but not all the series.

Not content to withhold data for those countries for which we had either no reply or a negative reply from their NMS, we have compared station locations and data with those available in GHCNv3. Where the locations and most of the data agreed, we deemed that we could release these data because they were already available through GHCNv3. WMO Resolution 40 (http://www.wmo.int/pages/about/Resolution40_en.html) requires that all monthly-mean temperature data "necessary to provide a good representation of climate" should be freely available, though the extent to which this is enforced in cases where NMSs do not make this data available is unclear. Furthermore, this is an agreement signed by the NMSs and WMO and not with other third parties. Data from the WMO's RBCN (Regional Baseline Climatological Network) should be freely available however they have been obtained. Additionally, data from CLIMAT, MCDW and WWR are freely available, just in different formats.

As a result of these efforts, we are able to make the station data for all the series in the CRUTEM4 network freely available, together with software to produce the gridded data (http://www.cru.uea.ac.uk/cru/data/temperature/ and http://www.metoffice.gov.uk/hadobs/). Note that in many cases these station data have been adjusted for homogeneity by NMSs; in order to gain access to the original raw (i.e. as measured data or daily and sub-daily measurements) it will be necessary to contact each NMS.

3. Transformation of the station data to a regular grid

All analyses of large-scale temperatures require strategies to reduce the biases in (e.g.) hemispheric averages and principal component patterns that would arise from uneven station density (i.e. biased to regions where station density is high) or from temporal variations in data coverage (e.g. a reduction in data from regions with cooler average temperature). These strategies typically include the representation of temperature anomalies on a regular grid (Peterson *et al.*, 1998): the most widely used method is termed the climate anomaly method (CAM, e.g. Jones, 1994), with the other two being the reference station method (RSM, Hansen *et al.*, 2010) and the first difference method (FDM, Peterson *et al.*, 1998).

Direct comparisons of the three approaches with the same basic data were discussed by Peterson *et al*. (1998) and Vose *et al*. (2005). Possible differences between the techniques and advantages/disadvantages of each are also discussed by Jones *et al*. (1999). In this study we use the CAM approach, which requires reducing all the station temperature data to anomalies, from a common period such as 1961-90 on a monthly basis. Grid-box anomaly values were then produced by simple unweighted averaging of the individual station anomaly values within each grid box.

The main disadvantage of CAM is that stations must have enough years with data within the 1961-90 period in order to be used. For some stations with incomplete data for 1961-90 it will be possible to use published 1961-90 normals (WMO, 1996), although care is required when doing this.

3.1 Development of 1961-90 normals and outlier checks

Monthly averages for 1961-90 (the latest WMO normal period) were calculated from the enhanced station dataset, accepting an average if at least 14 years of data are available. For stations where this was not possible, WMO (1996) normals were used, if available, for all months. For a further set of stations, 1961-90 normals were estimated using the 1951-70 period and adjusted by the difference between the grid-box averages for 1961–90 and 1951–70 from the earlier gridded data (see discussion in Jones and Moberg, 2003). Altogether 1961-90 normals were developed for 4842 stations, of which 4625 were calculated directly, 151 from WMO (1996), and 66 using 1951-70 averages. Temperature data for the remaining 741 stations without

1961-90 normals were not used in the subsequent gridding. In terms of station years (where a year with at least nine valid months counts as a year) over the 1850-2010 period, the amount of data not used totals only 4% of the overall station-year total.

The choice of 1961-90 rather than a later 30-year period (e.g. 1971-2000 or 1981-2010) ensures that as much data as possible are used. There would be a much greater amount of unused data if a more recent 30-year period were used. The period 1961-90 also ensures consistency with earlier analyses. Differences in base periods can also confuse users, especially the media (see Arguez and Vose, 2011).

Section 2 has extensively discussed the sources of the additional temperature data. Although many of the sources have undergone detailed homogeneity testing there is still the possibility of outliers, which might induce a longer-lived influence if they occur during the 1961-90 period. To assess outliers we have also calculated monthly standard deviations for all stations with at least 15 years of data during the 1941-90 period. If a station does not have standard deviation values then the station is not used in the subsequent gridding. This removes an additional 59 series, but all of these are relatively short in duration. All outliers in excess of five standard deviations from the 1961-90 mean were compared with neighbours and corrected or set to the missing code. After this step the 1961-90 normals and the 1941-90 standard deviations were recalculated. In the subsequent gridding (next section) outliers in excess of 5 standard deviations are omitted. As there are no outliers for 1961-90 values this step only applies to years before 1961 and after 1990.

3.2 Gridding and number of stations used through time

Each of the 4842 stations with normals were first associated with their 5º by 5º latitude/longitude grid box and grid-box anomaly values calculated by simple averaging of all available station anomaly values within each grid box for all months 1850-2010. All station outliers in excess of five standard deviations were omitted from the analysis. Apart from retaining the grid-box temperature values, we also retain the number of stations per grid box. This latter value will be necessary to calculate a 'variance-adjusted' version of the gridded dataset following the approach outlined in Jones et al. (2001), Jones and Moberg (2003) and Brohan *et al*. (2006). The approach used adjusts the variance of each grid-box time series to be compatible with the infinitely-sampled grid box (see Jones and Moberg, 2003 and Brohan *et al*. 2006). This version of the dataset is referred to as CRUTEM4v with the unadjusted version as CRUTEM4. CRUTEM4v reduces the impact on each grid-box time series of changing station availability through time. CRUTEM4v is recommended for use for small regions and individual grid-box time series, especially if users wish to consider changes in variance and/or extremes (at the monthly timescale – see for example Figure 5 of Jones *et al*. 1999). At the hemispheric scales, that will be discussed in the next section, there is very little difference between averages calculated with CRUTEM4 and CRUTEM4v. Brohan *et al*. (2006) additionally discusses reasons for appropriate usage of these two versions of the dataset. In the subsequent analyses in the next section we will use CRUTEM4 to calculate all the hemispheric and any regional series used.

Before moving to the next section, we first explain changes in the number of stations available through time. Figure 1 illustrates both the number of stations used each year and the % area coverage this produces for each hemisphere. The results are compared with the earlier analysis using CRUTEM3 in Brohan *et al*. (2006). The improvement in the station numbers is more dramatic for the NH compared to the SH. The big increase in 1895 represents the starting date for many of the stations in the contiguous United States, while there is a similar increase in station numbers in 1951 when the first of the 10-year WWR volumes became available. Numbers of stations reduce from a peak in the 1960s, occurring in a series of steps at the end of each decade indicative of the cause being changes in station availability in the WWR volumes. For the SH, there are few improvements in coverage, the main ones being due to improved use of the 1971-80 WWR volumes and the inclusion of more data after 2000.

In terms of percentage area coverage, the improvements have had a smaller effect than in terms of station numbers, with the increase being greater in the NH compared to the SH. The small step changes at the end of each decade (1980, 1990 and 2000) are due to the WWR volumes. There are generally enough contiguous US data for 2010 from the CLIMAT and MCDW sources, so the unavailability of USHCN data in 2010 does not affect the area coverage for the NH and the drop in the final year for the NH is principally due to missing Russian data. There is little change in area coverage for the SH in CRUTEM4 compared to CRUTEM3.

4. Analysis of the enhanced gridded land data

4.1 Hemispheric-scale averages and comparisons with CRUTEM3

Hemispheric average time series were produced using cosine weighting of grid-box values in each hemisphere (Jones, 1994). Averages were calculated for each month from January 1850 and then seasonal and annual averages calculated using the hemispheric-average monthly values. Standard three-month climatological seasons were used, with December of the previous counting towards the winter value for the current year. As December 1849 is not available all the seasonal and annual series for the Northern Hemisphere (NH) begin in 1851. For the Southern Hemisphere (SH), the first year is taken as 1856. Before this date there are less than 5 stations with data. Beginning with 1856 the number of available stations in the SH increases to 5 series, reaching 10 by 1860 (see Figure 1). In later figures (Figures 5b and 9b) we will highlight that uncertainty ranges of SH averages, calculated from so few stations, are substantial.

Figure 2 shows the seasonal and annual values for the NH and SH and an annual series for the global land together with 10-year Gaussian smoothed series. For comparison the smoothed CRUTEM3 series are also shown to see the impact of the additional or replaced series in the station database. The global land series is computed by weighting the two hemispheres approximately in proportion to the areas of their land masses (i.e. Global = [(2/3)NH + (1/3)SH]. This weighting is different to the equal hemispheric weighting applied by Brohan *et al*. (2006) for CRUTEM3. The new weighting has also been applied here to CRUTEM3.

The differences between the two sets of smoothed lines indicate excellent agreement from 1880 up to 2000 for the NH. At this decadal timescale, all the additions have made no discernible differences between the analyses, an initial indicator that for hemispheric-scale averages the analysis is very robust. CRUTEM4 is very slightly warmer since 2000 for the NH for the year and all seasons except summer. The likely reason for this is the additional data in the Arctic (particularly Russia) and this will be further investigated in the next section. Prior to 1880, CRUTEM4 is slightly cooler than CRUTEM3, more so in winter (DJF) and spring (MAM) than the other seasons. Again later analyses will be suggestive that this results from the additional Russian series. For the SH, differences between CRUTEM4 and CRUTEM3 are slightly greater earlier in the series and extend up to the early 20th century, particularly in the austral winter (JJA). CRUTEM4 is cooler than CRUTEM3 during 1861-1910, the exact period depending on the season. CRUTEM4 is very slightly warmer than CRUTEM3 since about 2005. Possible reasons for the differences in the 19th century in the SH will be investigated in the next section. Uncertainty ranges, calculated using the same approach as Brohan *et al*. (2006) will be shown in later figures (on the decadal scale in Figure 5 and on the interannual timescale in Figure 9).

For the NH, year-to-year variability is greatest during winter and least in summer. The slightly greater variability prior to 1880 in all seasons (except summer) is more likely to be due to sparser coverage then a real feature. This greater variability is marginally reduced by adjusting the individual grid-box time series for changing station data contribution (introduced in Jones *et al*., 2001 and the dataset produced here called CRUTEM4v) but the variance of regional averages has not been similarly adjusted for reduced grid-box availability. For the SH, year-to-year variability is more similar between the seasons.

All seasons and the annual series for both hemispheres show comparable century-scale warming from the beginning of the 20th century but there are differences in timing between them. Warming is significant in all seasons and annually for 1861-2010, 1901-2010 and 1979-2010 (except for May and December for the SH for 1979-2010). Table 2 provides the warming explained by a least squares linear fit to the monthly series for these three periods. Warming in all three periods tends to be greater in the NH compared to the SH, and the NH warming has a much more marked seasonal character than that for the SH. Table 2 also includes calendar year average values for CRUTEM3. CRUTEM4 warms more than CRUTEM3 for all three periods due to the cooler values before about 1880 (particularly in the SH) and slightly warmer values in the NH since about 2000.

The marked seasonality of the warming for 1861 to ~1900 (estimated by comparing the NH trend differences in Table 2 for 1861-2010 and 1901-2010) may be artificial due to the possible impacts of direct sunlight on the instruments, prior to the development of Stevenson-type screens, in higher northern latitudes during summer (see earlier discussion in relation to the HISTALP dataset, Böhm *et al*., 2010). The addition of the newly adjusted series in the GAR may be the reason for the slight difference between CRUTEM3 and CRUTEM4 before 1860 when coverage is sparse outside Europe. Böhm *et al*. (2010) and Brunet *et al*. (2011) are suggestive of this issue being much wider in scale across the mid and high latitudes of the NH. Alternatively, if this seasonal contrast is real, then it implies a marked change in continentality (greater winter/summer temperature differences) over part of the NH prior to 1880. Further work is required, but the studies reported above are clearly suggestive of screen exposures being the more likely cause.

4.1.1 Spatial comparisons between CRUTEM3 and CRUTEM4

In this section we compare spatial patterns between CRUTEM4 and the earlier CRUTEM3 dataset. In Figure 3, we plot the annual temperature anomaly for the decade 2001-2010, with respect to our base period of 1961-90, for both analyses and their difference. This difference clearly illustrates the improvement (i.e. outlined in black in panel (c)) in coverage in CRUTEM4 compared to CRUTEM3, particularly across the higher latitudes of Eurasia and North America. As this expansion of spatial coverage in the Northern Hemisphere has contributed to warmer temperatures in CRUTEM4, the 2001-2010 decade is warmer than CRUTEM3 for the NH (0.80°C compared to 0.73°C). There is much less coverage change across the Southern Hemisphere and the two corresponding averages are 0.43°C for CRUTEM4 and 0.40°C for CRUTEM3. Panel c of Figure 3 is mostly green, but differences do occur, particularly over the contiguous United States and Australia, where we have made many changes to the station data used (see discussion in Section 2).

In Figure 4, we show linear trend maps for annual temperature averages for 1951-2010 for both analyses and the difference. The panel (b) for CRUTEM4 shows the improvements in coverage, which can also be seen in panel (c) by the grid boxes outlined in black. Of the grid boxes in common between the two analyses, 499 boxes differ within ± 0.2°C in their total trends over the 60-year period, with 86 boxes indicating that the CRUTEM4 trend was > 0.2°C more than CRUTEM3 and 41 with CRUTEM3 having > 0.2°C more warming than CRUTEM4.

4.2 Assessment of the robustness of hemispheric averages omitting large numbers of stations

In the previous section, we illustrated the robustness of the large-scale averages by comparing this new version of the dataset (CRUTEM4) with the previous (CRUTEM3). Differences are relatively minor and well within the error ranges estimated by the earlier Brohan *et al*. (2006) study and re-calculated here. In this section we expand on this, by using considerably less station data while still producing essentially the same hemispheric series at the decadal time scale. We do this by using mutually exclusive subsets of the overall station data and secondly omitting all the station data from some large countries.

4.2.1 Using only a subset of the station data

For this exercise we took the 5583 stations and separated them into five subsets each containing a unique 20% of the data. The ordering of the stations in the station file uses the World Meteorological Organization (WMO) numbering system, with the exception of the 892 USHCN stations, which have been placed at the end. The first subset contained stations ordered 1, 6, 11, 16… etc in the list. The second contained stations ordered 2, 7, 12, 17…etc, with the fifth set containing the stations ordered 5, 10, 15, 20… In this separation into five subsets, no account was taken of whether the station had sufficient data for the 1961-90 reference period. Therefore, after removal of those station records with insufficient data during the 1961-90 reference period, the size of each subset may differ slightly. It will also differ back in time, since record length is also not considered when forming the subsets. For each subset the 20% of the data were gridded using the same method as described in section 3.2 and hemispheric seasonal and annual averages calculated as stated in section 4.1. Figure 5 shows the hemispheric averages from the five networks, by season and year, together with that of the complete CRUTEM4 network (i.e. 100%). Differences between the five networks are barely noticeable after the 19th century for the NH. For the SH there are larger differences, but for both hemispheres they are well within the error ranges calculated by Brohan *et al*. (2006) approach. For the 19th century, differences are only marked in the Southern Hemisphere, where coverage is poorer than in other parts of the world.

