Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 dataset

T. J. Osborn[1], P. D. Jones[1], D. H. Lister[1], C. P. Morice[2], I. R. Simpson[2], J. P. Winn[2], E. Hogan[2] and I. C. Harris[1,3]

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

[2]Met Office Hadley Centre, Exeter, EX1 3PB, UK

[3]National Centre for Atmospheric Science, University of East Anglia, NR4 7TJ, UK

Corresponding author: Timothy Osborn (t.osborn@uea.ac.uk)

Key Points:

- We describe the fifth major update of a dataset of global land air temperature, CRUTEM5
- CRUTEM5 is based on an expanded network of 7983 stations (up from 4842) and improved identification of outlier values.
- The estimated increase in global land air temperature from 1861–1900 to 2015–2019 is 1.6°C (or 1.7°C with an alternative gridding method)

Abstract

CRUTEM5 (Climatic Research Unit temperature, version 5) is an extensive revision of our land surface air temperature dataset. We have expanded the underlying compilation of monthly temperature records from 5583 to 10639 stations, of which those with sufficient data to be used in the gridded dataset has grown from 4842 to 7983. Many station records have also been extended or replaced by series that have been homogenized by national meteorological and hydrological services. We have improved the identification of potential outliers in these data to better capture outliers during the reference period; to avoid classifying some real regional temperature extremes as outliers; and to reduce trends in outlier counts arising from climatic warming. Due to these updates, the gridded dataset shows some regional increases in station density and regional changes in temperature anomalies. Nonetheless, the global-mean timeseries of land air temperature is only slightly modified compared with previous versions and previous conclusions are not altered. The standard gridding algorithm and comprehensive error model are the same as for the previous version, but we have explored an alternative gridding algorithm that removes the under-representation of high latitude stations. The alternative gridding increases estimated global-mean land warming by about 0.1°C over the course of the whole record. The warming from 1861–1900 to the mean of the last 5 years is estimated as 1.6°C using the standard gridding (with a 95% confidence interval on individual annual means of -0.11 to +0.10°C in recent years), while the alternative gridding gives a change of 1.7°C.

4 May 2020

1 Introduction

CRUTEM (Climatic Research Unit temperature) is a gridded dataset of monthly near-surface air temperature anomalies over the land surfaces of the world, running from 1850 to the present. We have undertaken the fifth major update (CRUTEM5.0) of this dataset since it was first published in the 1980s, and here we describe the changes since the previous version (CRUTEM4.0) was published in 2012 (Jones et al., 2012). This is a collaborative project between the Climatic Research Unit (CRU), the Met Office Hadley Centre and the National Centre for Atmospheric Science (NCAS). The temperature anomalies from CRUTEM form the land component of the global land and marine surface temperature dataset HadCRUT, with the Met Office Hadley Centre sea surface temperature (SST) dataset HadSST providing the marine component. Currently, HadCRUT4 (Morice et al., 2012) comprises land air temperatures from CRUTEM4 and SST from HadSST3 (Kennedy et al., 2011b, 2011a); HadCRUT5 will combine the new land air temperature dataset reported here with the recently-published HadSST4 (Kennedy et al., 2019).

There have been several important developments since 2012 for understanding and improving global (land-only and land-and-marine) temperature datasets. First, there have been further data rescue, data compilation and data homogeneity exercises at national, regional and global scales. Examples of the national and regional exercises are given later in this paper where we describe the acquisition of new or improved data into the CRUTEM5.0 compilation. The International Surface Temperature Initiative (ISTI; Rennie et al., 2014) and version 4 of the Global Historical Climatology Network (GHCN; Menne et al., 2018) provide updated global compilations of daily or monthly temperatures. Second, there is now better understanding of the sources of bias in global land temperature datasets, such as urbanization (Wang et al., 2015; Wickham et al., 2013) and lack of complete observational coverage (Cowtan et al., 2018; Cowtan & Way, 2014). Third, new global temperature datasets have been constructed using different methodological approaches, such as the Berkeley Earth and China Meteorological Administration (CMA) datasets (Rohde et al., 2013; Xu et al., 2018). Fourth, new reanalysis datasets are good enough to provide a useful and partially independent alternative for comparison with the traditional temperature datasets in recent decades. Reanalyses complement the traditional datasets because they utilise multi-variate observations (rather than only near surface temperature) and the physical processes represented within numerical models of the atmosphere (rather than statistical models) to obtain spatially complete fields.

There are multiple approaches to constructing a global temperature record and this enables some of the structural uncertainty, arising from choices of method (Thorne et al., 2011), to be sampled by making comparisons across the different datasets. Thus, it is important to continue to update (and improve) the dataset obtained using the CRUTEM approach as a contribution to this ensemble of structurally different datasets. It is useful, therefore, to list the general principles that guide the CRUTEM approach and to note where these differ from other global temperature datasets. There is now an almost 40-year history to this dataset and updates using the same overall approach (albeit with some modifications where improvements can be made) are valuable to allow comparisons to be made that depend mostly on updated data rather than methodological changes. For CRUTEM we do not apply global, statistical algorithms to identify and correct for inhomogeneities: instead we utilise homogenization efforts undertaken by national or regional initiatives, which may benefit from the knowledge of local circumstances or additional observing stations. We also use a simple gridding approach, with grid cell temperature anomalies based on station observations within the grid cell rather than relying on extra information from more distant stations. Though this reduces the spatial coverage of the dataset, the simplicity of the approach makes it more transparent and easier for others to reproduce. The bias introduced by incomplete global coverage (Cowtan & Way, 2014) will be addressed in the forthcoming HadCRUT5 dataset (Morice et al., submitted). Finally, the CRUTEM error model is quite comprehensive and was the first of its type applied to a global temperature dataset (Brohan et al., 2006), though other datasets have increasingly comprehensive error models (e.g. GISTEMP; Lenssen et al., 2019).

We start in section 2 by describing the new and updated data sources that we have included in the CRUTEM5.0 station temperature database (with more comprehensive listings given in the Supplementary Material, SM). We then develop improvements to the process for identifying and removing potential outlier observations (section 3) and consider the representation of high latitude stations when using a regular latitude-longitude grid that has longitudinally-slim grid cells at high latitudes (section 4). In section 5 we compare the effect on global-mean land temperature, in turn, of the changes to the station database, changes to the outlier checking and an alternative gridding method. Some results at continental or sub-continental scales are given in the SM.

2 Station data sources and updates

The CRUTEM station database comprises records obtained from global or regional compilations and records acquired from individual national meteorological and hydrological services (NMHS), as described by (Jones et al., 2012) and earlier CRUTEM papers (see list in Osborn & Jones, 2014). The database is updated monthly from CLIMAT and Monthly Climatic Data for the World (MCDW) circulations. Significant effort is expended to continually extend and improve the station database beyond these monthly updates. This effort is important as many long records in many regions do not report in real time over the CLIMAT network. More stations report sub-daily or daily values over the SYNOP network, and some groups (e.g. National Oceanographic and Aeronautical Administration, NOAA) extract monthly averages from SYNOP messages. The SYNOP data are not used here because they have been subject to less quality control (QC), the calculation of daily means may be incompatible (based on different or incomplete observation times) with the climate data shared later and decisions need to be made about the number of missing values in a month that will be allowed. Series acquired directly from NMHS are more likely to be based on complete observations and to have undergone more QC.

Data acquisitions can be in the form of stations not previously in CRUTEM, additional data to augment stations already in CRUTEM, or homogenized data to replace values already in CRUTEM. The latter are particularly important given the CRUTEM principle to utilise nationally-homogenised records in preference to applying global statistical algorithms to remove inhomogeneities. In some cases (Table 1) these homogenised series are consistently and regularly updated and we access them every one or two years. The sources for USA, Canada and Australia use homogenization schemes which are re-applied to each update or when additional data become available; these changes then become incorporated into CRUTEM each time we update those series. Other acquisitions are more irregular and typically arise from either specific regional homogenization or data rescue projects or personal contacts within NMHS; in some cases we identify specific issues (e.g. lack of routine updates or data sparseness) and focus on acquiring data to address them.

To facilitate updating of the series we utilise World Meteorological Organisation (WMO) ID codes where they exist (or assign a WMO-style CRUTEM ID code if not) and map these to the domestic ID codes used by some data sources, especially the larger NMHS. Some ID codes change over time, perhaps reflecting a composite series that has been homogenised. We rarely merge series from multiple nearby sites, though we occasionally combine series where a long record stops and is replaced by a new one with a different WMO identifier: in those cases, extra checks are undertaken with respect to the identifiers and locations to ensure that incompatible series are not merged. Using the current WMO ID codes enables the series to be updated routinely with CLIMAT and MCDW data. Similarly, where homogenisation has been undertaken, it is convenient to homogenise earlier data so that it is comparable to the most recent data (rather than vice versa), so that routine updates are compatible with the existing record. In practice, this is not always the case so routine monthly updates may subsequently be replaced by series received from the NMHS. An example is that many Chinese records in CRUTEM are based on the mean of the daily minimum (Tmin) and maximum (Tmax) temperatures, while the monthly mean temperatures from CLIMAT are calculated from 6-hourly observations, which tends to give lower values. The monthly CLIMAT updates therefore extend the records in near real time but with a relative cool bias; this bias is then removed when the annual or biennial acquisition of data from the CMA replaces the CLIMAT values in the CRUTEM database.

Biases in station data are discussed by Jones (2016) and are represented in the CRUTEM error model (Brohan et al., 2006). Of these biases, urbanization influences deserve particular attention in rapidly urbanizing regions such as China, and this influence can be exacerbated by unrepresentative observing networks (e.g. only 0.7% of the area of China is classified as urban yet 68% of stations are in urban locations; Wang et al., 2015). Sun et al. (2016) detect an urban warming signal in China of 0.09 °C/decade (1961–2013) that augments an inferred underlying warming of 0.18 °C/decade, indicating that a standard analysis of the available station network will overestimate the warming of this region by around 50%. In contrast, Wang et al. (2015) found a much smaller urban contribution in China, by appropriately weighting the land cover categories when averaging stations across China to reduce the urbanization bias. Their weighted series shows 0.23 °C/decade warming over 1955–2007, only 10 to 20% less than the warming exhibited by unweighted data or by a Chinese average formed from the CRU temperature data. These two studies (and others given in Table 1 of Wang et al., 2015) indicate the uncertainty in estimating the urban-warming component in warming across China.