The results shown in Figure 5 are not unexpected. A similar assessment of this kind was undertaken by Parker *et al*. (2009) using two networks of offset and non-adjacent 5° by 5° latitude/longitude grid boxes. The differences in the 19th century in the SH for Parker *et al*. (2009) were larger, but that was due to an even smaller set of stations (and hence grid boxes) being used. The simple reason that a small network of well-located sites can closely reproduce the series derived from a much greater station network is due to there being a limited number of independent spatial degrees of freedom (see Jones *et al*., 1997, where this concept was explored in considerable detail). That paper concluded that hemispheric and global average temperatures (at annual timescales and above) could be reliably estimated (i.e. within the error ranges shown in Figure 5) by as few as 50-100 sites. Reliably here means within the error range estimated by Brohan *et al*. (2006). The greater differences during the 19th century, especially for the SH, arise because the station network is so limited then, that separating it into five subsets results in each subset having insufficient stations to obtain a reliable SH temperature estimate. This point is discussed more in section 4.3.

There are a number of obvious asides that can be made once the concept is realised. For example, if resources became available for digitization of early temperature data then these would be best targetted at the data sparse regions, particularly in the Southern Hemisphere and the tropics. These issues are discussed further in Jones and Wigley (2010).

4.2.2 Omitting large countries

Another possible concern is that the CRUTEM4 station database might be unduly dominated by data from particular countries or regions. Gridding the data overcomes this to a large extent but the robustness of the CRUTEM4 data to this issue can additionally be assessed by considering the effect of removing series from different countries of the world. In the first part of this exercise we took the 5583 stations and separately removed all stations in the contiguous United States and Australia. Figure 6 shows the NH seasonal and annual averages based on all stations compared to averages omitting sites from the contiguous United States. The effect here is only noticeable in the 19th century and then mostly only in winter (DJF) and spring (MAM). In these seasons, and to some extent in the annual mean, omitting the contiguous US data lowers the earliest temperature estimates, implying that the mean US temperature anomalies are slightly warmer than the mean for the rest of the NH. Figure 7 shows similar plots omitting all Australian stations. This is a much more severe test than in Figure 6, as Australia is a much larger component of the SH landmass than the contiguous USA is of the NH. Removing Australian stations has a larger effect, particularly prior to 1900, but as with Figure 5, if error ranges were plotted these would easily encompass the differences seen. The sign of the difference arising from the removal of Australian temperatures varies between seasons and with time, indicating no systematic difference with the mean of the rest of the SH. In the annual mean, removing Australian data warms the SH mean around 1860 and in the 1940s, but cools it during the 1880s.

Although both Australia and the contiguous United States are very large areas, we now go a stage further and omit two larger regions: first Russia and second the former Soviet Union (fUSSR). The results are shown in Figures 7 and 8. As expected the effects of removing fUSSR are slightly more apparent than when removing just Russia, though the periods of the differences tend to be similar (as Russia was a large component in terms of area of the fUSSR). Removal of either tends to make the NH slightly warmer in the 19th century, particularly in the winter (DJF) and spring (MAM) seasons. As we have added large numbers of extra stations in both Russia and the Arctic (particularly the Russian Arctic) this is probably the principal reason for the slightly cooler NH temperatures during the 19th century and to a lesser extent the slightly warmer temperatures in the last ten years in CRUTEM4 compared to CRUTEM3. The similarity of the seasonal differences between Figure 1 and Figures 7 and 8 is very suggestive of this being the most likely cause. Additional data in other parts of the world (principally Europe in the 19th century) are also probably factors. The negligible effects of omitting large regions (and consequently large numbers of stations) are a direct result of the remaining stations still being adequate for monitoring hemispheric averages by sampling the most important spatial degrees of freedom, across the world’s land areas.

There are also issues with the exposure of early instrumental data prior to about 1910 over parts of Australia (Nicholls *et al*., 1996). It is important that resources be found to objectively estimate the necessary adjustments, so that pre-1910 data can be used with more confidence. Biases due to different exposures of early thermometers are also important in Europe, particularly for the period before 1870 (Böhm *et al*., 2010). Issues with the different exposure properties (from pre-louvred-screen locations) are only beginning to be incorporated into global temperature databases. Traditional approaches to station homogenization are unable to detect the problem as all sites within a region are likely similarly affected by the same problem (see discussion in Jones and Wigley, 2010). In this study we have included 107 series from the GAR that have been adjusted to attempt to compensate for changes in exposure, but it is apparent that stations in other mid and high latitude regions probably need adjustment during the summer months (typically to cool the earliest temperature estimates relative to the modern data). For the NH, the effect principally occurs for the period before about 1880, so the regions of the world where additional assessment is needed is Europe, Russia and Iceland/Greenland. Canada and Alaska are also likely to be affected, but there are few stations beginning before 1880. Assessment will be difficult as all series are likely to be similarly affected. Approaches such as the rebuilding of the screens from the 19th century (e.g. Brunet *et a*l., 2010) and taking parallel measurements is a possible avenue to follow.

4.3 Comparison of annual hemispheric series with the results of analyses by other groups

In this section the two hemispheric land-only averages are compared with two other analyses: series developed by the National Climatic Data Center (NCDC, Smith and Reynolds, 2008) and the Goddard Institute for Space Studies (GISS, Hansen *et al*., 2010). Our present study uses a base period of 1961-90 while NCDC currently uses 1901-2000 and GISS 1951-80 for their published series. For direct comparison we have adjusted both series to our 1961-90 base period on a monthly basis. Figure 9 shows hemispheric seasonal and annual series from CRUTEM4, additionally plotting decadally-smoothed series for the two US analyses. For both the NH and SH, CRUTEM4 tends to more closely follow NCDC than GISS, even though all three show similar amounts of long-term warming since 1880. The reason why CRUTEM4 more closely follows NCDC has been discussed before (Vose *et al*. 2005) and relates to these two analyses using the same 5° by 5° latitude/longitude grid boxes compared to the 40 equal area boxes used per hemisphere by GISS. Correlations between CRUTEM4 and NCDC/GISS are 0.984/0.980 for the NH and 0.950/0.927 for the SH (for the 1880 to 2010 period) and support the findings of Vose *et al*. (2005). Differences between the three analyses are greater in the SH compared to the NH, particularly before about 1920. Differences are not sustained right back to the start of records, however, as the lines move closer together again in the 1880s. The uncertainty ranges for the SH are larger than the NH due to more missing boxes (particularly over the Antarctic) and fewer stations per grid box over Africa and South America than the northern continents.

4.4 Comparisons with ERA-Interim and ERA-40 Reanalyses

In this section we compare CRUTEM4 at the hemispheric resolution with similar land averages calculated from two versions of the European Centre for Medium-Range Weather Forecasting Reanalyses (ECMWF) Reanalyses (ERA-40 and ERA-Interim). ERA-40 covers the period 1958-2001 and ERA-Interim (which uses 4D variational assimilation compared to the 3D schemes in ERA-40) the period from 1979 to 2010. For a discussion of the ECMWF Reanalyses see Simmons *et al*. (2004, 2010) and Uppala *et al*. (2005). A common period for both Reanalyses is 1981-2000 so we reduce their absolute land temperature values to anomalies from this base period. Figure 10a shows seasonal and annual comparisons between the two Reanalyses and CRUTEM4. As with the earlier plots we show seasonal and annual values of CRUTEM4 (from the 1961-90 base period) with the two ECMWF Reanalyses as smoothed series using a 10-year Gaussian smoother. For the NH, both ERA-40 and ERA-Interim track one another very well over their period of overlap (1979-2001) and are offset from CRUTEM4 by an amount that relates to the difference between the 1961-90 and 1981-2000 periods. In Figure 11 we compare ERA-Interim with CRUTEM4 for the Northern Hemisphere on the monthly timescale from 1979. For this plot, the base period of 1979-2010 is used for both series. The agreement between the two series is excellent. ERA-Interim warms slightly more than CRUTEM4 over this period, which is probably due to greater warming in the Arctic land grid boxes in ERA-Interim that are missing in CRUTEM4.

For the SH in Figure 10b, there are marked differences between both Reanalyses during their overlap period. ERA-Interim is closer to CRUTEM4 but the similarity of the smooth curves is markedly less good particularly in the austral autumn (MAM) and winter (JJA). ERA-40 is further offset from CRUTEM4 before about 1980 in all season except austral summer, and this is due to a cold bias in the climate model used by both Reanalayses over the Antarctic (Uppala *et al*., 2005). To illustrate this further, we have calculated averages for both the SH 0-60°S and for the Antarctic (60-90°S) for all three series (Figure 12). For ERA-Interim, the time series agreement (for the SH 0-60°S) is almost as good as the NH land but for ERA-40, there is a significant divergence before the early 1970s with warmer ERA-40 temperatures in all seasons. This difference was commented upon by Simmons *et al*. (2004) and was shown to be due to ERA-40 being given little input data for Australia prior to the early 1970s. With little input data to correct model biases, the Reanalyses tends to the model simulation which for Australia is a model that is biased warm (see further discussion in Simmons *et al*., 2004 and Uppala *et al*., 2005). For the Antarctic, the cold bias in the climate model used by ERA-40 is clearly evident, particularly so in all seasons, although it is smaller in the austral summer (DJF). Figure 13 repeats Figure 11 but for the SH 0-60°S showing good agreement between CRUTEM4 and ERA-Interim, but this is less good than the NH for the 12-month Gaussian smoothed lines.

5. Conclusions

In this paper we have detailed the developments to the CRUTEM4 dataset available from the Climatic Research Unit. The improvements to the quality of the grid-box dataset have been made possible by better availability of the basic station data. The homogeneity of the station data has been improved by investments of effort by a number of research groups and particularly by a number of NMSs around the world. We undertook much homogeneity work in the 1980s, but recommended at that time that this work be best undertaken by NMSs. This is beginning to come to fruition and we hope that more can find the resources to complete this task. In the 1980s, we adjusted 312 station series (then about 10% of the overall total of stations). Replacement of many of these series by improvements from NMSs means that there are only 219 stations (4.6% of the new total of stations with normals) that we adjusted almost thirty years ago. The major bias issue that still affects the dataset relates to exposure of the thermometers before louvred screens were introduced between 1870 and 1880. Three studies (Böhm *et al*., 2010, Brunet *et al*., 2011 and Nicholls *et al*., 1996) have considered the problem (summer temperatures are probably biased warm by up to 0.5°C) and provided adjusted data in the case of the Greater Alpine Region, which we have used. We urge more studies of these kinds to be undertaken using the parallel measurement approach developed by Brunet *et al*. (2011).

Differences in the hemispheric averages produced by the new version (CRUTEM4) compared to the earlier (CRUTEM3) are relatively small and well within the error ranges developed using the techniques described in Brohan *et al*. (2006). This result is not unexpected and confirms a number of other studies by the groups producing these datasets. To illustrate this robustness further we carried out two sets of analyses, focussing on the hemispheric-scale averages that result. Firstly, we separated the station data into five independent samples each comprising 20% of the basic station series. Secondly, we separately omitted all the station series from large countries (contiguous United States, Australia, Russia and the former Soviet Union). For both sets there were differences between the analyses, but they were barely visible on time-series plots after 1900 for the Northern Hemisphere (NH) and after about 1920 for the Southern Hemisphere (SH), so effects are only for periods where coverage becomes markedly sparse. Even then, differences were well within the range of the error estimates we have developed in an earlier study (Brohan *et al*., 2006).

Finally, we compared the hemispheric averages with estimates derived from Reanalysis products (ERA-40 and ERA-Interim) developed by the European Centre for Medium-Range Weather Forecasts. ERA-40 covers the period 1958-2001 and ERA-Interim 1979-2010. For the NH, the agreement between the two Reanalyses and CRUTEM4 was excellent. For the SH, agreement was considerably poorer, but if the SH was restricted to 0-60°S then it was markedly improved. Problems with Reanalyses over the Antarctic are well known, though ERA-Interim is a considerable improvement over ERA-40 for the Antarctic region.

Acknowledgements

The authors thank all the scientists and NMSs who have collected and made available the monthly-mean temperature data referred to in Table 1. The authors thank David Parker and Adrian Simmons for comments on an earlier draft. This work has been supported by the USDoE (Grant DE-SC0005689), JISC (ACRID project) and by the University of East Anglia. CM was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

References

Aguilar, E., I. Auer, M. Brunet, T.C. Peterson and J. Wieringa (2003) Guidelines on Climate Data and Homogenization. WCDMP-53, World Meteorological Organization, WMO/TD No. 1186, Geneva, Switzerland, 52pp (http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP-53.pdf)

Arguez, A. and Vose, R.S., 2011: The definition of the standard WMO Climate Normal: The key to deriving alternative climate normals. *Bull. Amer. Met. Soc*. 93, 699-704, doi:10.1175/2010BAMS2955.1.

Auer, A., R. Böhm, and W. Schöner (2001), Austrian long-term climate 1767–2000: Multiple instrumental climate time series from Central Europe. Publ. 397, Österreichische Beiträge zu Meteorogie und Geophysik Heft 25, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria, 147 pp.

Bekryaev R., I. Polyakov, and V. Alexeev, (2010) Role of polar amplification in long-term surface air temperature variations and modern Arctic warming, *J Climate*, 23,3888–3906

Böhm, R., P.D. Jones, J. Hiebl, D. Frank, M. Brunetti, and M. Maugeri (2010) The early instrumental warm-bias: a solution for long Central European temperature series, 1760-2007, *Climatic Change*, 101, 41-67.

Brohan, P., J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones (2006) Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, *J. Geophys. Res.*, 111, D12106, doi:10.1029/2005JD006548.

Brunet, M., J. Asin, J. Sigró, M. Bañon, F. García, E. Aguilar, J.E. Palenzuela, T.C. Peterson, and P.D. Jones (2011) The minimization of the *screen bias* from ancient Western Mediterranean air temperature records: an exploratory statistical analysis, *Int. J. Climatol*. 31, 1879-1885.

Bureau of Meteorology (2000) *Climate Data: Australia, Version 1.01d,* CD-ROM. Climate Services, Bureau of Meteorology, Kent Town, Australia.

Camuffo, D. and P.D. Jones (Editors) 2002: *Improved Understanding of Past Climatic Variability from Early Daily European Instrumental Sources,* Kluwer Academic Publishers, Dordrecht, 392pp.

Cappeln, J. (2010) DMI Monthly Climate Data Collection 1768-2009, Denmark, The Faroe Islands and Greenland, Technical Report No. 11-05, Danish Meteorological Institute, Copenhagen, 52pp.

Cappeln, J. (2011) DMI Monthly Climate Data Collection 1768-2010, Denmark, the Faroe Islands and Greenland, Technical Report No. 11-05, Danish Meteorological Institute, Copenhagen, 54pp.

Frich, P. and Coauthors (1996) North Atlantic Climatological Dataset (NACD Version 1). Scientific Rep. 96-1, Danish Meteorological Institute, Copenhagen, Denmark, 47 pp. plus appendixes and data files.

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

Jones, P.D. (1994) Hemispheric surface air temperature variations: a reanalysis and an update to 1993. *J. Climate* 7, 1794-1802.

Jones. P.D. and D. Lister, (2004) The development of monthly temperature series for Scotland and Northern Ireland, *Int. J. Climatol*., 24, 569-590.

Jones, P.D. and T.M.L. Wigley (2010) Estimation of global temperature trends: What's important and what isn't, *Climatic Change*, 100, 59-69.

Jones, P.D., S.C.B. Raper, B.D. Santer, B.S.G. Cherry, C.M. Goodess, P.M. Kelly, T.M.L. Wigley, R.S. Bradley and H.F. Diaz (1985) A grid point surface air temperature data set for the Northern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TR022*, 251pp (Available at www.cru.uea.ac.uk/st/TR022.pdf).

Jones, P.D., S.C.B. Raper, B.S.G. Cherry, C.M. Goodess and T.M.L. Wigley (1986) grid point surface air temperature data set for the Southern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, *Technical Report TR027*, 73pp (Available at www.cru.uea.ac.uk/st/TR027.pdf).

Jones, P.D., T.J. Osborn and K.R.Briffa (1997) Estimating sampling errors in large-scale temperature averages, *J. Climate*, 10, 2548-2568.

Jones, P.D., M. New, D.E. Parker, S. Martin, and I.G. Rigor (1999) Surface air temperature and its changes over the past 150 years, *Reviews of Geophysics*, 37, 173-199.

Jones, P.D., T.J. Osborn, K.R. Briffa, C.K. Folland, B. Horton, L.V. Alexander, D.E. Parker and N. Rayner (2001) Accounting for sampling density in grid-box land and ocean surface temperature time series, J. *Geophys. Res*., 106, 3371-3380.

Menne M.J., C.N. Williams Jr., and R.S. Vose (2009), The United States Historical Climatology Network Monthly Temperature Data – Version 2, *Bull. Amer. Meteor. Soc.*, 90, 993-1007.

Nicholls N, R. Tapp, K. Burrows, and D. Richards (1996) Historical thermometer exposures in Australia, *International Journal of Climatology*, 16, 705-710.

Parker, D. E. (2004), Large-scale warming is not urban, *Nature*, 432, 290–290, doi:10.1038/432290a.

Parker, D. E. (2006), A demonstration that large-scale warming is not urban, *J. Clim.*, 19, 2882–2895, doi:10.1175/JCLI3730.1.

Parker, D. E. (2010), Urban heat island effects on estimates of observed climate change, *Wiley Interdisciplinary Reviews: Climate Change*, 1, 123–133, doi:10.1002/wcc.21.

Parker D. E., P. D. Jones, T. C. Peterson, J. Kennedy (2009) Comment on "Unresolved issues with the assessment of multidecadal global land surface temperature trends" by Roger A. Pielke Sr. et al., *J. Geophys. Res*., 114, D05104, doi:10.1029/2008JD010450.

Peterson, T.C. and R.S. Vose (1997) An overview of the Global Historical Climatology Network Temperature Database, *Bull. Am. Met. Soc*., 78, 2837-2848.

Peterson, T.C., T.R. Karl, P.F. Jamason, R. Knight and D.R. Easterling (1998) The first difference method: Maximizing station density for the calculation of long-term global temperature change, *J. Geophys. Res*., 103, 25967-25974.

Razuvaev, V and O. Bulygina (2009) In Regional Aspects of Climate-Terrestrial-Hydrologic Interactions in Non-boreal Eastern Europe *NATO Science for Peace and Security Series C: Environmental Security* (edited by P. Ya. Groisman and S. Ivanov) 2009, 1, 17-22, DOI: 10.1007/978-90-481-2283-7_3. Book is ISBN 978-90-481-2241-7

Simmons, A.J., P.D. Jones, V. da Costa Bechtold, A.C.M. Beljaars, P.W. Kållberg, S. Saarinen, S.M. Uppala, P. Viterbo and N. Wedi (2004) Comparison of trends and low-frequency variability in CRU, ERA-40 and NCEP/NCAR analyses of surface air temperature, *J. Geophys. Res*., 109, D24115, doi:10.1029/2004JD006306.

Simmons, A.J., K.M. Willett, P.D. Jones, P.W. Thorne, and D. Dee (2010) Low-frequency variations in surface atmospheric humidity, temperature and precipitation: Inferences from reanalyses and monthly gridded observational datasets, *J. Geophys. Res.*, 115, D01110, doi:10.1029/2009JD012442.

Smith, T. M., R. Reynolds, T.C. Peterson and J. Lawrimore (2008) Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880-2006), *J. Climate*, 21, 2283-2293.

Trenberth, K.E., P.D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden and P. Zhai, 2007: Observations: Surface and Atmospheric Climate Change. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)], pp235-336, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Trewin, B. (2010) Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Clim Change* 1: 490-506. doi: 10.1002/wcc.46

Tuomenvirta, H., A. Drebs, E. Førland, O. E. Tveito, H. Alexandersson, E. Vaarby Laursen, and T. Jonsson (2001) Nordklim data set 1.0—Description and illustrations. KLIMA Rep. 08/01, Norwegian Meteorological Institute, Oslo, Norway, 26 pp. [CDROM available from A. Drebs, Finnish Meteorological Institute, P.O. Box 503, FIN-0010 Helsinki, Finland.]

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc*., 131, 2961–3012, doi:10.1256/qj.04.176

Van den Besselaar, E., A. M. G. Klein Tank, G. van der Schrier and P.D. Jones (2011) Synoptic messages to extend climate data records. *J. Geophys. Res*. (submitted).

Vincent, L.A. and D.W. Gullet (1999) Canadian historical and homogeneous temperature datasets for climate change analyses. *Int. J. Climatol.*, 19, 1375-1388.

Vincent, L.A., X. Zhang, B.R. Bonsal and W.D. Hogg (2002) Homogenization of daily temperatures over Canada. *J. Climate*, 15, 1322-1334.

Vinther, B.M., K.K. Andersen, P.D. Jones, K.R. Briffa, and J. Cappelen, (2006) Extending Greenland temperature records into the late-18th century, *J. Geophys. Res*., 111, D11105, doi:10.1029/2005JD006810.

Vose, R.S., D. Wuertz, T.C. Peterson, and P.D. Jones, (2005) An intercomparison of trends in surface air temperature analyses at the global, hemispheric and grid-box scale, *Geophys. Res. Letts*., 32, L18718, doi:10.1029/200GL023502.

Williams, M.W. and V.G. Konovalov (2008) *Central Asia Temperature and Precipitation Data, 1879-2003,*

WMO (1996) Climatological Normals (CLINO) for the period 1961-1990, WMO/OMM – No. 847, World Meteorological Organization, Geneva, Switzerland, 768pp.

WMO (2011) Guide to Climatological Practices, 3rd Edition, WMO-No. 100, World Meteorological Organization, Geneva, Switzerland, 117pp (http://www.wmo.int/pages/prog/wcp/documents/Guidefulltext.pdf).

Table 1: Source codes and number of stations from each

Code	**Station Count**	**Regions**	**Sources (paper, project acronym or website)**
10	2444	Global	Jones (1994), Jones and Moberg (2003) and Brohan *et al*. (2006); homoge sites assessed in Jones *et al*. (1985, 1986).
30	440	Europe, East Asia, Africa, USA, S. America and Australia	GHCNv2 (adjusted series) http://www.ncdc.noaa.gov/ghcnm/v2.php
31	113	Middle East, E. Asia and N. Africa	GHCNv2 (adjusted series) – added at a different time than Code 30 http://www.ncdc.noaa.gov/ghcnm/v2.php
33	11	North Atlantic/Fennoscandia	NACD project: Frich *et al*. (1996)
34	18	Fennoscandia	NORDKLKIM project: Tuomenvirta *et al*. (2001)
35	1	Long European series	IMPROVE project: Camuffo and Jones (2002)
36	1	Canadian climate series	CHTD: Vincent and Gullett (1999). Replaced by code 42 stations
37	63	Asia, Central and S. America	WWR, data added in 2006, mostly for the 1981-1990 decade
38	60	Mali, DR. Congo plus a few others	Series given to CRU by various academic visitors
40	98	Australia	Homogenized series, Bureau of Meteorology, Australia ftp://ftp.bom.gov.au/anon/home/ncc/www/change/HQdailyT/HQdailyT_in
41	13	New Zealand	Homogenized series , NIWA, New Zealand http://www.niwa.co.nz/our-science/climate/news/all/nz-temp-record
42	207	Canadian (updated version of #36)	AHCCD (Vincent et al., 2002) http://www.ec.gc.ca/dccha-ahccd/
43	372	Russia	RIHMI-WDC: Razuvaev and Bulgina (2009) http://meteo.ru/climate/sp_clim.php
44	1064	Contiguous United States	USHCN : Menne *et al*. (2009) http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html
45	13	United Kingdom	UK Met Office and SNIFFER (Jones and Lister, 2004) http://www/metoffice.gov.uk/climate/uk/stationdata
47	1	Greenland (Qaqortoq)	Vinther *et al*. (2006) and Cappeln (2010, 2011) http://www.dmi.dk/dmi/tr10-05 and http://www.dmi.dk/dmi/tr11-05
48	6	Denmark and Faroe Islands	Cappeln (2010, 2011) - http://www.dmi.dk/dmi/tr10-05 and http://www.dmi.dk/dmi/tr11-05

49	107	Greater Alpine Region (GAR)	HISTALP: Böhm *et al*. (2010) http://www.zamg.ac.at/histalp
50	61	Central Asian stations mostly at high elevation	NSIDC: Williams and Konovalov (2008) http://nsidc.org/data/docs/noaa/g02174_central_asia_data/index.html
51	213	Russia (includes some fUSSR) – daily series	RIHMI -WDC http://meteo.ru/english/climate/d_temp.php
52	12	Swiss climate series combined by HISTALP	http://www.meteosuisse.admin.ch/web/en/research/events/archive/foko_2007_2.Par.0008.DownloadFile.tmp/beggertfoko20072.pdf
53	30	Various NMSs	Data from various NMSs received in 2010
54	235	Arctic Series	IARC: Bekryaev *et al*. (2010) and see data series at: http://people.iarc.uaf.edu/~igor
	5583	**Total**	

Table 2: Total temperature change (°C) for CRUTEM4 described by linear least squares regression lines fitted over three periods: 1861-2010, 1901-2010 and 1979-2010. Comparative annual values for CRUTEM3 are shown at the bottom.

	1861-2010		**1901-2010**		**1979-2010**	
	NH	**SH**	**NH**	**SH**	**NH**	**SH**
Jan.	1.39	0.94	1.12	0.84	1.02	0.39
Feb.	1.48	0.96	1.55	0.77	1.21	0.37
Mar.	1.69	0.92	1.62	0.83	1.40	0.38
Apr.	1.25	0.91	1.33	0.78	1.24	0.35
May	1.06	1.11	1.15	0.90	1.00	0.15
Jun.	0.69	0.92	1.00	0.78	0.99	0.53
Jul.	0.52	1.10	0.84	0.86	1.00	0.58
Aug.	0.67	1.07	0.82	0.95	1.04	0.40
Sep.	0.71	0.89	0.75	0.88	0.98	0.64
Oct.	1.20	0.92	0.89	0.95	1.16	0.65
Nov.	1.54	0.86	1.10	0.81	1.43	0.43
Dec.	1.47	0.76	1.27	0.74	0.81	0.18
Year	1.14	0.94	1.12	0.84	1.11	0.42
CRUTEM3	1.05	0.77	1.06	0.82	1.02	0.39

Figure Captions

Figure 1: Comparison of station counts and percent area coverage (of the entire hemisphere including oceans) for CRUTEM4 (thick) and CRUTEM3 (thin) for the NH and SH.

Figure 2: Seasonal and annual averages by hemisphere for CRUTEM4, with the smoothed lines showing decadal-filtered series for CRUTEM4 (thick) and CRUTEM3 (thin). Hemispheric temperature averages for the land areas are expressed as anomalies (in degrees Celsius from the base period of 1961-90). The decadal smoothing uses a 13-term Gaussian filter, padded at the ends with the mean of the adjacent 6 values. (a) NH, (b) SH and (c) global for the annual average. The global average is calculated as [(2/3)NH + (1/3)SH].

Figure 3: Comparison of annual mean temperature anomalies from (a) CRUTEM3 and (b) CRUTEM4 for the period 2001-2010 (degC anomalies from 1961-90). Grid boxes with less than 50% data coverage (5 years) are left white. (c) Difference (b)-(a) to compare CRUTEM3 and CRUTEM4 means over this period. Grid boxes with insufficient data (<5 years during 2001-2010) in CRUTEM3 but sufficient data in CRUTEM4 are outlined in black; black crosses indicate the reverse situation.

Figure 4: Comparison of linear trends fitted to (a) CRUTEM3 and (b) CRUTEM4 annual temperatures for the period 1951-2010. Trends are expressed as the degC linear trend change over the 60 year period. Grid boxes with less than 80% data coverage (48 years) are left white. Boxes or regions outlined in black are those where the trend slopes are significantly different from zero, with 95% confidence taking into

account first-order autocorrelation. (c) Difference (b)-(a) to compare CRUTEM3 and CRUTEM4 trends over this period. Grid boxes with insufficient data (<48 years during 1951-2010) in CRUTEM3 but sufficient data in CRUTEM4 are outlined in black; black crosses indicate the reverse situation.

Figure 5: Seasonal and annual averages by hemisphere for CRUTEM4 compared to 5 sets of independent station data (each representing roughly 20% of the total station dataset). The five different subsets are referred to as A to E, indicated by different coloured lines. The data are plotted smoothed using a 21 point binomial filter as used in Brohan *et al.* (2006). (a) NH and (b) SH. The green swathe is the uncertainty range from 2.5 to 97.5% calculated using Brohan et al. (2006) at this smoothing timescale.

Figure 6: Seasonal and annual averages by hemisphere for CRUTEM4 compared to (a) excluding all stations from the contiguous United States from the NH and (b) excluding all stations from Australia from the SH. Smoothing and linestyles as in Figure 2.

Figure 7: Seasonal and annual averages for the NH for CRUTEM4 compared to excluding all stations from Russia. Smoothing and linestyles as in Figure 2.

Figure 8: Seasonal and annual averages for the NH for CRUTEM4 compared to excluding all stations from the former Soviet Union. Smoothing and linestyles as in Figure 2.

Figure 9: Seasonal and annual averages for CRUTEM4 compared to similar series developed by NCDC (Smith and Reynolds, 2008) and GISS (Hansen *et al*., 2010). Only the smoothed series from NCDC and GISS are shown. The smoothing here is the same as Figure 2, but the green swathe encompasses the 2.5 and 97.5% uncertainty range calculated at the interannual timescale using the approach of Brohan *et al*. (2006). (a) NH and (b) SH.

Figure 10: Seasonal and annual averages for CRUTEM4 compared to two versions of the ECMWF Reanalyses (red ERA-40 from 1958-2001 and blue ERA-Interim from 1979-2010). The two reanalyses have been set to a base period of 1981-2000, so are offset slightly cooler than CRUTEM4, which uses a base period of 1961-1990. Smoothing as in Figure 2. (a) NH and (b) SH.

Figure 11: Monthly time series for ERA-Interim and CRUTEM4 (both set as anomalies by month based on the period 1979-2010) for the NH. The smoothed line is a 12-term Gaussian filter. The least-squares linear trend during the 1979-2010 overlap period (using annual averages) is shown for both series, together with its slope.

Figure 12: As Figure 8 but for (a) SH 0-60°S and (b) Antarctica (60-90°S).

Figure 13: As Figure 9, but for the SH 0-60°S.