Some simple, subjective tests are applied to newly acquired historical climate datasets prior to merging them into the CRUTEM archive. Annual and seasonal timeseries of the new and existing series are inspected visually; any apparent spikes or steps are considered more closely (e.g. by comparison with nearby series). If there is an overlap period, we compute differences on a monthly basis between new and existing series to locate systematic offsets (which might vary seasonally or occur suddenly) indicative of an inhomogeneity in one series that has been corrected in the other series or to identify other potential problems (e.g. to avoid overwriting with the wrong station if the new series has been wrongly labelled). In most cases the full length of a newly acquired series is used, overwriting existing data, rather than just adding a few years to the end of the data we already hold. This reduces the likelihood that we add a few years of incompatible data to the end of an existing series. When a newly received series can potentially be combined with an existing CRUTEM series to create a longer series, the resultant series is only retained in full if the existing data appears to be consistent with the newly received series, based on simple tests described earlier. If these tests identify any obvious inhomogeneities then the early part of the series is not used.

2.1 New data incorporated since CRUTEM4.0

Through appending CLIMAT and MCDW values, the station database and then the gridded, global and hemispheric series have been updated monthly (with no change in version number). Separate updates (approximately annually) amalgamate updates/acquisitions from more disparate sources and a change in version number (from 4.0 to 4.1, etc.) is used to indicate the non-routine nature of some of the changes. The update to CRUTEM5.0 documented here combines all these updates from 4.0 (released in 2012) to 4.6 (released in 2017) together with a further round of updates (from 4.6 to 5.0). Thus many of the station database changes reported here are already present in the publicly available CRUTEM4.6 dataset; although those changes have been documented informally (via the Met Office website https://www.metoffice.gov.uk/hadobs/crutem4/data/versions.html), this paper represents the formal publication of this significant update to the CRUTEM dataset.

Table 1 lists the sources accessed on an annual or biennial basis to update large subsets of data with series that are homogenised at a national level. These updates not only add recent observations but also improve or increase earlier data. All these have been used for the latest update from version 4.6 to 5.0. The details of the many other acquisitions are given in Supplementary Tables 1 to 7 and make the scope of this effort clear. A summary in terms of significant sources and numbers of series is given in Table 2. This illustrates our priorities in acquiring new, updated or improved data: regions with sparse data and benefitting from homogenization projects in particular.

2.2 Changes in station temporal and latitudinal coverage

The new database (CRUTEM5.0) now contains almost twice as many stations (10639) as were in CRUTEM4.0 (5583). The majority of the new acquisitions were already included by version 4.6 or earlier. Alongside additional stations, the extensions, updates and replacement with improved data have been significant. Figure 1 gives an overall picture by time and by latitude band of the changes from 4.0 to 5.0. Note that each latitudinal band has a different scaling according to the maximum observation count in each band; by “observation” we mean a monthly average temperature from one station, so the observation count equals the station count for an individual month. There were few changes prior to 1890 so only the period since then is shown.

The CRUTEM4.0 station database ended in 2011, so of course all values since then are new gains for the CRUTEM5.0 database. However, even prior to 2011 there are significant increases in observation counts in all latitude bands, sometimes doubling the number of available values. Sources for some of these are given in Table 2, such as the ECA&D project for Europe that contributes especially to the increases from latitudes 30 to 70°N. The gains have offset some of the previous decrease in observation counts since the 1970s, so that counts now peak in the 1990s or later in some latitude bands. However, not all observations in the station database are actually used in the generation of the gridded CRUTEM temperature anomaly dataset because stations with insufficient data prior to 1990 to estimate their mean and standard deviation are not used. Some of the new acquisitions do not currently meet this requirement (the high proportion of missing values is apparent in dark blue in Figure 1), so the higher station counts do not fully translate into greater gridded coverage (illustrated later).

Some existing CRUTEM4.0 station observations have been replaced by improved estimates in CRUTEM5.0 (e.g. through their replacement by homogenized data obtained from national projects). These changes (labelled 'Different' in Figure 1) are present in all latitude bands except for Antarctica and represent a large proportion of observations in bands 30 to 50°S and 30 to 50°N. The latter arises in part from continual updates to the United States Historical Climatology Network (USHCN) homogenization which changes as data series lengthen (Menne & Williams, 2009), while the former reflects various South American (Table 2) and Australian (Table 1) homogenization initiatives. A few observations have been removed if they had been identified as duplicates or as inhomogeneous (brown in Figure 1).

3 Removal of outlier values

3.1 Introduction and limitations of the CRUTEM4 methods

The process to construct the CRUTEM dataset includes multiple layers of QC to identify and either correct or ignore dubious values. This begins with the QC checking by the originating NMHS, followed by additional checks implemented in various data compilations that we access (see section 2; e.g. Durre et al., 2010). The CLIMAT data, which constitute the main source for the regular monthly updates, are QC'd by the Met Office prior to inclusion in the CRUTEM station database. Common CLIMAT coding errors are first corrected, if detected. Subsequently, an automated check compares each value to neighbouring stations or the mean of daily values from SYNOP reports and flags it for inspection if it differs by more than a threshold amount. Values are also flagged for inspection if they lie outside climatological confidence intervals for that station. Flagged values are then manually inspected and not used in CRUTEM if considered erroneous. However, if a correct value can be confidently determined by inspection of the SYNOP mean daily values, or the mean of the max and min temperatures in the CLIMAT message, or by knowledge of basic coding errors (e.g. a factor of ten error), that is used instead. These stages in QC are unchanged from CRUTEM4.

After compilation of the station database, CRUTEM4 and all earlier versions (summarised in Osborn & Jones, 2014) then applied a simple standard deviation (SD) based check to identify and remove outliers prior to creating the gridded dataset. Monthly temperature values were flagged as outliers if they lay more than 5 SD from the 'normal' value, where the SD and normal (i.e. time mean) were calculated separately for each month of the year and for each station from data during the reference periods 1941–1990 (SD) and 1961–1990 (normal).

In line with the dataset construction principle that methodological changes should be minimised (section 1), this outlier check has remained almost unchanged since at least CRUTEM1 (Osborn & Jones, 2014). However, a count of outlier removals per year ("SD" brown lines in Figure 2) illustrates two limitations of this check. First, almost no outliers are identified during the 1941–1990 period over which the station SD values are calculated. This is despite the fact that an occasional gross outlier has been found to be present in the database during this period (e.g. at three stations in St Kitts, Colombia and Romania). Sensitivity checks showed that in some cases (e.g. where only 15 to 20 values are available to calculate the SD and normal) physically impossible values can pass this test if they occur during this period. Second, there is a clear trend outside the 1941–1990 period with more cold outliers excluded prior to 1941 and more warm outliers excluded after 1990 (and the proportion of warm outliers increases to the present). This behaviour is expected when outliers are identified relative to a fixed normal in the presence of an ongoing warming trend. Although the excluded outliers represent less than 1% of the data values in any one year (and only 0.03% of values overall), the effect will grow as warming continues and already adversely affects some extremely warm months (e.g. June 2003 in Europe: Supplementary Figure 2). A third limitation of the CRUTEM4 outlier check is that if there are insufficient data to compute the SD (or the normal) in any month of the year then the station is entirely discarded. This may throw away usable data that we would prefer to retain.

These behaviours are clear in the very different total number of outliers before, during and after the 1941–1990 period (Figure 3; bars show the SD outlier totals). Only 21 cold and 5 warm outliers are identified in the entire 1941–1990 period; prior to this, there are 2.5 times more cold than warm outliers found (448 cf. 179). After 1990, there are almost five times more warm than cold outliers found (1328 cf. 279).

Revised outlier checks were developed (described in the following sections) for CRUTEM5 to address these three limitations. First, a physical plausibility test is applied to screen out any obvious outliers. As this is applied in all cases, we relax the minimum data requirement for the subsequent outlier check so that we do not discard usable data. Second, we replace the subsequent SD outlier check with one based on the interquartile range (IQR) because this is less sensitive to outliers occurring during the reference period. Finally, the IQR test is relaxed in the presence of regional extremes that affect many neighbouring stations, which also partly addresses the trend towards fewer cold and more warm extremes being excluded.

3.2 Checking for physical plausibility

The aim of this new outlier check is to pick up any very large errors that do not seem physically plausible. It is not intended to be a stringent test because the main outlier check is applied afterwards (section 3.3). The overall range of physically plausible values for monthly-mean temperature depends on multiple factors, but the three most influential factors are month of the year, station latitude and station elevation (Rohde et al., 2013). For each month of the year (m), we compute in each 5° latitude band (j) the median ($\widetilde{N}_{j,m}$) of all station normals ($N_{s,m}$ for each station s) and the median ($\tilde{E}_{j,m}$) of all station elevations (E_s). The median normal by latitude band is illustrated for March by the white line in Figure 4. The deviations of each individual station's normal and elevation from their respective latitudinal-band medians are used in a linear regression to determine an empirical global lapse rate (L):

$$\left(N_{s,m} - \widetilde{N}_{j,m}\right) = L\left(E_s - \tilde{E}_{j,m}\right) + c + r_{s,m} \qquad (1)$$

where c is a constant and r a residual from the regression. The average of the 12 monthly lapse rates is the same as the lapse rate obtained using annual-mean normals. This value (L = -3.91 K/km) is used solely for the physical plausibility outlier check.