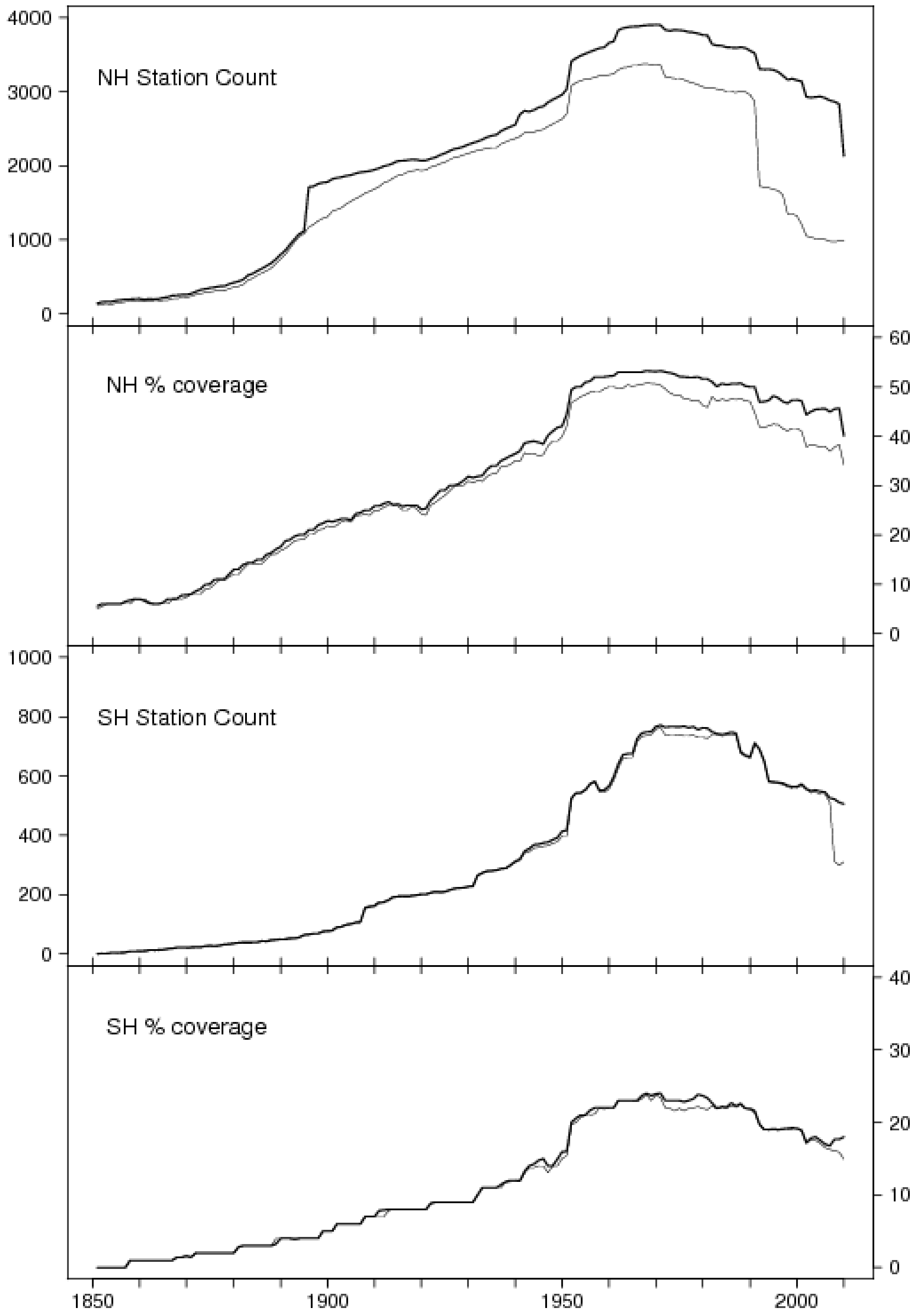


Figure 1: Comparison of station counts and percent area coverage (of the entire hemisphere including oceans) for CRUTEM4 (thick) and CRUTEM3 (thin) for the NH and SH.

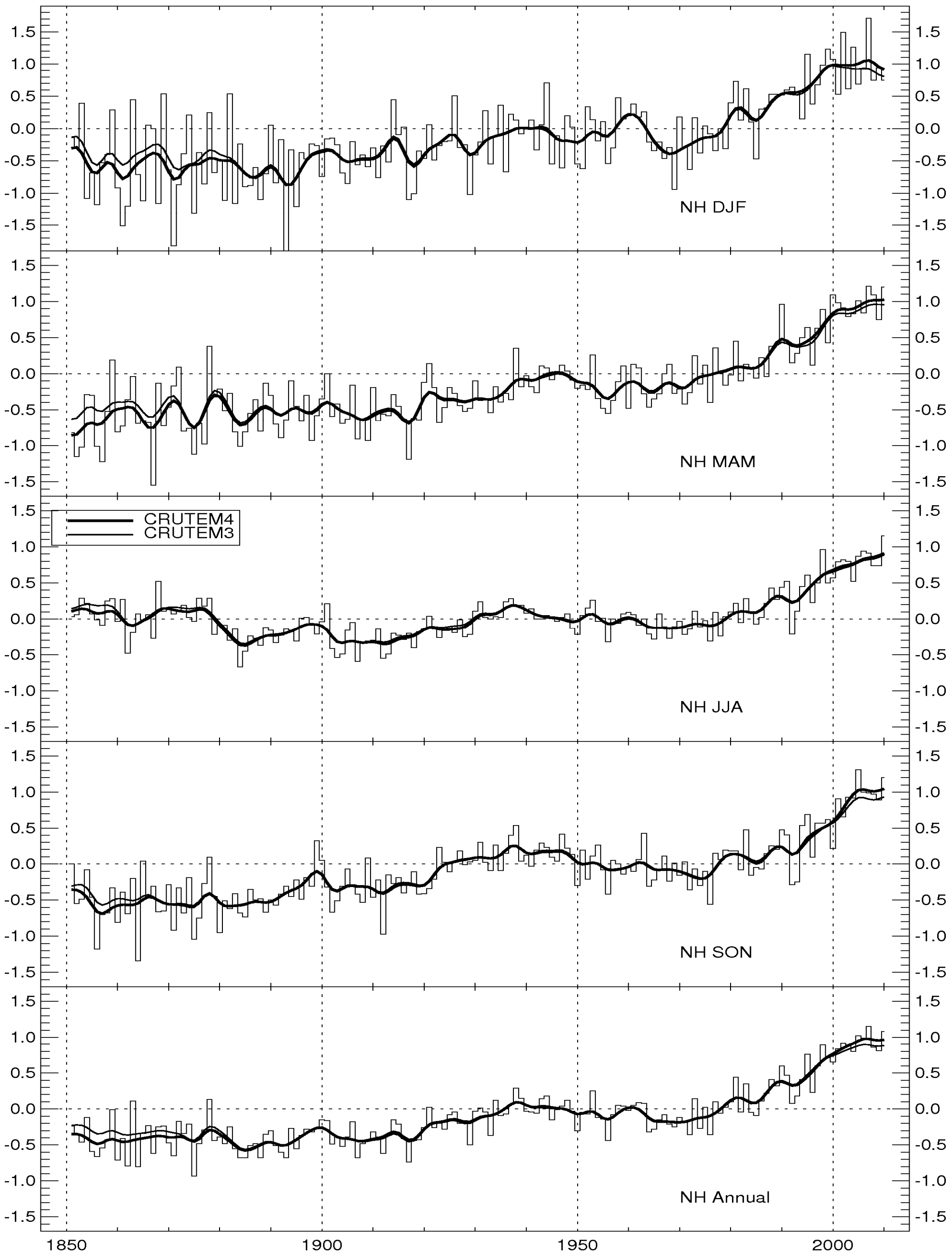


Figure 2: Seasonal and annual averages by hemisphere for CRUTEM4, with the smoothed lines showing decadal-filtered series for CRUTEM4 (thick) and CRUTEM3 (thin). Hemispheric temperature averages for the land areas are expressed as anomalies (in degrees Celsius from the base period of 1961-90). The decadal smoothing uses a 13-term Gaussian filter, padded at the ends with the mean of the adjacent 6 values. (a) NH, (b) SH and (c) global for the annual average. The global average is calculated as [(2/3)NH + (1/3)SH].

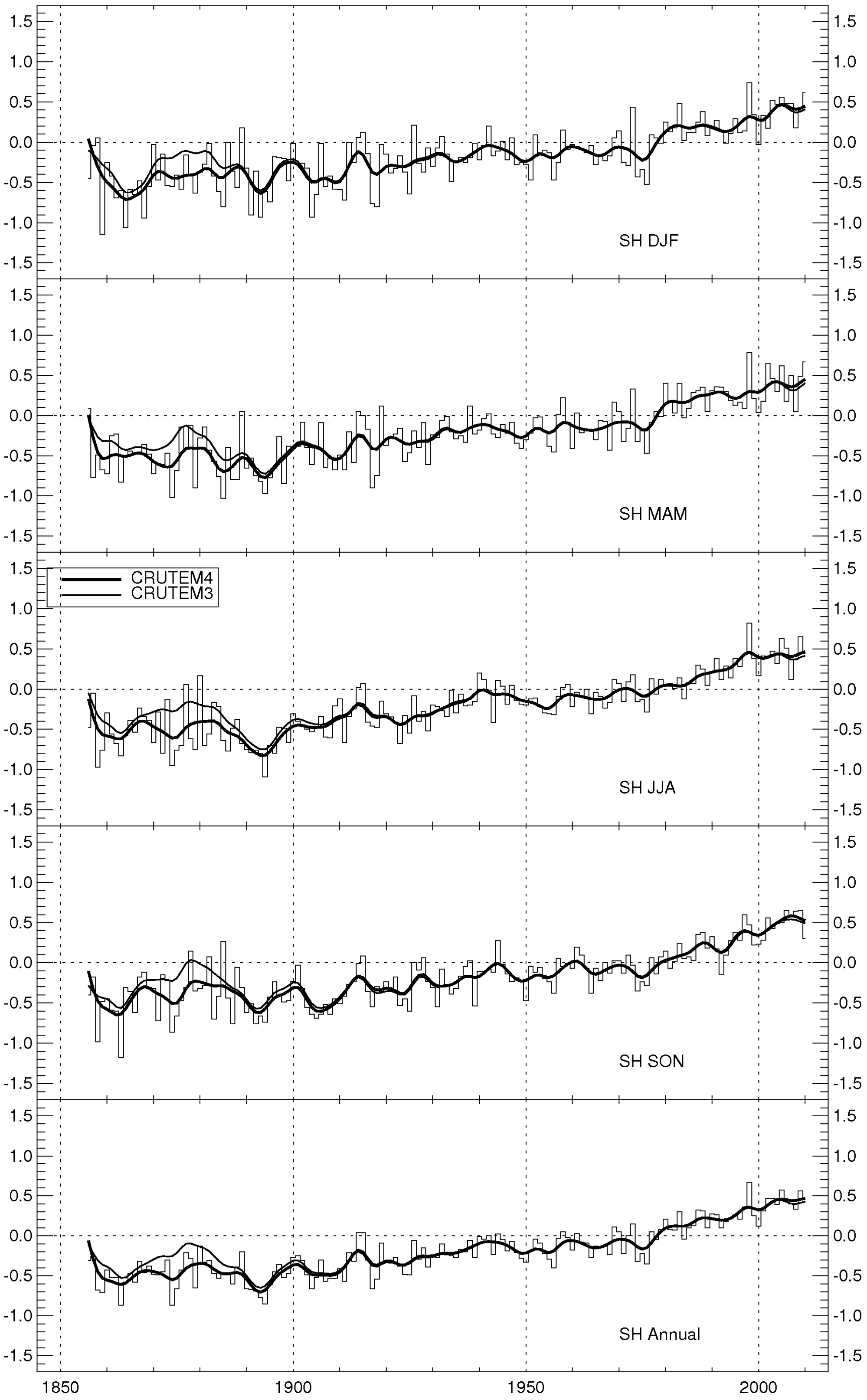


Figure 2b: see Figure 2a

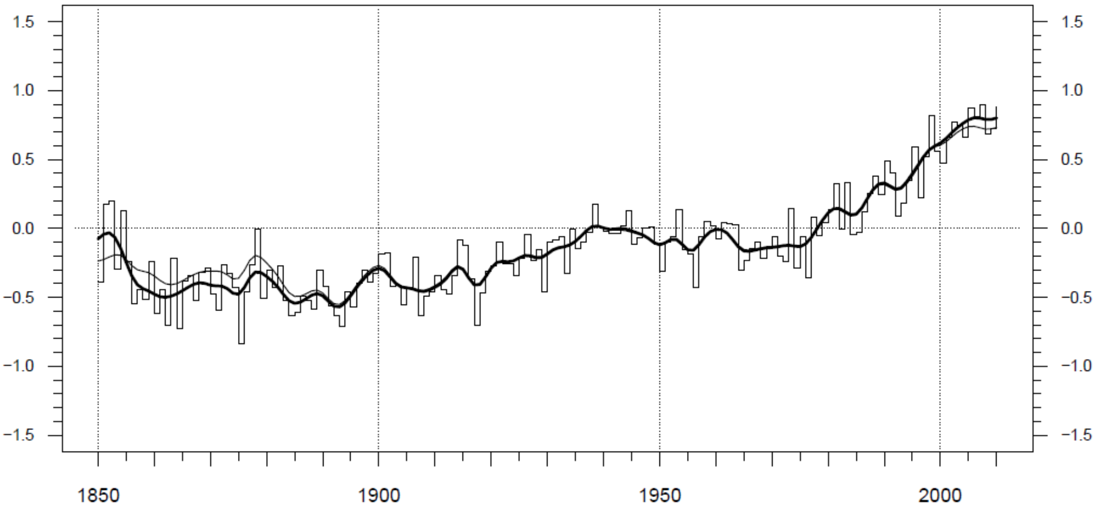
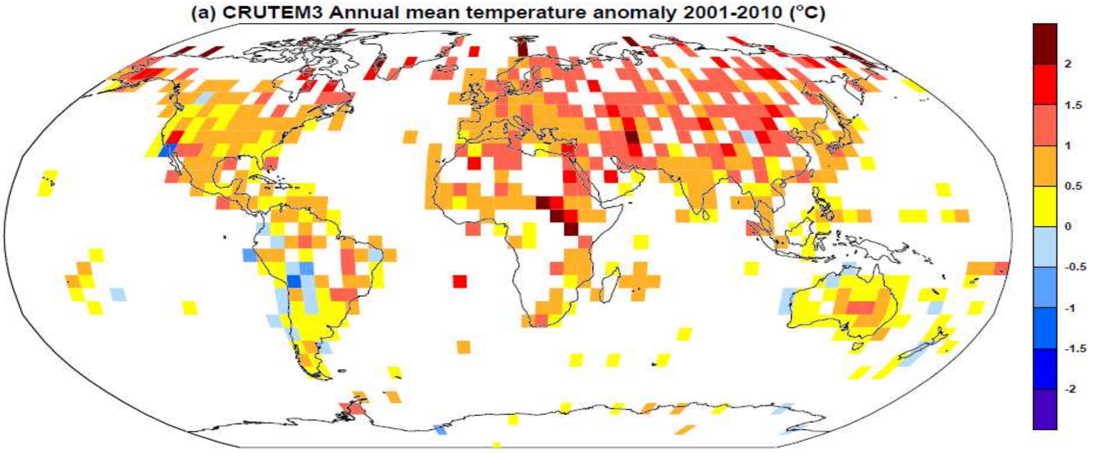
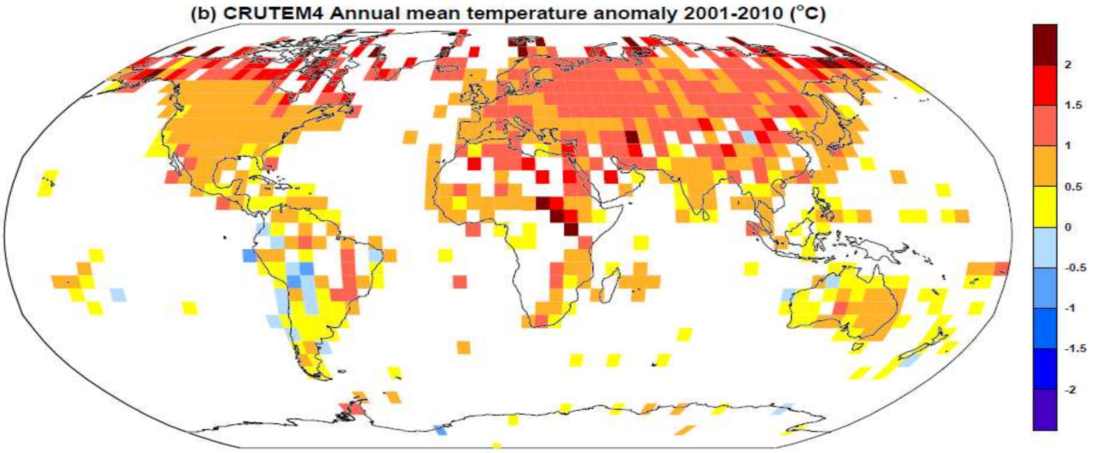


Figure 2c: see Figures 2a

(a) CRUTEM3 Annual mean temperature anomaly 2001-2010 (°C)



(b) CRUTEM4 Annual mean temperature anomaly 2001-2010 (°C)



(c) CRUTEM4-CRUTEM3 Annual difference in mean temperature (°C)

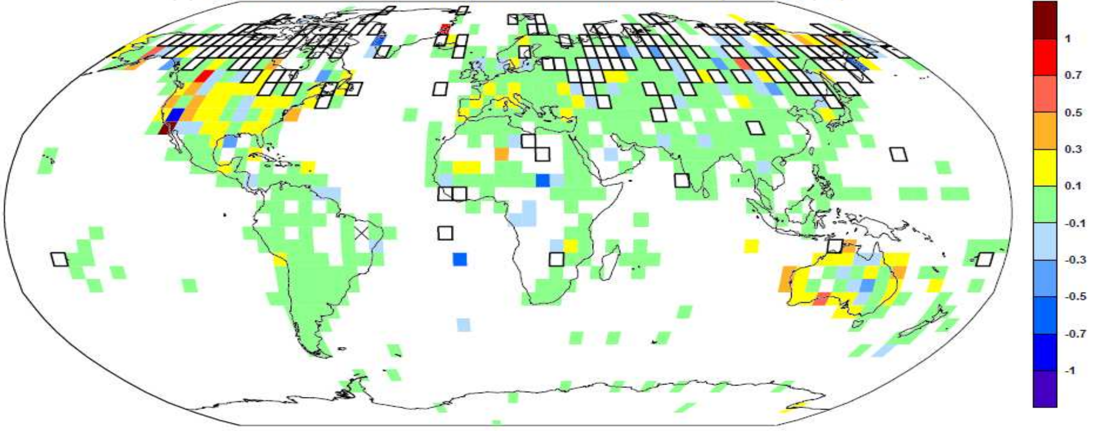


Figure 3: Comparison of annual mean temperature anomalies from (a) CRUTEM3 and (b) CRUTEM4 for the period 2001-2010 (degC anomalies from 1961-90). Grid boxes with less than 50% data coverage (5 years) are left white. (c) Difference (b)-(a) to compare CRUTEM3 and CRUTEM4 means over this period. Grid boxes with insufficient data (<5 years during 2001-2010) in CRUTEM3 but sufficient data in CRUTEM4 are outlined in black; black crosses indicate the reverse situation.

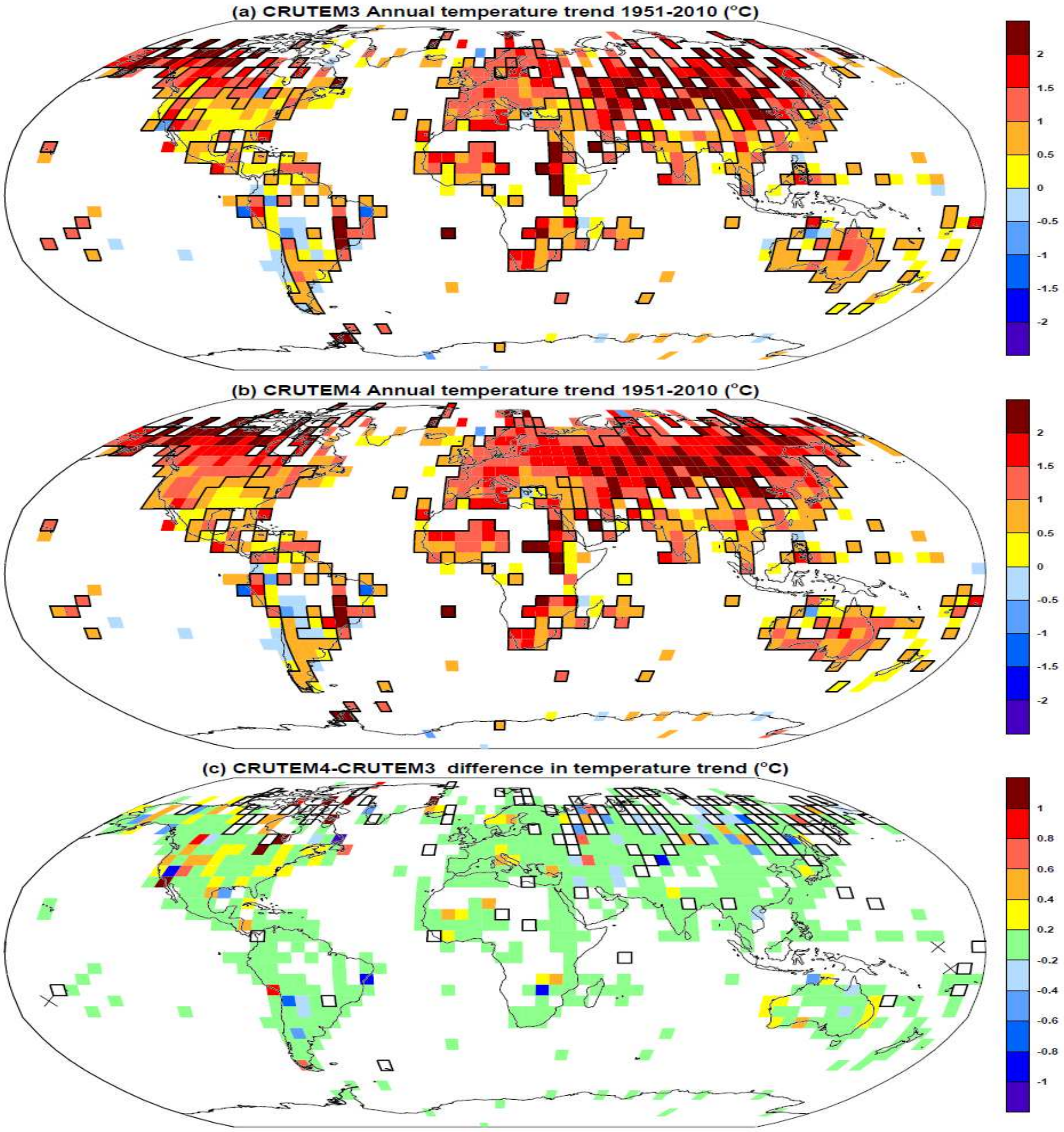


Figure 4: Comparison of linear trends fitted to (a) CRUTEM3 and (b) CRUTEM4 annual temperatures for the period 1951-2010. Trends are expressed as the degC linear trend change over the 60 year period. Grid boxes with less than 80% data coverage (48 years) are left white. Boxes or regions outlined in black are those where the trend slopes are significantly different from zero, with 95% confidence taking into account first-order autocorrelation. (c) Difference (b)-(a) to compare CRUTEM3 and CRUTEM4 trends over this period. Grid boxes with insufficient data (<48 years during 1951-2010) in CRUTEM3 but sufficient data in CRUTEM4 are outlined in black; black crosses indicate the reverse situation.

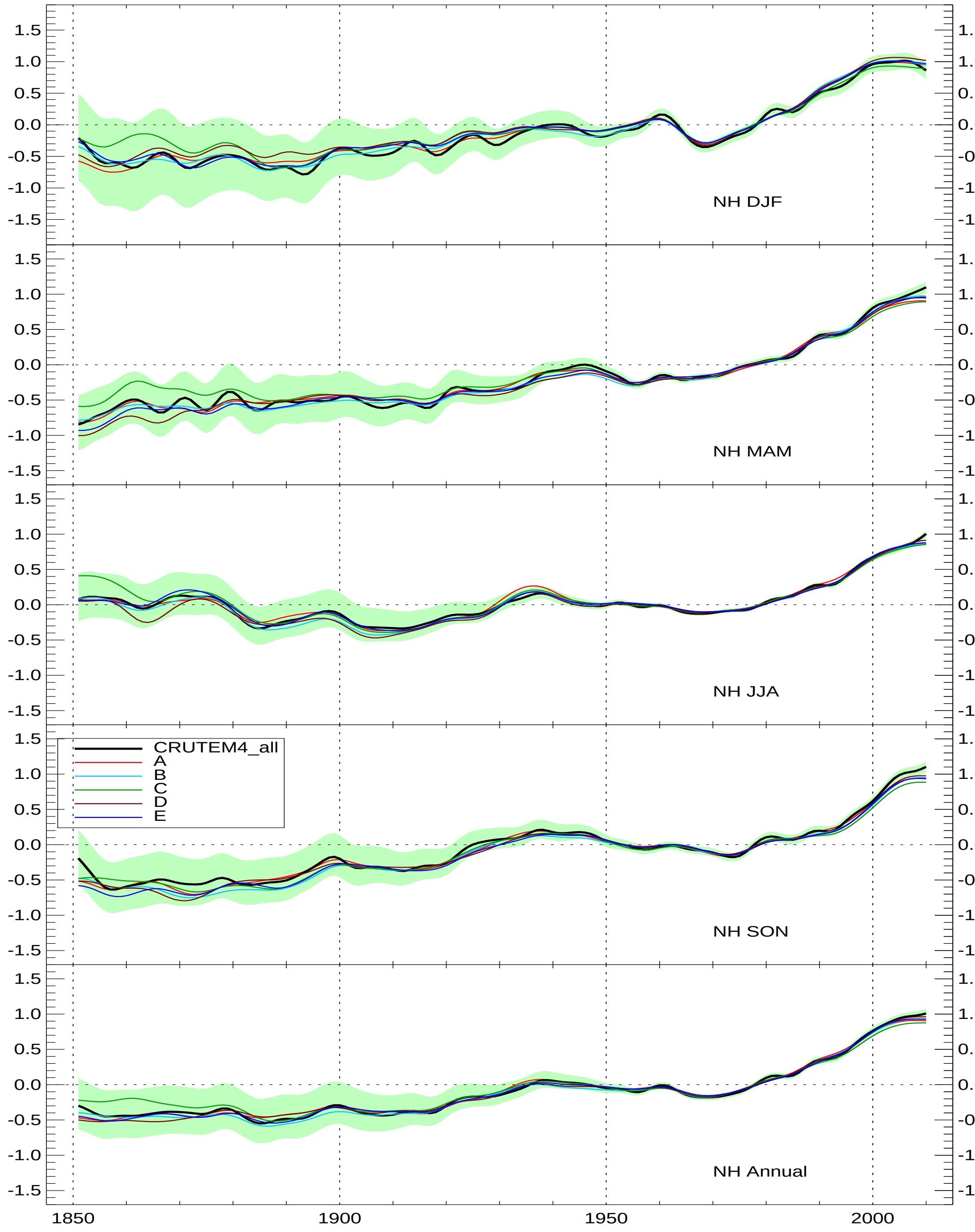
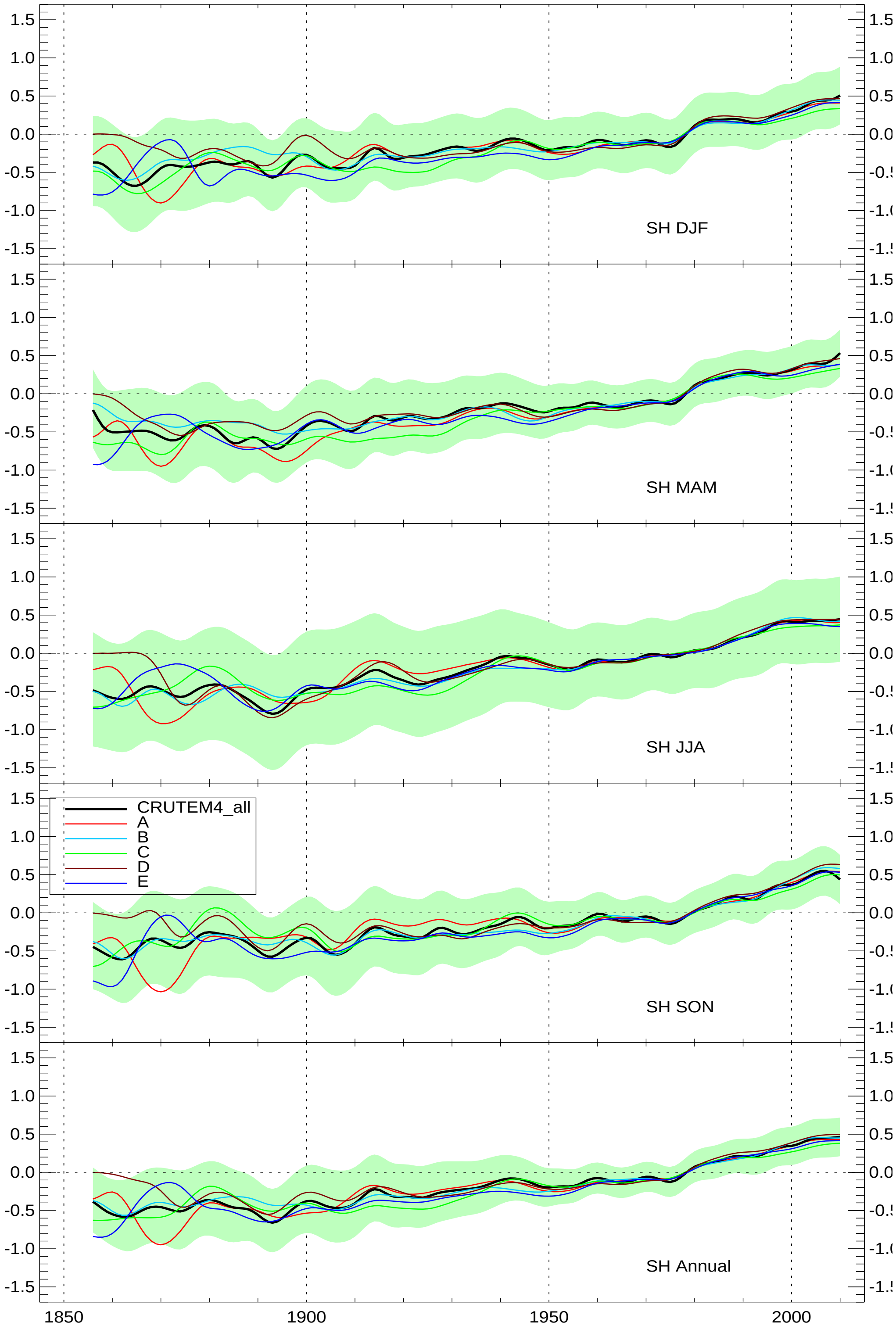


Figure 5a: Seasonal and annual averages by hemisphere for CRUTEM4 compared to 5 sets of independent station data (each representing roughly 20% of the total station dataset). The five different subsets are referred to as A to E, indicated by different coloured lines. The data are plotted smoothed using a 21 point binomial filter as used in Brohan *et al.* (2006). (a) NH and (b) SH. The green swathe is the uncertainty range from 2.5 to 97.5% calculated using Brohan et al. (2006) at this smoothing timescale.