The distribution by latitude of the individual station temperature values (Figure 4a shows March as an example) illustrates the spread of values through time and location. Each temperature value ($T_{s,t,m}$ in year t) is then adjusted for latitude and elevation of the station according to:

$$\hat{T}_{s,t,m} = T_{s,t,m} - \widetilde{N}_{j,m} - L\left(E_{s,m} - \tilde{E}_{j,m}\right) \tag{2}$$

This expresses each value relative to an expected norm considering the station latitude, elevation and month of the year. These are shown for March in Figure 4b, illustrating the overall range of these latitude- and elevation-adjusted values. The spread of these values represents the variability (spatial and temporal) of observed temperatures (e.g. it is largest in the mid-to-high latitudes of the winter hemisphere). The results were used to subjectively draw boundaries within which all the physically plausible values are thought to lie. Any values in the existing CRUTEM station database lying outside these boundaries were inspected and the boundaries were made more liberal if there was any doubt that the values might be genuine. The blue lines in Figure 4 show these boundaries for March.

The physical plausibility check is applied to the CRUTEM5 station database to identify values that are implausible. Only 549 values were identified as being outside the physically plausible range (203 too cold, 346 too warm), less than 0.007% of the values checked. All 549 values are excluded from the subsequent analysis (and are flagged in the underlying database). The CRUTEM4 outlier check had previously correctly identified (and thus excluded) many of these, except most of those in the 1941–1990 period.

3.3 Quartile-based thresholds

As noted above and illustrated in Figure 2, the CRUTEM4 SD-based outlier check identified few outliers during the 1941–1990 reference period (0.0008% values flagged as outliers, compared with 0.03% from 1850–1940 and 0.08% from 1991–2018). Outliers present during 1941–1990 inflate the SD. If occurring during 1961–1990, then they also bias the normal towards the outlier value. These effects are particularly large if the number of values used to compute the SD and the normal is relatively small (e.g. 15 or only slightly more). The inflated SD and biased normal increase the chance that the outlier value will lie within 5 SD of the normal. In some test cases, the effect is so limiting on the power of the outlier test that even a value of 1000°C passes the test if it occurs within the 1961–1990 period.

We explored several potential improvements to the SD-based outlier check but none resolved all the issues. We looked at the ratio of each SD value to the SD of other months or of neighbouring stations to identify those that might be inflated by outliers, but no simple criteria that could be applied without manual intervention were identified. The SD used for testing the value in year t could be calculated using all values except the one in year t, but this still failed if there were two outliers in the data sequence for the same month at that station.

Instead, we found that an outlier test based on the IQR provides a more robust test (Tukey, 1977). Outliers were identified as those values lying outside the range (sometimes called the upper and lower "fences"):

$$(\mathrm{LQ} - n\ \mathrm{IQR}) \ \text{ to } \ (\mathrm{UQ} + n\ \mathrm{IQR}) \qquad (3)$$

where LQ is the lower quartile and UQ the upper quartile of the data, IQR = UQ – LQ, and n is a multiplier. The LQ and UQ are calculated for each monthly data sequence at each station from values in the same 1941–1990 period as used previously for the calculation of SD, again requiring a minimum of 15 non-missing values. The quartile and IQR values are more robust to the presence of erroneous values and the IQR-based test is able to identify potential outliers during the 1941–1990 period that the SD-based test let through.

The choice of n is somewhat arbitrary, in the same way as is the choice of 5 SD rather than, say, 4.5 SD, because there is no specific value to separate genuine from erroneous values. Instead it is a balance between discarding too many genuine values and including too many erroneous values. Assuming the previous 5 SD test provides this suitable balance (except during the reference period where the SD test is inadequate), we can select n in the IQR-based test to yield the same number of outliers. For normally distributed data, $n = 3.206$ is equivalent to normal ±5 SD. On testing, this captured considerably more outliers than the 5 SD test did, because the data are not normally distributed (e.g. in many regions, especially Siberia, monthly temperature anomalies are negatively skewed) and the sample SD, normal and quartiles are sometimes poor estimates of their population values. Trialling a range of values for n, the total number of outliers (outside the 1941–1990 period) is closest to the 5 SD test when $n = 3.7$ (Figure 3). The IQR-based test also identifies many outliers during the 1941–1990 period, which the SD test failed to do, including the three cases mentioned earlier.

Although the 3.7 IQR and the 5 SD tests identify a similar total number of outliers, they do not always designate the same values as outliers. In fact only about 50% of outliers are common to both tests. Manual inspection of some cases suggests that the IQR outliers may be closer to what would be considered erroneous values (based on expert judgement or regional clustering). For example, 3.7 IQR designates as outliers far fewer high values in the June 2003 European heatwave than does 5 SD (Supplementary Figure 2). Given the very warm anomalies across this region, many of these may be genuine values rather than outliers. The 3.7 IQR test also slightly reduces the trends in designated outliers compared with the 5 SD test (Figure 2b), though a trend towards more frequent designation of warm outliers is still present.

3.4 Allowance for regional extremes

Inspection of the outliers identified by the 3.7 IQR test indicates that there are cases where many stations in a region have extreme values. In many instances, regional clusters imply that some (or all) of the values designated as outliers may in fact be genuine values (there are some exceptions to this, e.g. if all the stations from one country are mis-reported in a particular month then a regional anomaly can occur which is erroneous despite agreement between neighbouring stations). To address this, the IQR test is modified to take into account the values reported simultaneously at other stations in the vicinity. This also partly addresses the issue of a trend towards more frequent designation of warm outliers as the climate warms, since the climatic warming is expressed at the neighbouring stations too.

This is achieved by modifying the IQR test in eq. (3), replacing n by $(n - fn')$ for the lower fence and by $(n + fn')$ for the upper fence. The strength of the modification is given by parameter f (f = 0 reverts to the standard IQR test), while n' is a regional mean of surrounding station values normalised to IQR units. This normalisation is analogous to the common transformation of subtracting the mean and dividing by SD, but instead a quartile is subtracted and then the division is by the IQR. If the value being tested is below the median temperature for that station, all neighbouring station temperature values are normalised relative to their LQ; otherwise they are all normalised relative to their UQ. The normalised values represent how many IQRs each station value is below or above their relevant quartile. When applying the IQR outlier check to each monthly temperature at a station, the average of the normalised values from the nearest 15 stations is used for n' (though only neighbours within 1200 km are considered, the typical correlation decay length of monthly land air temperatures; Harris et al., 2014).

This regionally-modified IQR-based outlier check was applied to the CRUTEM5 database with f = 0.3, after the removal of values that fell outside the physically plausible ranges. Shifting the fences by the regional normalised values from surrounding stations results in fewer values being labelled as outliers. On the basis that the overall stringency of the CRUTEM4 5 SD outlier check had been considered to give a good balance between keeping bad values versus excluding good values, n was reduced to 3.6 so that the number of outliers (outside the 1941–1990 period) remained close to the number found previously. These choices make no practical difference to large spatial average temperature timeseries, but do affect local temperature anomalies in some months.

Using these parameters, 2389 further values (0.03% of those tested) were flagged as outliers (972 cold, 1417 warm) and excluded from the subsequent analysis. Those values that could not be checked (due to insufficient values to compute the quartiles) are now used because they have passed the new physical plausibility test that removes gross errors (in CRUTEM4 they were excluded). In practice, some will later be excluded because they also have insufficient values to compute a normal. The adjustment for regional extremes has, as intended, reduced the number of designated outliers during some extremely cold (e.g. December 1879, Supplementary Figure 1) or warm (e.g. June 2003, Supplementary Figure 2) events. It has also reduced the trends in outlier counts (Figure 2) for cold outliers prior to 1941 and for warm outliers after 1990, compared with the simple IQR- or SD-based checks. Unlike the SD-based check, it is effective in designating outliers during the 1941–1990 period. However, errors that affect a set of stations in a region may now pass the modified outlier test (such as when a data source provided erroneous August 2015 values for all stations in Turkey, Supplementary Figure 3) and so regional clusters of outliers that were previously flagged but are now let through must be manually checked (the Turkish station values were set to missing for August 2015).

After removal of outliers, the normal (1961–1990 means) and SD (1941–1990) are recalculated using the retained data values.

4 Generating gridded fields of temperature anomalies

4.1 Gridded anomalies using the standard CRUTEM method

The standard CRUTEM5 method used to generate gridded fields of temperature anomalies is the same as used for CRUTEM4, with the details given by Osborn & Jones (2014) and the background to this choice discussed by Jones et al. (2012). This is the climate anomaly method and has two steps: (1) convert the monthly temperatures at each station into anomalies from their 1961–1990 means ("normals"); and (2) use these station anomalies to estimate temperature anomalies on a grid over the land surface of the world. For the second step, the CRUTEM approach is to form the arithmetic mean of any station anomalies that lie within each grid cell of a regular latitude-longitude grid with 5° resolution. Grid cells that do not contain any station anomalies are left missing. An alternative gridding with better high latitude representation is explored in a later section. Unlike some other methods (e.g. Cowtan & Way, 2014; Rohde et al., 2013), neither the standard nor alternative CRUTEM5 gridding utilises estimates of the spatial covariance of temperature anomalies.

The uncertainty model for the gridded temperatures is unchanged from CRUTEM4 (Brohan et al., 2006; Jones et al., 2012; Morice et al., 2012) and so it is not described here. Normals were not calculated for stations with insufficient data to meet our criterion. For CRUTEM4, this criterion had to be met for every month of the year otherwise normals were not calculated for *any* month for that station. This effectively excluded such stations from the creation of the gridded dataset (unless normals were obtained separately), whereas for CRUTEM5 we use stations for any month for which a normal can be calculated. This allows the inclusion of 277 extra stations with partial coverage. After calculation of normals, we adopt the same method as CRUTEM4 to infill some missing normals from World Meteorological Organization (1996) or estimated from different periods and then adjusted to represent the 1961–1990 mean (Jones et al., 2012). The total number of stations with normals and SDs, and thus available for gridding, is 7983, up from 4842 in CRUTEM4.0.

4.2 An alternative gridding method with better representation of high latitude stations

The overall rationale for CRUTEM gridding is that observations contribute to grid cells that they lie within. Thus the covariance between locations further afield, that might be used in kriging, kernel smoothing or covariance-based methods is not utilised (see section 1 for a justification of our choice, including that structural uncertainty is better sampled with each global temperature dataset taking different approaches). In CRUTEM, therefore, a station's influence is not linked to the covariance structure of temperature, but only to its geographical location. Arguably, under such a scheme, each station should contribute the same representation (weight) to the global field and global mean (except of course where we have redundant information from multiple stations in one small area, which gridding is designed to deal with). However, the standard CRUTEM gridding approach causes high latitude stations to be *under-represented* because the longitudinal extent of a grid cell decreases like the cosine of its latitude and each station can only contribute to a single grid cell.