1165

1166 Figure 5b: See Figure 5a

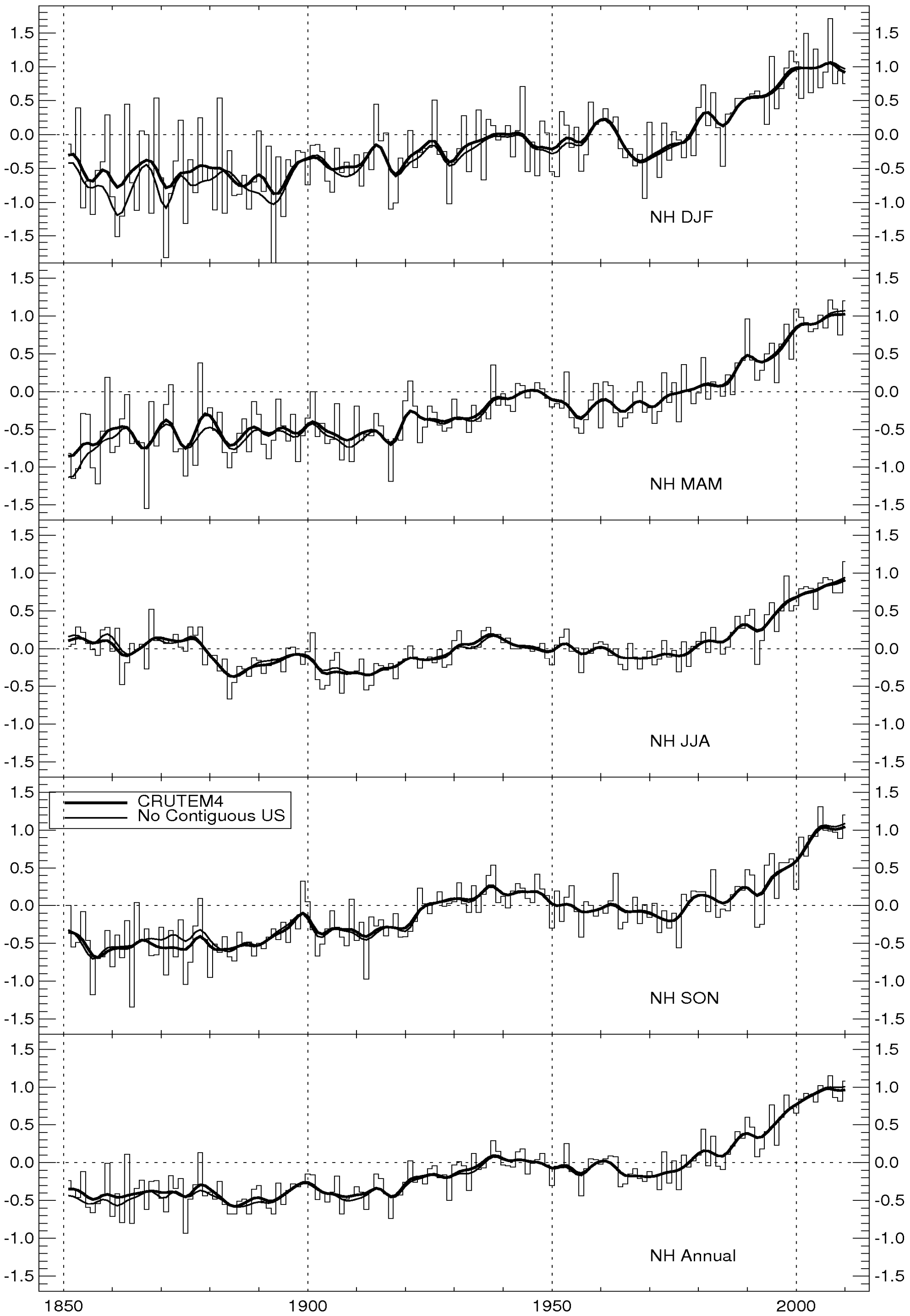


Figure 6a: Seasonal and annual averages by hemisphere for CRUTEM4 compared to (a) excluding all stations from the contiguous United States from the NH and (b) excluding all stations from Australia from the SH. Smoothing and linestyles as in Figure 2.

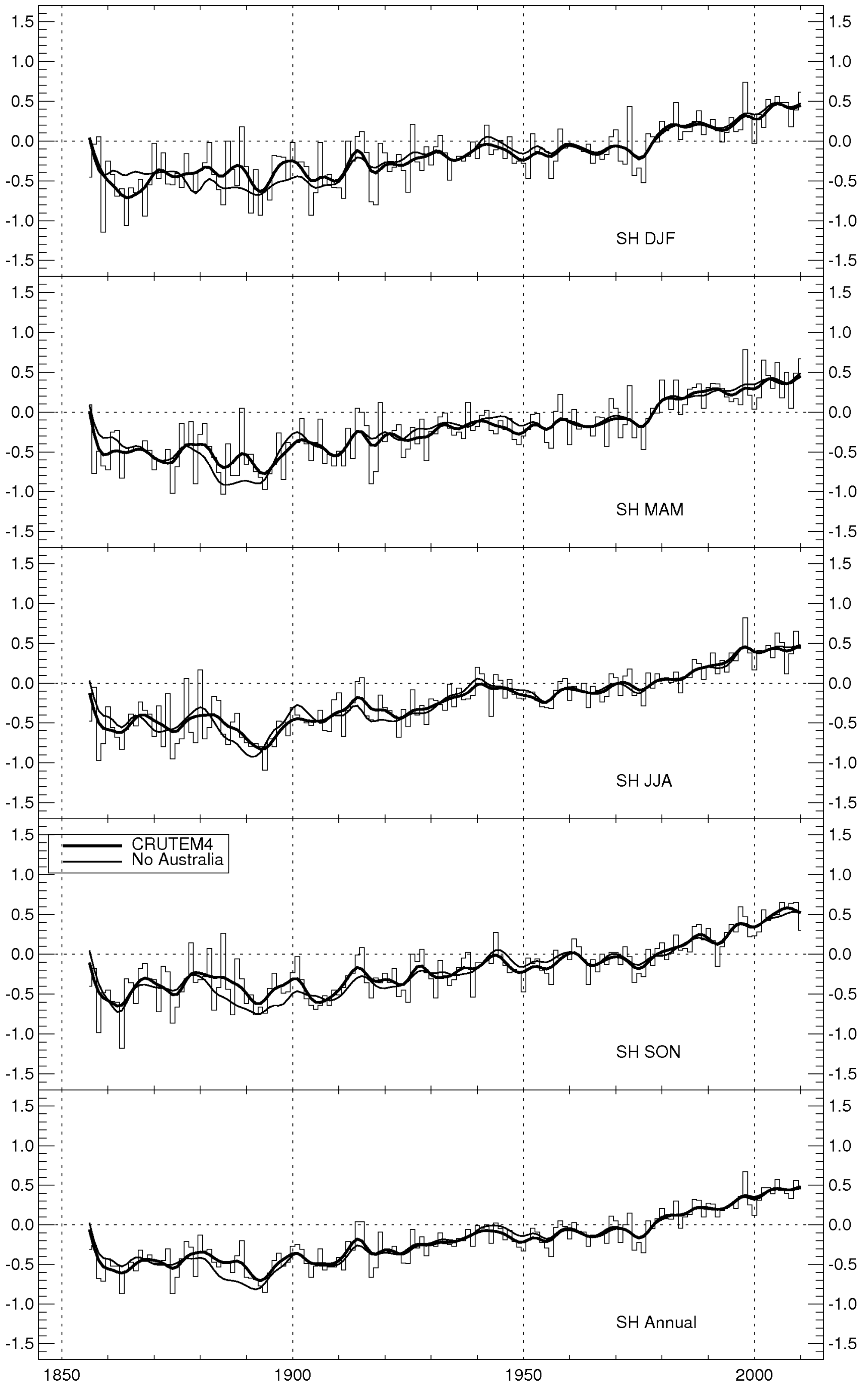


Figure 6b: Without Australia – see Figure 6a.

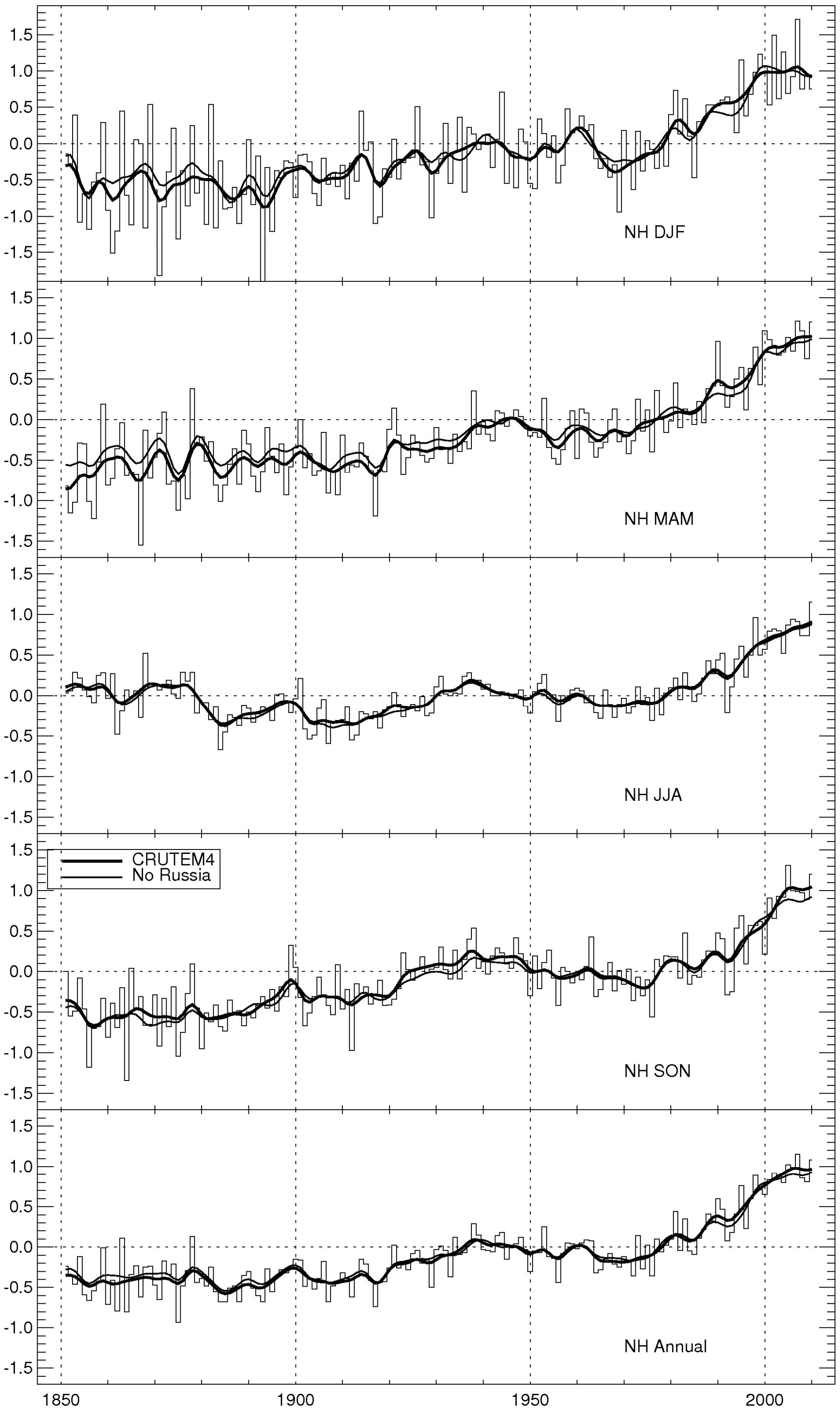


Figure 7: Seasonal and annual averages for the NH for CRUTEM4 compared to excluding all stations from Russia. Smoothing and linestyles as in Figure 2.

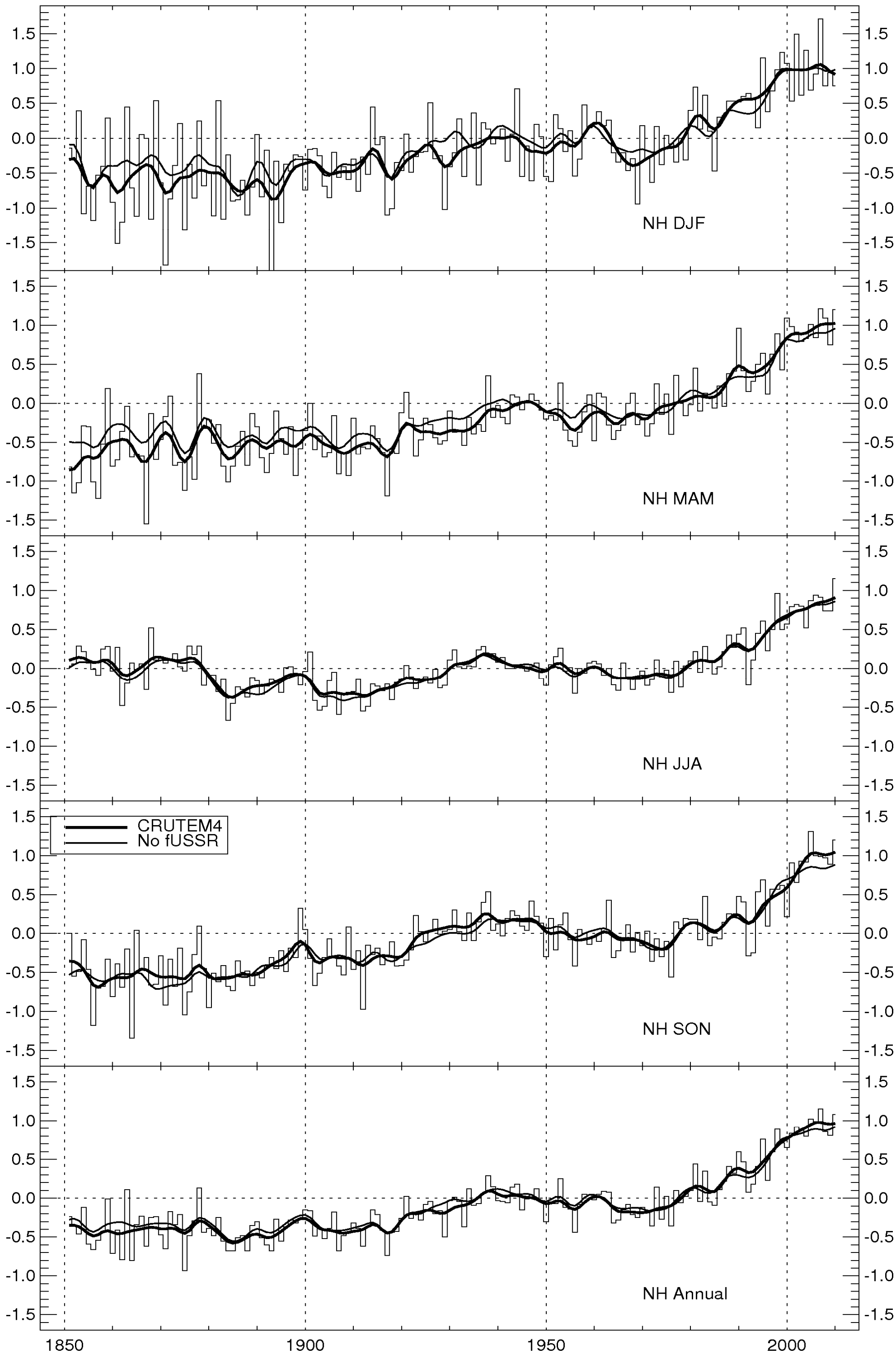


Figure 8: Seasonal and annual averages for the NH for CRUTEM4 compared to excluding all stations from the former Soviet Union. Smoothing and linestyles as in Figure 2.