This is a different issue to the potential bias in estimates of global-mean temperature due to non-random incomplete coverage (Cowtan et al., 2018), such as when temperature changes are not estimated over those areas (e.g. the Arctic Ocean) that are warming faster (Simmons & Poli, 2015). Even if a global-mean temperature estimate is not required, the under-representation of high-latitude information can be problematic. For example, a data-model comparison where the simulated data is correctly masked to match the observed data coverage (and hence properly taking into account the incomplete coverage) will nevertheless be biased towards the agreement or disagreement at low latitudes if the high latitudes are under-represented.

An alternative gridding method has been designed that addresses this issue while following as closely as possible the standard CRUTEM gridding. The modification is that a station is allowed to contribute to M adjacent grid cells where $M = 1/\cos(latitude)$ rounded to the nearest whole number and the latitude of the grid cell centre is used. For example, at 72.5°N, M = 3. These M cells are those longitudinally adjacent cells centred most closely on the station's longitude. Each 5° by 5° grid cell temperature anomaly is now the arithmetic mean of any station anomalies that can *contribute to* that grid cell (even if they lie in a neighbouring longitudinal cell when cells are narrow at higher latitudes). Other approaches were considered but all had disadvantages. For example, using a non-regular equal-area grid would be more complex for users familiar with a regular grid, comparing with other datasets on regular grids, or using software designed for regular grids. Allowing a station to contribute to all cells within a fixed longitudinal distance would give more influence to those located near grid cell boundaries. The chosen method is simple, retains the regular grid, and reduces the link between a station's location and its influence on the gridded dataset. The South Pole station is assigned to all grid cells in the southernmost row of the grid.

The outcome of this alternative gridding method is illustrated for some example monthly fields in Figure 5. The benefits of this gridding can be seen visually in the SH polar projection maps: the high latitude coverage is more closely equivalent to that at the equator (right column) compared with the slim grid cells of the standard gridding (left column). The geographical structure of circum-Arctic temperature anomalies is also much clearer, whether it is for the more uniformly warm case of August 2016 or for the strong gradient between a very cold European sector and very warm at other longitudes in February 1963. The effect of gridding on global-mean land air temperature is considered in section 5. Although the alternative gridding method addresses the under-representation of high-latitude temperature anomalies, it is not intended to supplant the standard CRUTEM gridded dataset because the CRUTEM uncertainty model applies to that gridding method.

4.3 Generating global-mean temperature timeseries

Global and hemispheric mean timeseries are calculated using the same method as for CRUTEM4 (Jones et al., 2012; Osborn & Jones, 2014). Hemispheric series are computed as the area-weighted mean of grid cell temperature anomalies, requiring a minimum of five grid cells. The global series is then computed as (2 NH + SH) / 3, reflecting the relative land areas in each hemisphere. The requirement for at least five grid cells in a hemisphere currently restricts the SH and global series to begin in January 1857, whereas the NH series covers our entire study period from January 1850 to the present. The SH records available in 1857 provide sampling of four different regions (South America, South Africa, SE Australia and New Zealand) but all are at similar latitudes (between 26 and 38°S).

The uncertainty model for the global and hemispheric temperature anomaly timeseries is almost unchanged from CRUTEM4 (Brohan et al., 2006; Morice et al., 2012), grouped into four components. (1) Uncertainty in grid cell temperature anomalies that is uncorrelated between grid cells (e.g. due to measurement error or incomplete sampling of a grid cell). (2) Uncertain biases associated with residual homogenisation error and uncertainty in climatological normals, which are systematic for individual stations but independent between stations. (3) Systematic biases that are correlated between grid cells and persistent in time (e.g. urbanization or exposure changes). (4) Coverage uncertainty due to incomplete sampling of the land surface in each hemisphere. These components are combined into overall confidence intervals. The only change for CRUTEM5 is that the coverage uncertainty, which is estimated by subsampling a spatially complete dataset, is now based on the European Centre for Medium-Range Weather Forecasts reanalysis version 5 (ERA5; Hersbach et al., submitted). A previous implementation error has also been corrected (the exposure and urbanisation biases are now correctly treated as independent, adding them in quadrature), resulting in slightly narrower confidence intervals for CRUTEM5 than for CRUTEM4. Note that for HadCRUT4 (Morice et al., 2012), the combined land and marine global temperature dataset, the same underlying error model is used to generate an ensemble of realizations rather than the central estimate and confidence intervals reported here.

5 Analysis of CRUTEM5

5.1 Comparing CRUTEM4 and CRUTEM5

We consider the expansion and improvements to the station database separately from the improved algorithms for identifying and removing outliers, by first calculating the global-mean temperature anomalies using the CRUTEM4 methods but with the updated station database (Figure 6). The significant expansion in the station database (from 5583 to 10639 stations) led to a 65% increase in the number of stations actually used (from 4842 to 7983, i.e. after application of outlier checks and removal of stations without normals or SD). The count of individual monthly station temperature anomalies increased by 57%. The majority of this expansion had already been incorporated into version CRUTEM4.6 first released in 2017 (compare brown and black lines in Figure 6). Increases in observation counts are particularly large from the 1960s to the present, though even the period 1880–1950 shows a useful increase. The increase from the CRUTEM4.6 to 5.0 station databases is mostly in the 2017–2019 period, with modest increases prior to that. The station observation counts peak in the 1970s and decrease by about 25% by the 2000s. The underlying station database (Figure 1) already includes data that could address this decline, but these extra data for the 1980s to 2000s are from stations without 1961–1990 normals so they are not used with the current CRUTEM methods.

Despite the large increase in station counts, the coverage of grid cells with temperature anomalies is only moderately expanded (by about 10% from CRUTEM4.0 to 5.0, with most of this increase already achieved by version 4.6; Figure 6). This is because most of the station acquisitions are in already-sampled regions. Nevertheless, this extra sampling improves the estimates in those regions and will reduce their uncertainty, as well as providing about 10% extra coverage. The inclusion of more nationally-homogenised data (section 2) will also improve the reliability of regional temperature anomaly estimates, though this is not measured by the station or grid cell observation counts.

Turning to the global-mean land temperature anomalies themselves (Figure 6, upper and middle), we find that the station database expansion has little effect. This is expected because prior work has shown that global estimates are robust and can be estimated from a relatively small number of observations. There are some differences as large as 0.1°C in the early decades when coverage is poor (with CRUTEM5.0 often cooler than 4.0), the difference peaking around 1870 and again in 1885 (pink line in upper-right panel of Figure 6). In the recent period (middle row), station database updates tend to raise global estimates by up to 0.05°C in the final couple of years. This is because the monthly updates are biased relatively low in regions such as China where the CLIMAT data are inconsistent with our preceding series based on the mean of Tmin and Tmax; the less frequent updates then correct this bias by replacing the values with those estimated more consistently (see section 2).

The modification of methods (improved outlier identification and allowing stations to be used for any months with normals, even if they do not have normals for all 12 months) affects the global-mean land temperature series even less (Figure 7). This is expected because these modifications were intended to improve local estimates during some extreme events rather than to have a global-mean effect (also note that this figure shows 12-month running means rather than individual months). The changes give a slight improvement in coverage (1.6% increase in station observation counts and 0.4% increase in grid cell observation counts), but changes in global land annual anomalies are less than 0.01°C except in the early part of the record.

The impact of the change in outlier identification is apparent for some individual regional extreme events, such as December 1879, June 2003 and August 2015 (right-hand columns of Supplementary Figures 1–3). Some grid cell anomaly estimates for June 2003 are more than 0.5°C warmer with the regionally-modified IQR-based outlier check compared with the old SD-based outlier check, and a central European average of 15 grid cells is 0.16°C warmer. Such differences can be important when quantifying the increased risk of such events attributable to human-induced climate change (Stott et al., 2004). The impacts of the changes to the station database and the outlier identification method are more apparent at regional scales than at the hemispheric and global scales, where they are negligible. Timeseries of continental and sub-continental average temperature anomalies are shown in Supplementary Figures 4 to 10 and include a comparison of the CRUTEM4.6 and CRUTEM5.0 results.

5.2 Comparing standard and high-latitude gridding

That the alternative gridding (section 4.2) provides more uniform representation of stations regardless of their latitude has already been shown for two individual months (Figure 5) and four more examples are given in Supplementary Figures 11 and 12. A good illustration is August 2016 (Figure 5): single stations at St Helena in the South Atlantic (16°S) and at Halley on the Antarctic coast (75.5°S) provide very different coverage (and hence contributions to any area-weighted analysis) with the standard gridding but much more similar coverage with the alternative gridding.

The alternative gridding increases the estimated global land warming by about 0.1°C over the course of the whole record (top-right panel of Figure 8), with about half of that additional estimated warming occurring since 2000. This places the global series diagnosed from the alternative gridding near the upper edge of the 95% confidence interval from the standard gridding result during the last decade (Figure 8). The greater warming estimated with the alternative gridding arises from the NH series (in fact the overall estimated warming is reduced by ~0.05°C in the SH since 1975), as expected because the longitudinally-slim high latitude grid cells under-represent the northern polar stations with standard gridding, and this is where temperature has increased the most (Simmons & Poli, 2015). With the alternative gridding, pre-1890 values are about 0.04°C lower and the warming trends from 1910 to 1940 and from 1990 to present are slightly enhanced.