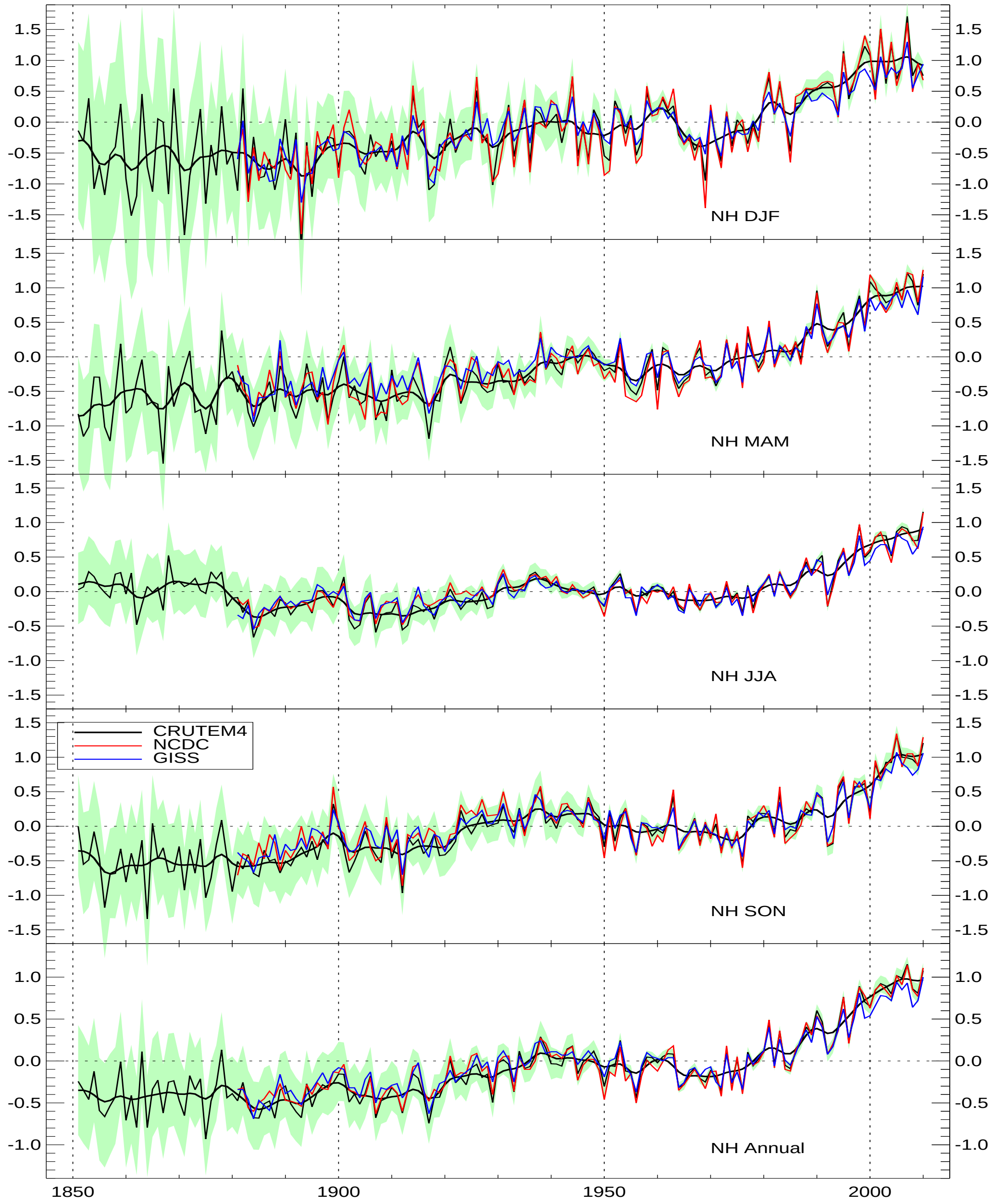


Figure 9a: Seasonal and annual averages for CRUTEM4 compared to similar series developed by NCDC (Smith and Reynolds, 2008) and GISS (Hansen *et al*., 2010). Only the smoothed series from NCDC and GISS are shown. The smoothing here is the same as Figure 2, but the green swathe encompasses the 2.5 and 97.5% uncertainty range calculated at the interannual timescale using the approach of Brohan *et al*. (2006). (a) NH and (b) SH.

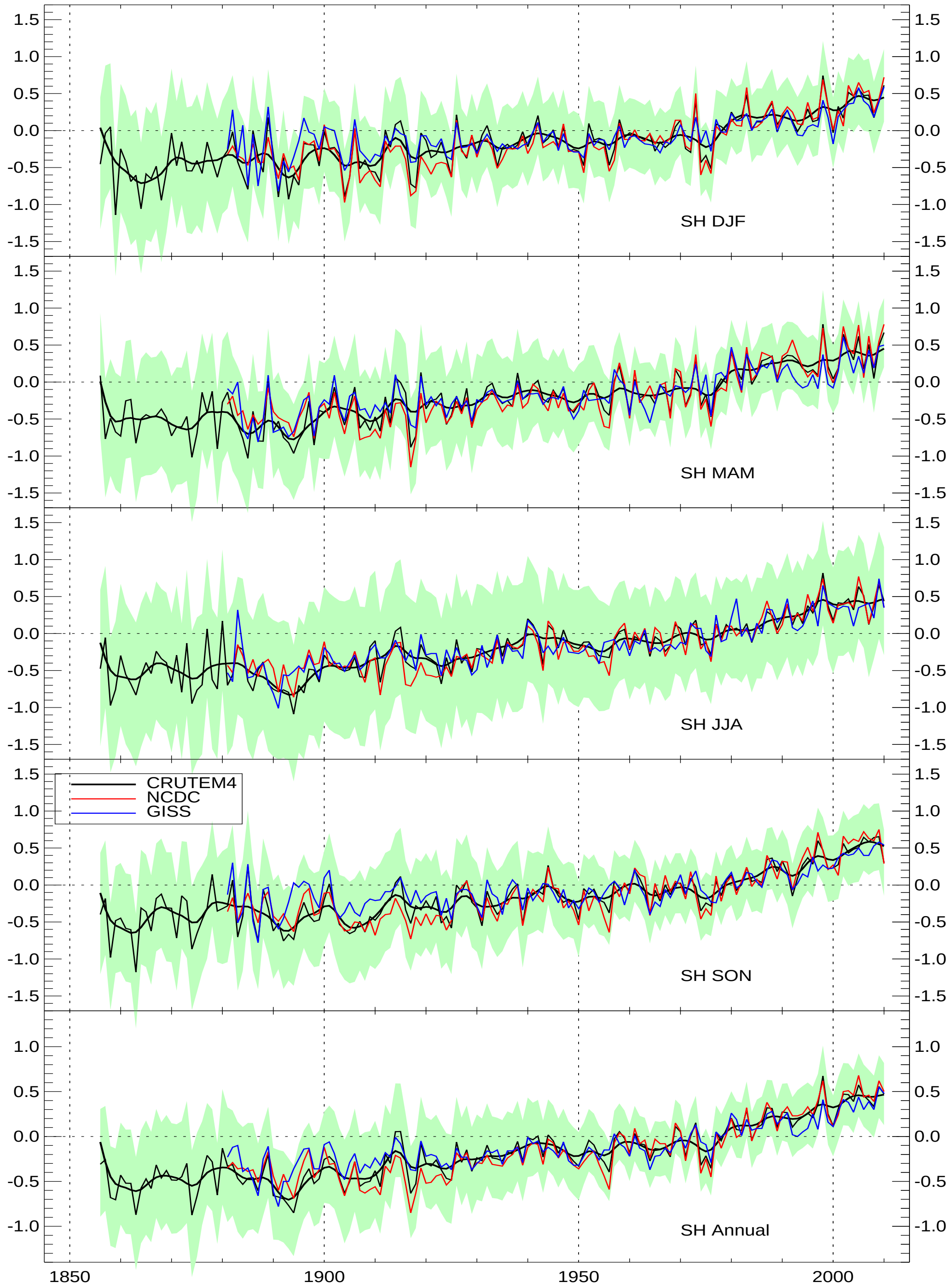


Figure 9b: see Figure 9, but for the SH

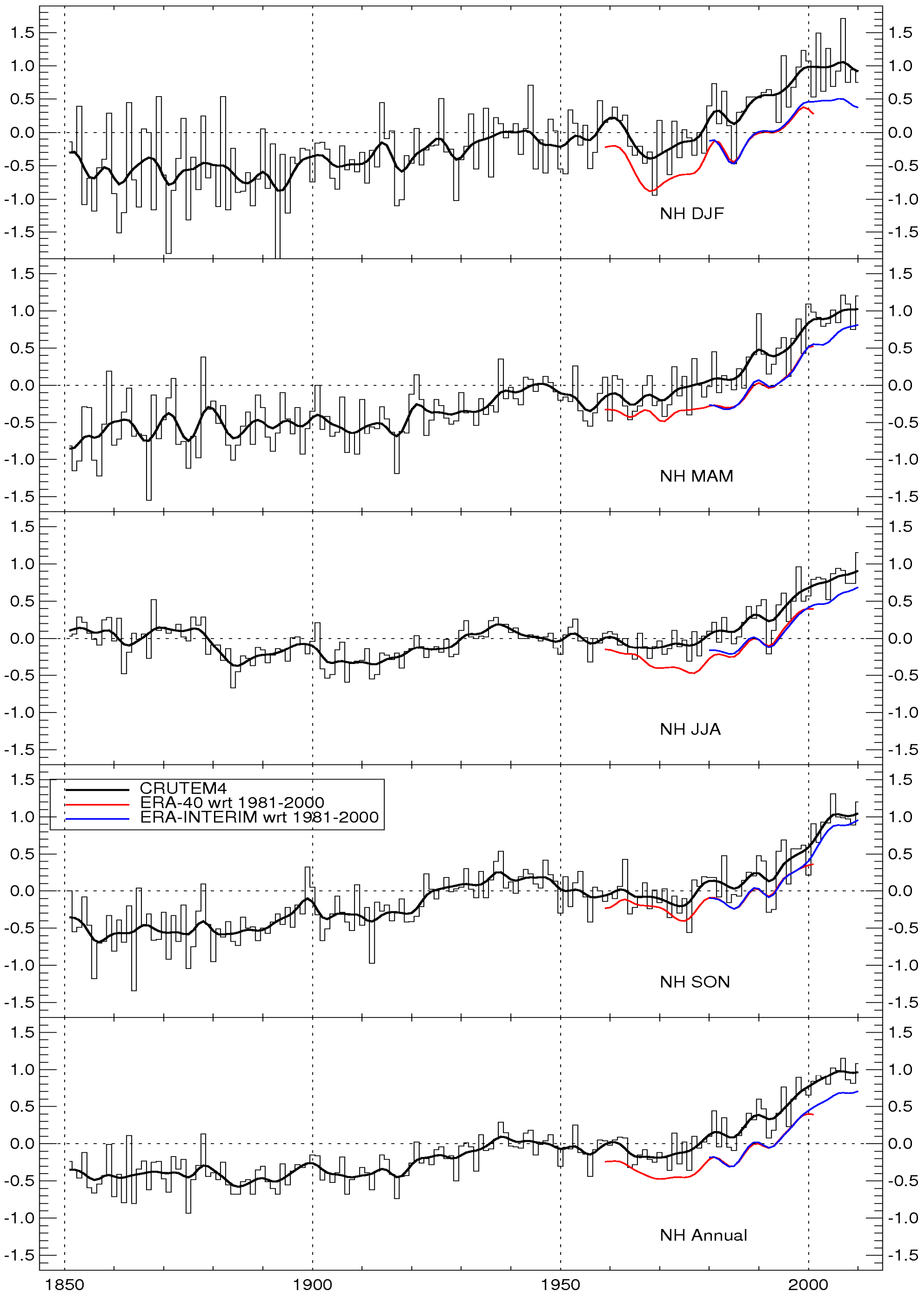
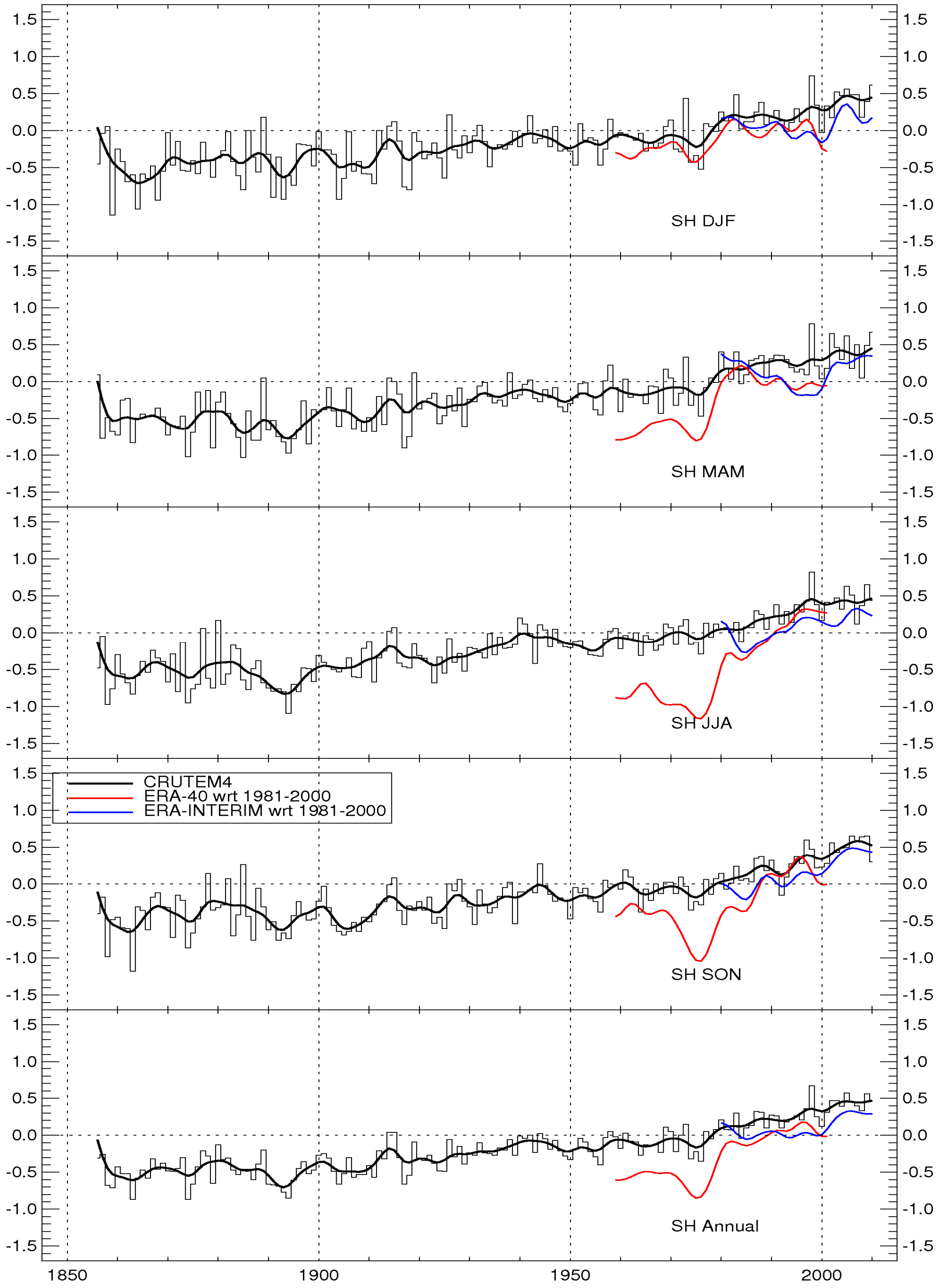