With the standard gridding, the overall warming from the 1861–1900 mean to the mean of the last 5 years is estimated to be 1.6°C (with a 95% confidence interval on individual annual means of -0.11 to +0.10°C in the recent period). With the alternative gridding it is 1.7°C, while with CRUTEM4.6 it was 1.6°C. Given that the underlying station database is the same for both gridding methods and that the modification to the gridding is relatively minor, the errors in global-mean values are likely to be quite similar. However, the errors of adjacent high latitude grid cells will be more strongly correlated with the alternative gridding because a station can now contribute to multiple grid cells, and the coverage error will be affected by the greater number of grid cells with estimates of temperature anomalies (bottom-right of Figure 8). Therefore, the CRUTEM error model does not apply directly to the alternative gridding, and for this reason the standard gridding version of CRUTEM5.0 will remain as the preferred dataset.

An important point to make is that the alternative high-latitude gridding introduced here is not intended to address the broader issue of incomplete spatial coverage due to lack of observations in some regions. The biases introduced in an estimate of the full global-mean warming by not sampling a rapidly warming region such as the Arctic Ocean (Cowtan et al., 2018) are better addressed by other approaches such as with reanalyses or making spatially-more-complete estimates and require consideration of the land, ice-free and ice-covered oceans together. As such, this land-only paper is not an appropriate place to investigate this, but it is addressed in the new HadCRUT5 dataset (Morice et al., submitted) formed by combining CRUTEM5.0 and HadSST4.0 (Kennedy et al., 2019).

The alternative gridding version of CRUTEM5.0, with better representation of high-latitude data, could be useful for (e.g.) model-observation comparisons where the model data are masked to match the coverage of the observation dataset. With the standard gridding, this mask will unduly limit the high latitude area retained and might bias the model-observation comparison to the lower latitude areas (this is obvious from Figure 5). Masking and comparing with the alternative gridding would reduce this problem.

5.3 Comparing CRUTEM5 with other land air temperature datasets

The two versions of CRUTEM5.0 (standard and alternative gridding) show close agreement with other land air temperature datasets at the global scale. Figure 9 compares these series with global-mean land series from GISTEMP (NASA Goddard Institute for Space Science; Lenssen et al., 2019), NOAAGlobalTemp V5 (Zhang et al., 2019), Berkeley Earth

(Rohde et al., 2013) and ERA5. Each annual series is very highly correlated (*r* > 0.98 for all series, > 0.99 for the two CRUTEM5.0 series) with the mean of the other series. The root-mean-squared difference between each annual series and the mean of the other series is between 0.05 and 0.12°C (for CRUTEM5.0 it is 0.08°C with standard gridding and 0.06°C with alternative gridding).

These small differences are comparable to the estimated one sigma uncertainties of the CRUTEM5.0 annual-mean values (which are smaller than ±0.2°C since 1870 and then smaller than ±0.1°C since 1930). However, there are also some small systematic differences visible in the intercomparison (Figure 9). CRUTEM5.0 with standard gridding tends to lie at the bottom of the group of series since 2000, Berkeley Earth tends to lie at the top of the group in most years since 1940, with the other series lying more centrally within the spread of results. Some of these differences likely arise from spatial coverage and masking to a common geographical region reduces them (not shown here), or to the different methods of calculating the global mean provided by each group.

6 Conclusions

In this paper we have detailed the further development of the CRUTEM global land air temperature dataset and present the new version CRUTEM5.0. The key aspects of this work and its implications for our knowledge of regional and global temperature change over the land surfaces of the Earth are as follows:

1. The underlying work to strengthen the CRUTEM station database is important because it allows us to benefit from improved availability of station observations and from better assessments of their long-term homogeneity. Also, data coverage could gradually decrease if only monthly CLIMAT updates are used because some stations close or stop reporting through the routine compilations; with our continued non-routine work, we are able to incorporate new stations in their place. We note, however, that there is a growing body of stations that we are not currently using to generate the gridded dataset because they do not have sufficient data to calculate their normals (compare station counts in Figures 1 and 6). This will need to be addressed with methodological changes in future versions.

2. Compared with CRUTEM4.0, the CRUTEM5.0 station database is expanded in terms of station numbers (by 91% in total, and by 65% in terms of those that can be used in the gridded dataset), expanded in terms of monthly observation counts (by 59%, though part of this increase is because the dataset now runs to 2019; for 1850–2010, the expansion is 49%). Alongside this expansion, many values have been replaced (yellow in Figure 1) with the products of improved national homogeneity exercises.

3. Most of the data acquisitions are in already-sampled regions, where they improve the temperature estimates and reduce their uncertainty. Despite the large increase in station counts, the coverage of grid cells with temperature anomalies is only moderately expanded (for 1850–2010 there are 9% more grid cell temperature anomalies in CRUTEM5.0 than in 4.0).

4. Improved outlier checking has been applied to the updated station database, providing better removal of physically implausible values especially during the reference period, retention of some extreme values when they occur in regional clusters, and reducing the trends in outlier

removal that arise from the underlying climatic warming. In future work we may utilise spatially interpolated grids (Morice et al., submitted) to identify outliers relative to regional information or relative to a time-evolving climatology. This could more completely address the issue of a warming climate causing high extremes to be more frequently mis-classified as outliers.

5. The mean temperature timeseries for global land is refined but not significantly changed. This is because global land temperature estimates are already quite robust to changes in datasets and across datasets. Uncertainty in the global series would be reduced most by acquiring stations in unsampled regions rather than more in well-sampled regions, and by further evaluation of the biases related to changes in exposure in the 19th century (see discussion in Jones, 2016).

6. A caveat to the previous conclusion is that it is the mean temperature of the global *sampled-area* that appears to be robust. Estimates of the full global-mean land temperature *including the unsampled areas* may be less robust and can also be biased when calculated as the mean of the sampled region, though bias has been more clearly demonstrated for the global land and marine temperature (Cowtan et al., 2018) rather than land-only. Bias can arise if temperature changes are very different between sampled and unsampled regions. This is especially the case for the sea-ice region of the Arctic Ocean, but the strong warming of the circum-Arctic land also needs to be properly sampled to reduce bias in the global-mean land temperature. We partly mitigated this bias previously (from CRUTEM3 to CRUTEM4; Jones et al., 2012) by expending effort to acquire previously unused data from the high northern latitudes. We further mitigate it here by providing a second estimate based on an alternative gridding method which removes the under-representation of high latitude stations: this increases our estimates of global-mean land warming by about 0.1°C. Linear trends (°C/decade) over the last 40 years (1980–2019) are 0.28 (0.30) globally, 0.34 (0.37) for the northern hemisphere and 0.17 (0.17) for the southern hemisphere using the standard (or alternative) gridding.

7. Related to the previous paragraph, many analyses (e.g. comparisons of models with observations) of this and other global land temperature datasets should ideally focus on the observed region. Infilling via various statistical estimators is best considered in the combined land and marine context (see Morice et al., submitted) rather than here, not least because the outcome is sensitive to the choice of estimating water or air temperature anomalies in sea ice regions. Nevertheless, infilling does not solve the issue with unobserved regions, and a common structural error in all datasets is the lack of observations from Antarctica and the continental interiors of Africa and South America and some parts of tropical/subtropical Asia prior to the 1950s.

Acknowledgments

The authors thank all the scientists and NMHS who have collected and made available the monthly mean temperature data referred to in Tables 1 and 2. TJO, DHL and ICH were supported by UK NERC (grant number NE/N006348/1 SMURPHS and NE/S015582/1 GloSAT) and UK National Centre for Atmospheric Science. CPM and IRS were supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. We thank the three reviewers and associate editor for providing suggestions to improve this paper. We acknowledge John McLean for identifying four outlier values in the CRUTEM4 station database. We thank these organizations for making available the other land air temperature datasets: NASA GISS for GISTEMP, NOAA for NOAAGlobalTemp V5, Berkeley Earth and ECMWF for ERA5 (Generated using Copernicus Climate Change Service information 2019) and the KNMI Climate Explorer for providing access to ERA5 data.

Data availability statement

The underlying station database, the gridded temperature anomalies, the global and hemispheric timeseries and their uncertainty intervals will be available from the Met Office website (i.e. an update to the current CRUTEM4 webpages https://www.metoffice.gov.uk/hadobs/crutem4/), the CRU website https://crudata.uea.ac.uk/cru/data/temperature/) and via a Google Earth interface (https://crudata.uea.ac.uk/cru/data/crutem/ge/). In addition, the dataset will be deposited with the CEDA repository for long-term preservation and re-use (https://catalogue.ceda.ac.uk/uuid/eeabb5e1ff2140f48e76ea1ffda6bb48, doi to be provided prior to publication).

Other datasets were obtained from:

CLIMAT (after QC by the Met Office): http://hadobs.metoffice.com/crutem4/data/climat_summary/

MCDW: https://www.ncei.noaa.gov/data/monthly-climatological-data-of-the-world/access/

GISTEMP land: https://data.giss.nasa.gov/gistemp/graphs_v4/graph_data/Temperature_Anomalies_over_Land_and_over_Ocean/graph.csv

NOAAGlobalTemp V5 land: https://www.ncei.noaa.gov/data/noaa-global-surface-temperature/v5/access/timeseries/aravg.mon.land.90S.90N.v5.0.0.201911.asc

Berkeley Earth land: http://berkeleyearth.lbl.gov/auto/Global/Complete_TAVG_complete.txt

ERA5 land: https://climexp.knmi.nl/selectfield_rea.cgi

References

Ashcroft, L., Karoly, D. J., & Gergis, J. (2014). Southeastern Australian climate variability 1860-2009: a multivariate analysis. *International Journal of Climatology*, *34*(6), 1928–1944. https://doi.org/10.1002/joc.3812

van den Besselaar, E. J. M., van der Schrier, G., Cornes, R. C., Iqbal, A. S., Klein Tank, A. M. G., Besselaar, E. J. M. van den, et al. (2017). SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia. *Journal of Climate*, *30*(14), 5151–5165. https://doi.org/10.1175/JCLI-D-16-0575.1

Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., & Rondanelli, R. (2018). CR2MET: A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile. In *EGU General Assembly*. Retrieved from https://ui.adsabs.harvard.edu/abs/2018EGUGA..2019739B/abstract

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes : a new dataset from 1850. *Journal of Geophysical Research*, *111*(D12), D12106. https://doi.org/10.1029/2005JD006548

Brunet, M., SaladiÉ, O., Jones, P. D., SigrÓ, J., Aguilar, E., Moberg, A., et al. (2006). The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850–2003). *International Journal of Climatology*, *26*(13), 1777–1802. https://doi.org/10.1002/joc.1338

Bulygina, O. N., & Razuvaev, V. N. (2012). *Daily Temperature and Precipitation Data for 518 Russian Meteorological Stations (1881 - 2010)*. Oak Ridge, TN (United States). https://doi.org/10.3334/CDIAC/cli.100

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. https://doi.org/10.1002/qj.2297

Cowtan, K., Jacobs, P., Thorne, P., & Wilkinson, R. (2018). Statistical analysis of coverage error in simple global temperature estimators. *Dynamics and Statistics of the Climate System*, *3*(1). https://doi.org/10.1093/climsys/dzy003

Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., & Vose, R. S. (2010). Comprehensive Automated Quality Assurance of Daily Surface Observations. *Journal of Applied Meteorology and Climatology*, *49*(8), 1615–1633. https://doi.org/10.1175/2010JAMC2375.1

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3). https://doi.org/10.1002/joc.3711

Hersbach, H. Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., et al. (submitted) The ERA5 global reanalysis. Submitted to *Q. J. Roy. Meteorol. Soc.*

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., et al. (2017). Identifying, attributing, and overcoming common data quality issues of manned station observations. *International Journal of Climatology*, *37*(11), 4131–4145. https://doi.org/10.1002/joc.5037

Jones, P. D. (2016). The reliability of global and hemispheric surface temperature records. *Advances in Atmospheric Sciences*, *33*(3), 269–282. https://doi.org/10.1007/s00376-015-5194-4

Jones, P. D., Lister, D. H. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. P. (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, *117*(D5), D05127. https://doi.org/10.1029/2011JD017139

Jones, P. D., Harpham, C., Harris, I., Goodess, C. M., Burton, A., Centella-Artola, A., et al. (2016). Long-term trends in precipitation and temperature across the Caribbean. *International Journal of Climatology*, *36*(9), 3314–3333. https://doi.org/10.1002/joc.4557

Jovanovic, B., Braganza, K., Collins, D., & Jones, D. (2012). Climate variations and change evident in high-quality climate data for Australia's Antarctic and remote island weather stations. *Australian Meteorological and Oceanographic Journal*, *62*, 247–261.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (2011a). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research*, *116*(D14), D14103. https://doi.org/10.1029/2010JD015218

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., & Saunby, M. (2011b). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research*, *116*(D14), D14104. https://doi.org/10.1029/2010JD015220

Kennedy, J. J., Rayner, N. A., Atkinson, C. P., & Killick, R. E. (2019). An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set. *Journal of Geophysical Research: Atmospheres*, *124*(14), 7719–7763. https://doi.org/10.1029/2018JD029867

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522

Menne, M. J., & Williams, C. N. (2009). Homogenization of Temperature Series via Pairwise Comparisons. *Journal of Climate*, *22*(7), 1700–1717. https://doi.org/10.1175/2008JCLI2263.1

Menne, M. J., Williams, C. N., & Vose, R. S. (2009). The U.S. historical climatology network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, *90*(7), 993–1007.

https://doi.org/10.1175/2008BAMS2613.1

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., Lawrimore, J. H., Menne, M. J., et al. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *Journal of Climate*, *31*(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, *117*(D8), n/a-n/a. https://doi.org/10.1029/2011JD017187

Morice, C. P., et al. An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. Submitted to *Journal of Geophysical Research*.

Mullan, B., Salinger, J., Renwick, J., & Wratt, D. (2018). Comment on "A Reanalysis of Long-Term Surface Air Temperature Trends in New Zealand." *Environmental Modeling & Assessment*, *23*(3), 249–262. https://doi.org/10.1007/s10666-018-9606-6

Osborn, T. J., & Jones, P. D. (2014). The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth. *Earth System Science Data*, *6*(1), 61–68. https://doi.org/10.5194/essd-6-61-2014

Penalba, O. C., Rivera, J. A., & Pántano, V. C. (2014). The CLARIS LPB database: constructing a long-term daily hydro-meteorological dataset for La Plata Basin, Southern South America. *Geoscience Data Journal*, *1*(1), 20–29. https://doi.org/10.1002/gdj3.7

Rennie, J. J., Lawrimore, J. H., Gleason, B. E., Thorne, P. W., Morice, C. P., Menne, M. J., et al. (2014). The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. *Geoscience Data Journal*, *1*(2), 75–102. https://doi.org/10.1002/gdj3.8

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., & Mosher, S. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, *01*(02). https://doi.org/10.4172/2327-4581.1000103

van der Schrier, G., van Ulden, A., & van Oldenborgh, G. J. (2011). The construction of a Central Netherlands temperature. *Climate of the Past*, *7*(2), 527–542. https://doi.org/10.5194/cp-7-527-2011

van der Schrier, G., van den Besselaar, E. J. M., Klein Tank, A. M. G., & Verver, G. (2013). Monitoring European average temperature based on the E-OBS gridded data set. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5120–5135. https://doi.org/10.1002/jgrd.50444

Simmons, A. J., & Poli, P. (2015). Arctic warming in ERA-Interim and other analyses. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1147–1162. https://doi.org/10.1002/qj.2422

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, *432*(7017), 610–614. https://doi.org/10.1038/nature03089

Sun, Y., Zhang, X., Ren, G., Zwiers, F. W., & Hu, T. (2016). Contribution of urbanization to warming in China. *Nature Climate Change*, *6*(7), 706–709. https://doi.org/10.1038/nclimate2956

Thorne, P. W., Lanzante, J. R., Peterson, T. C., Seidel, D. J., & Shine, K. P. (2011). Tropospheric temperature trends: history of an ongoing controversy. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), 66–88. https://doi.org/10.1002/wcc.80

Trewin, B. (2018). *The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2*.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesely.

Vincent, L. A., Wang, X. L., Milewska, E. J., Wan, H., Yang, F., & Swail, V. (2012). A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research: Atmospheres*, *117*(D18), n/a-n/a. https://doi.org/10.1029/2012JD017859

Wang, F., Ge, Q., Wang, S., Li, Q., & Jones, P. D. (2015). A New Estimation of Urbanization's Contribution to the Warming Trend in China. *Journal of Climate*, *28*(22), 8923–8938. https://doi.org/10.1175/JCLI-D-14-00427.1

Wickham, C., Rohde, R., Muller, R., Wurtele, J., Curry, J., Groom, D., et al. (2013). Influence of Urban Heating on the Global Temperature Land Average using Rural Sites Identified from MODIS Classifications. *Geoinformatics & Geostatistics: An Overview*, *01*(02). https://doi.org/10.4172/2327-4581.1000104

World Meteorological Organization. (1996). *Climatological Normals (CLINO) for the period 1961–1990*. Geneva, Switzerland.

Xu, W., Li, Q., Wang, X. L., Yang, S., Cao, L., & Feng, Y. (2013). Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *Journal of Geophysical Research: Atmospheres*, *118*(17), 9708–9720. https://doi.org/10.1002/jgrd.50791

Xu, W., Li, Q., Jones, P. D., Wang, X. L., Trewin, B., Yang, S., et al. (2018). A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dynamics*, *50*(7–8), 2513–2536. https://doi.org/10.1007/s00382-017-3755-1

Zhang, H.-M., Lawrimore, J. H., Huang, B., Menne, M. J., Yin, X., Sánchez-Lugo, A., et al. (2019). The NOAA Global Surface Temperature Dataset Version 5: Employing Improved Data Coverage and Historical Homogenization. *Eos*.