Figure 10a: Seasonal and annual averages for CRUTEM4 compared to two versions of the ECMWF Reanalyses (red ERA-40 from 1958-2001 and blue ERA-Interim from 1979-2010). The two reanalyses have been set to a base period of 1981-2000, so are offset slightly cooler than CRUTEM4, which uses a base period of 1961-1990. Smoothing as in Figure 2. (a) NH and (b) SH.

1205



1206

1207 Figure 10b: see Figure 10a

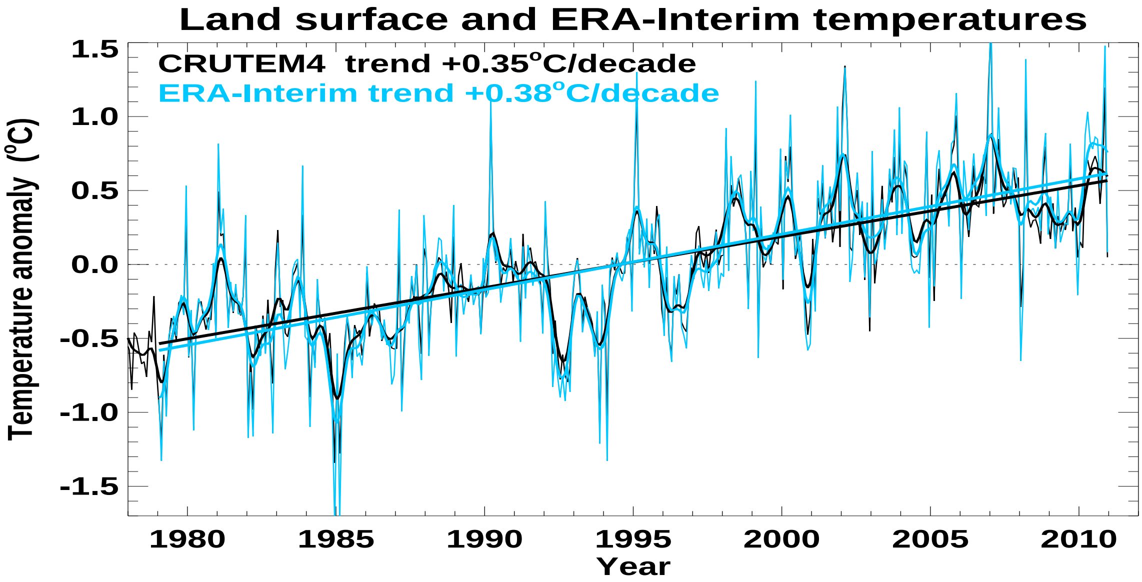


Figure 11: Monthly time series for ERA-Interim and CRUTEM4 (both set as anomalies by month based on the period 1979-2010) for the NH. The smoothed line is a 12-term Gaussian filter. The least-squares linear trend during the 1979-2010 overlap period (using annual averages) is shown for both series, together with its slope.

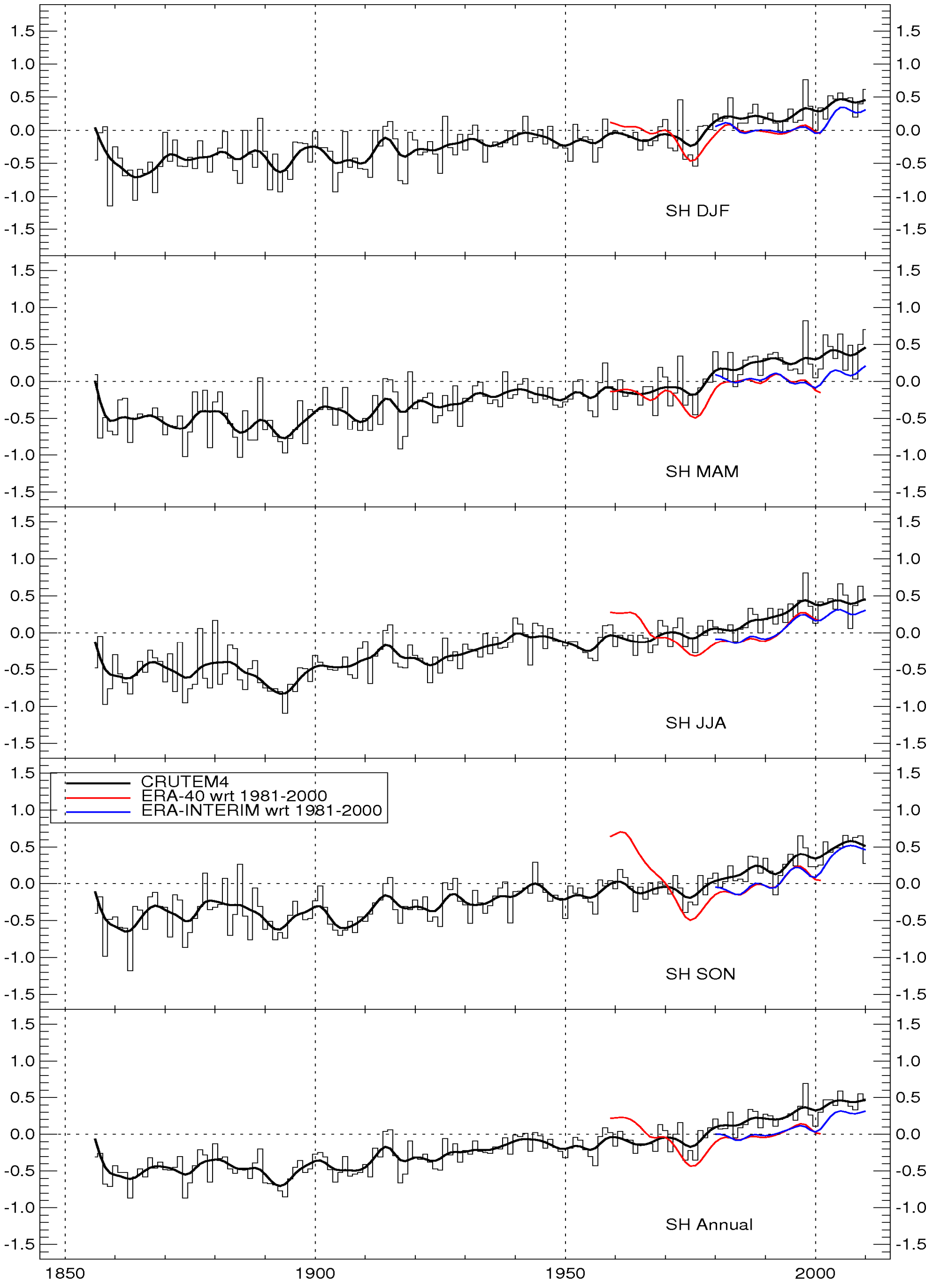


Figure 12a: As Figure 8 but for (a) SH 0-60°S and (b) Antarctica (60-90°S).

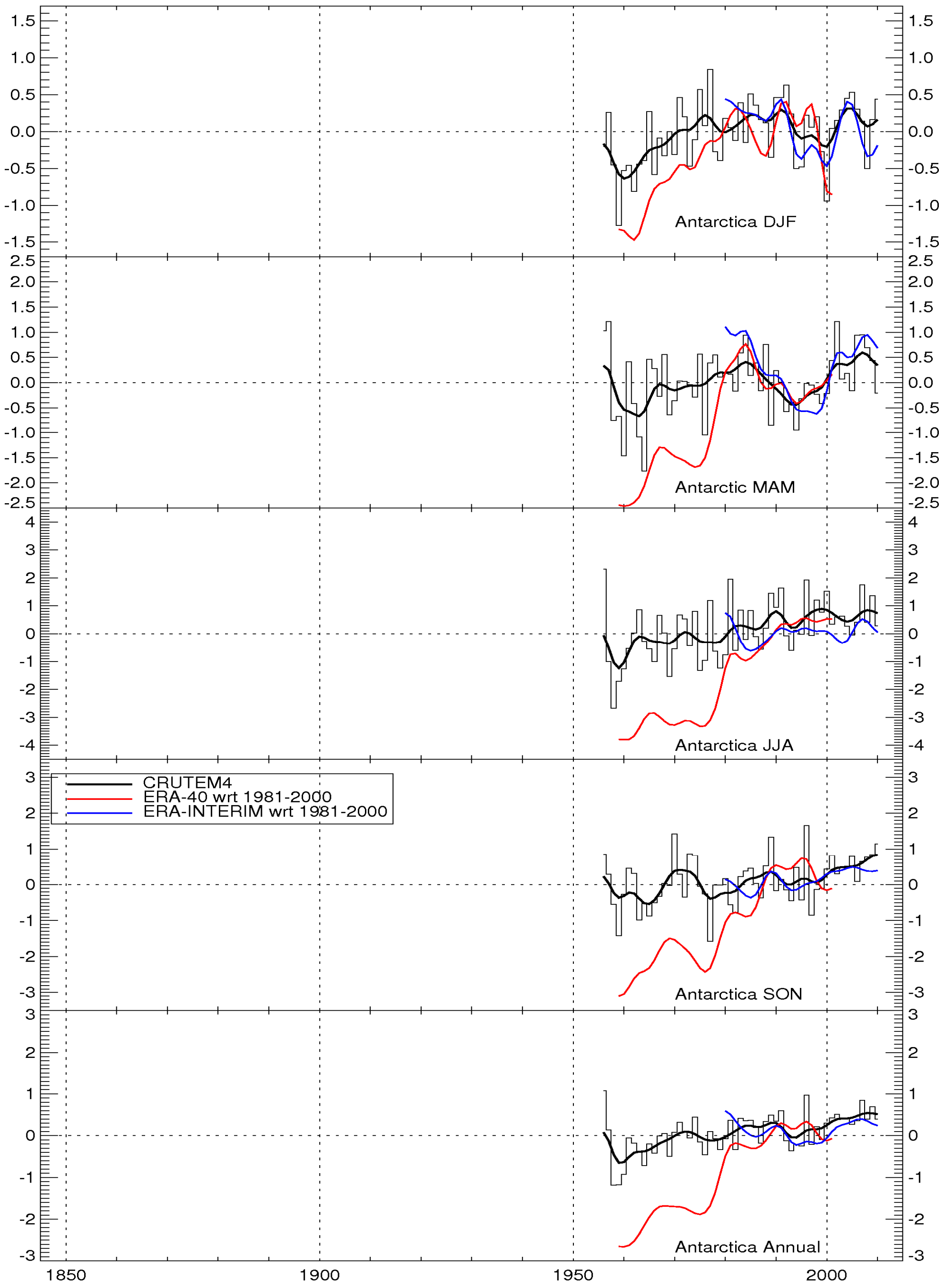


Figure 12b: see Figure 12a

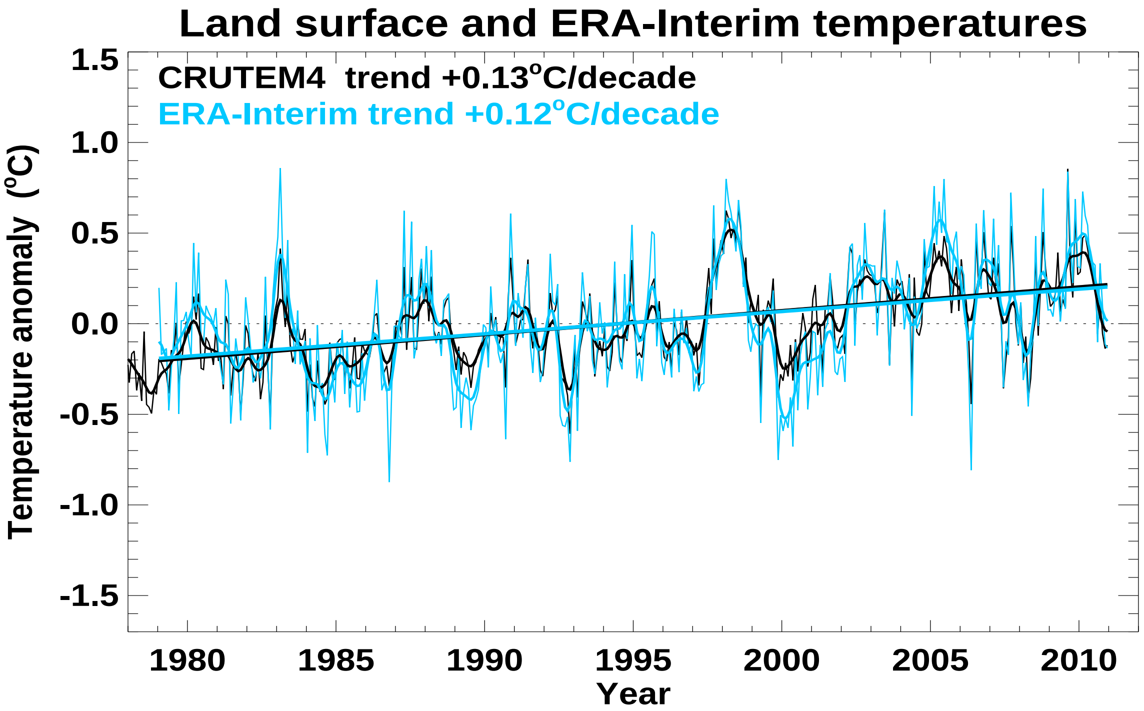


Figure 13: As Figure 9, but for the SH 0-60°S.