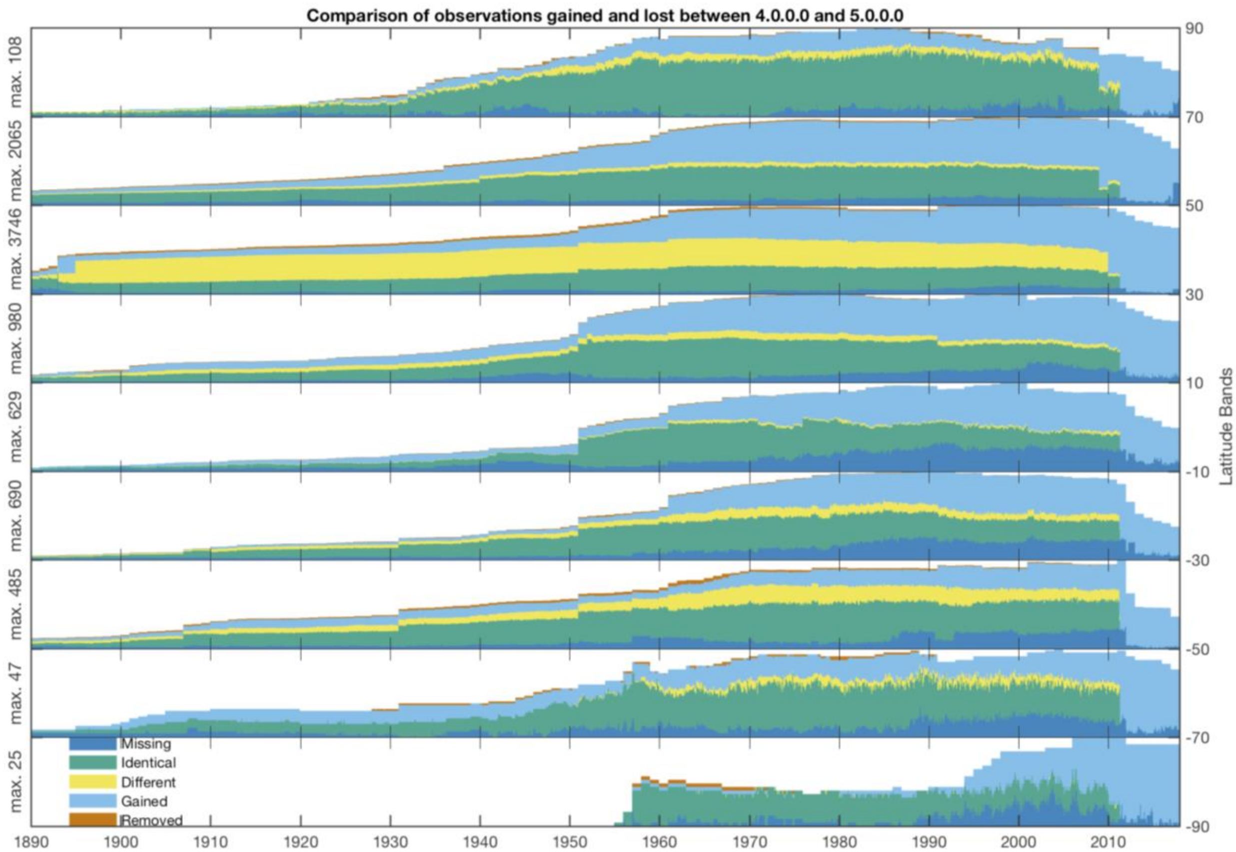


Figure 1. Station counts (i.e. number of monthly temperature observations) by time and latitude band, showing changes from CRUTEM4.0 to CRUTEM5.0 station databases. Values that are recorded for the same station in both databases are yellow if they differ or green if they are unchanged; those present only in CRUTEM4.0 are brown and those present only in CRUTEM5.0 are pale blue. Missing values that lie within a station's overall period of record are dark blue. Counts are shown in 20° latitude bands and the vertical axis of each band covers the range from zero to the maximum station count (indicated on the left-hand axis) in that band.

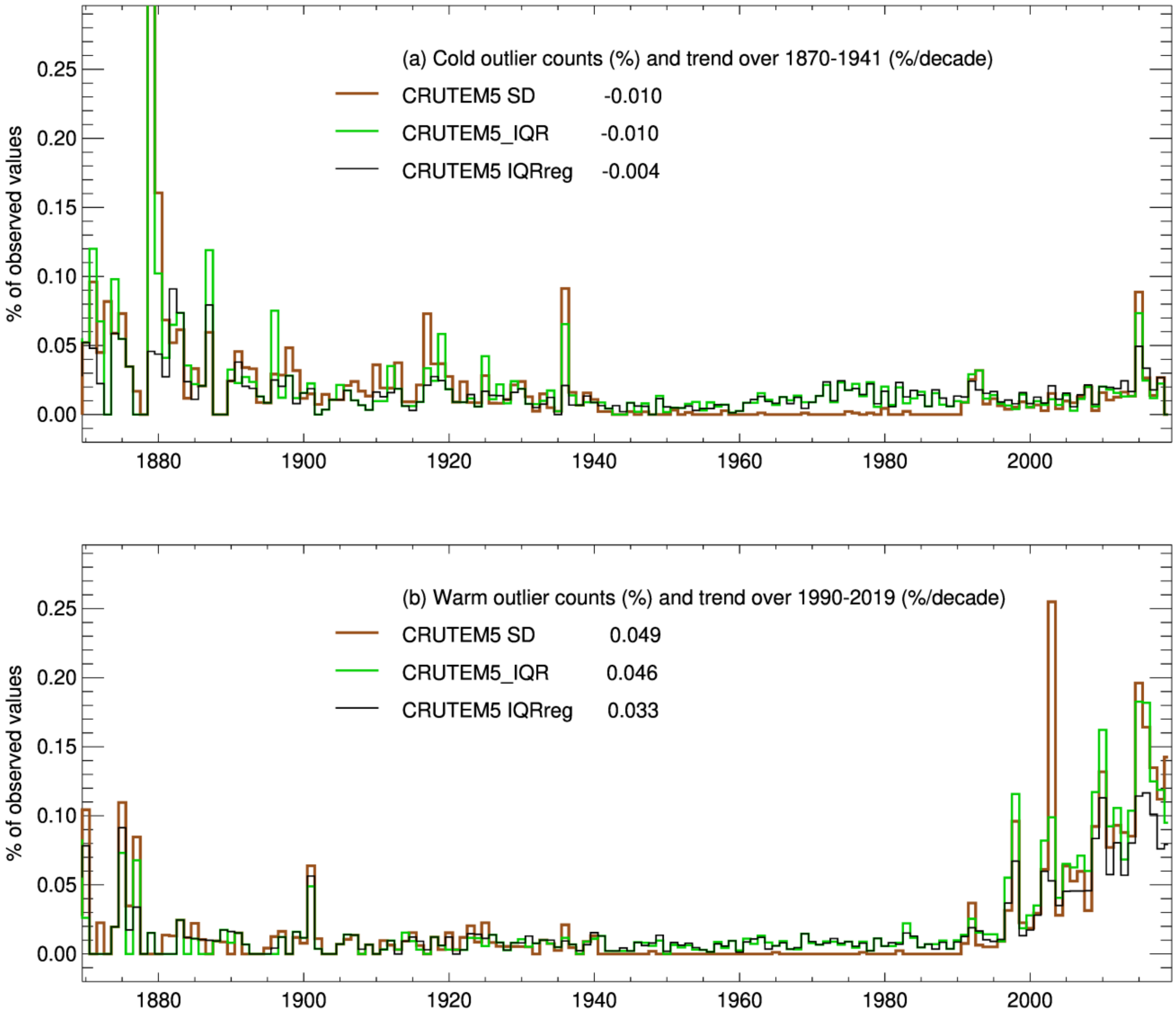


Figure 2. Timeseries of candidate (a) cold and (b) warm outlier counts obtained using the standard deviation (SD) method (brown), the interquartile range (IQR) method (green) and the IQR method with modification of the fences to account for regionally-coherent anomalies (black). Values show the percentage of each year's observations that are flagged as outliers from 1870 to 2019. Legends show the trends in outlier counts (%/decade) from each method during 1870-1941 for cold outliers and during 1990-2019 for warm outliers.

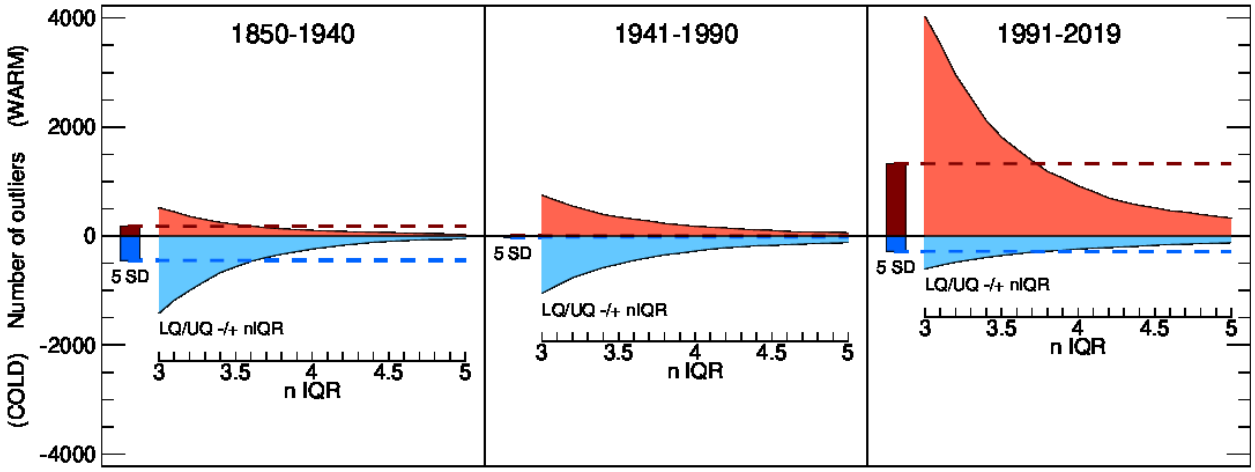


Figure 3. Outlier counts prior to (1850-1940), during (1941-1990) and after (1991-2019) the reference period used to define the outlier test parameters for the standard deviation (SD) method (bars and horizontal dashed lines) and the interquartile range (IQR) method (curved regions) as a function of the strictness of the IQR test (n in equation 3). Warm outlier counts are positive (red), cold outlier counts are negative (blue).

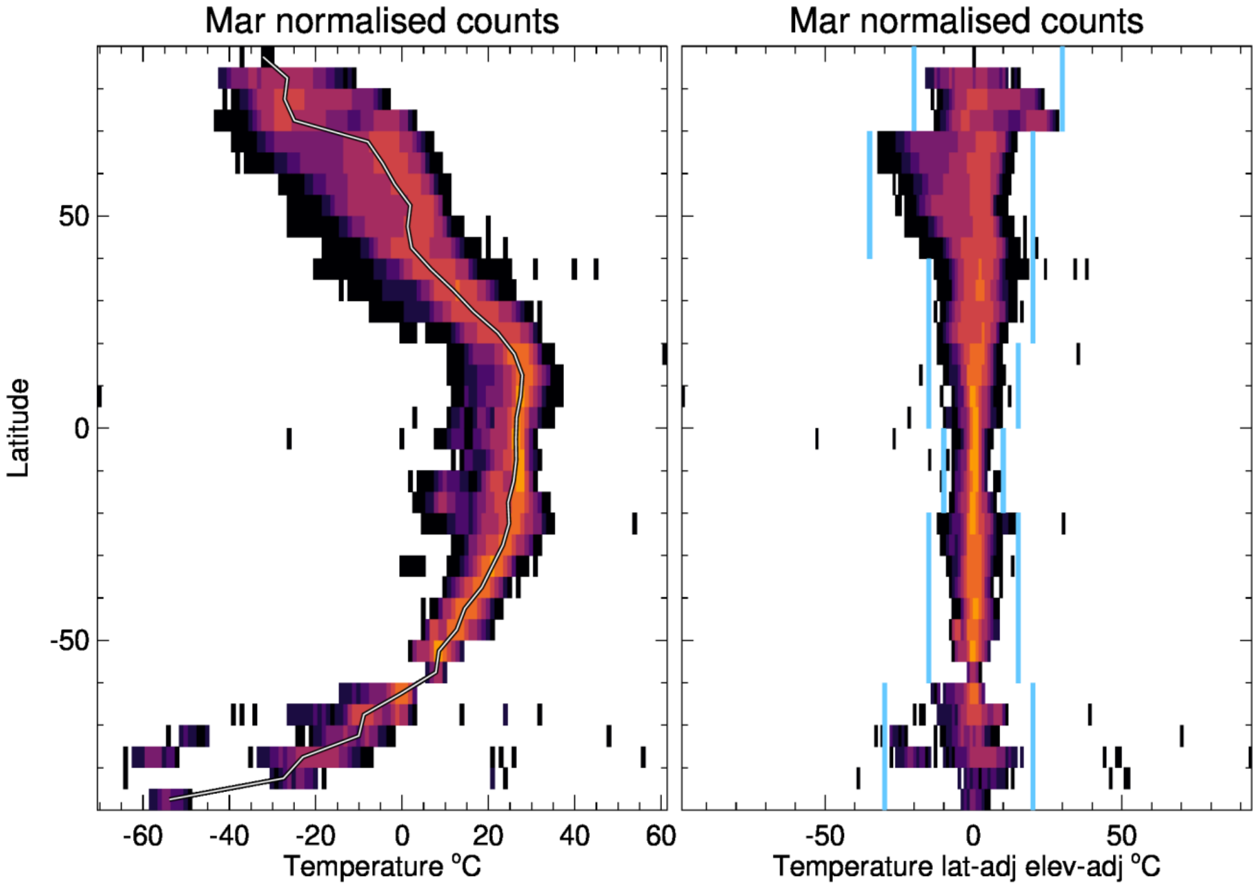


Figure 4. Left panel: heat map of all station monthly temperatures (°C) in the CRUTEM5 database as a function of latitude for March. Brighter (darker) colours indicate more (less) frequent values. The median of the March temperature normals (1961–1990 means) for each 5° latitude band as a white line. Right panel shows: heat map for the same station data, but after adjustment for station latitude and elevation, together with the ranges (vertical blue lines) used to identify physically implausible values (those lying outside these ranges).

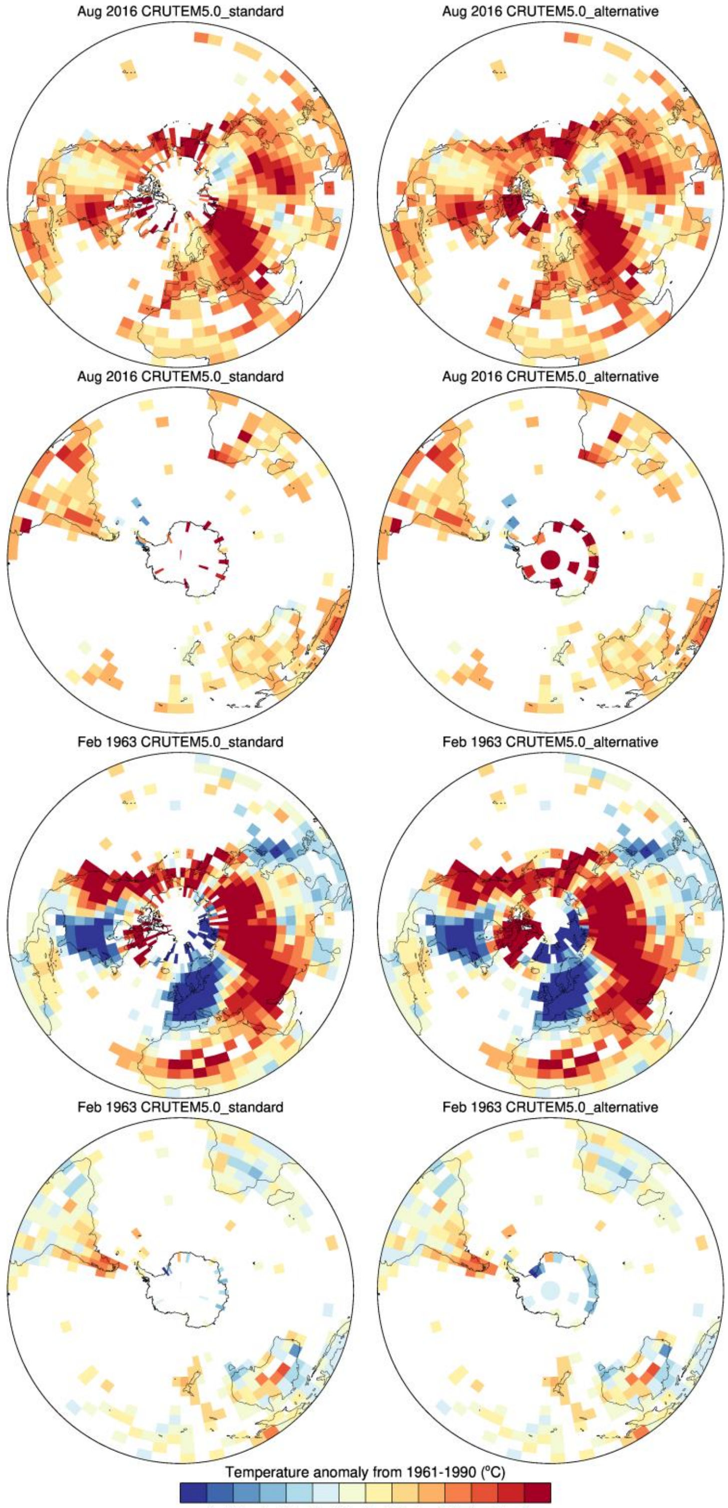


Figure 5. Gridded temperature anomaly (°C relative to the 1961-1990 mean) maps for two example months (August 2016 top, February 1963 bottom) for standard (left) and alternative (right) gridding.

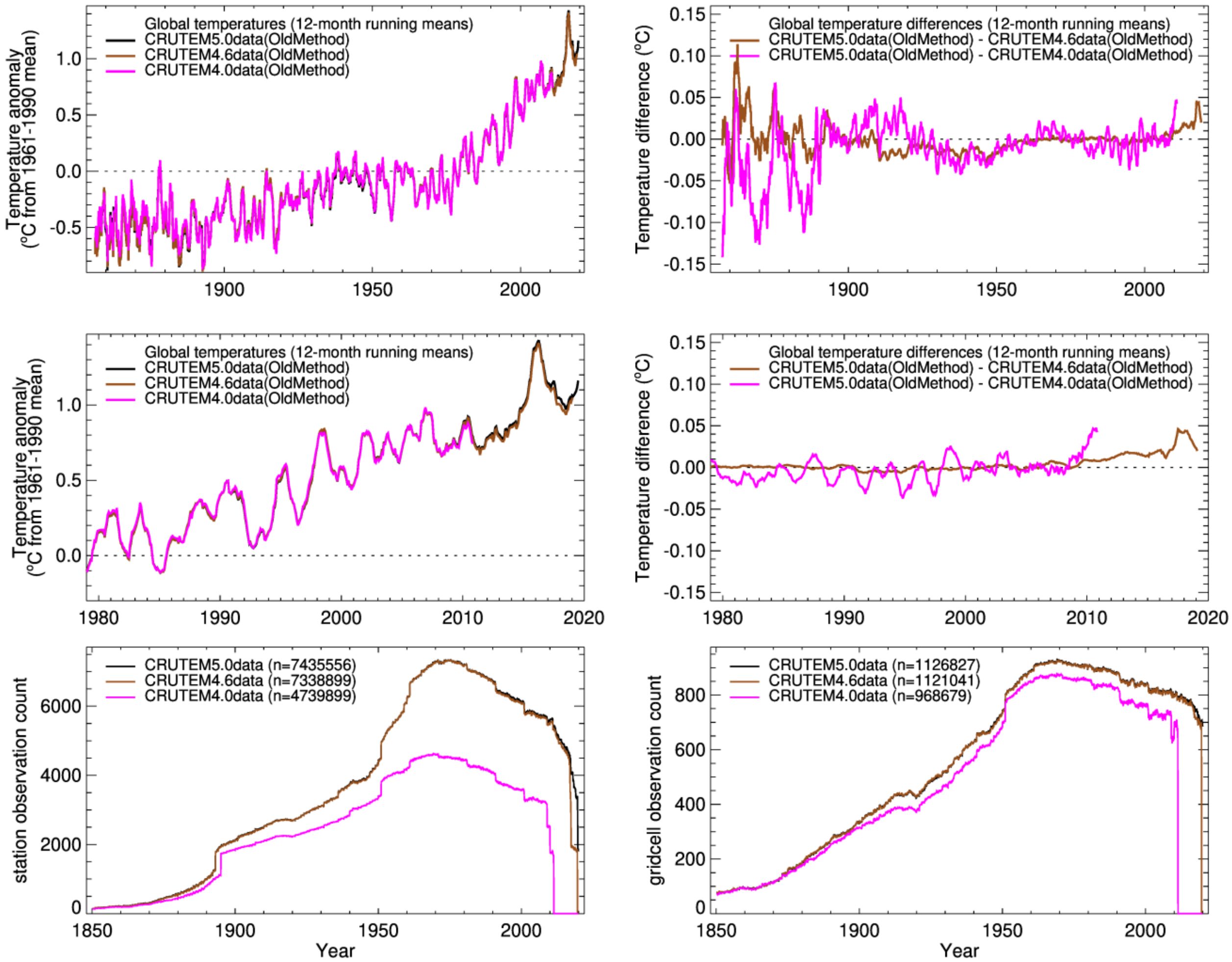


Figure 6. Comparison of global-mean land temperature series from CRUTEM4.0 (pink), CRUTEM4.6 (brown) and CRUTEM5.0 (black) station databases and the same construction methods (the CRUTEM4 methods). Top: 12-month running mean temperature anomalies (°C from the 1961-1990 mean) for each series (left) and their differences (right). Middle: as top but for the period since 1979. Bottom: timeseries of counts for stations (left) and grid cells (right) containing data, with total monthly observations indicated in the legends. Observation counts are after the removal of outliers and stations without normals. Note that the black lines are obscured by the brown lines where the values are close.

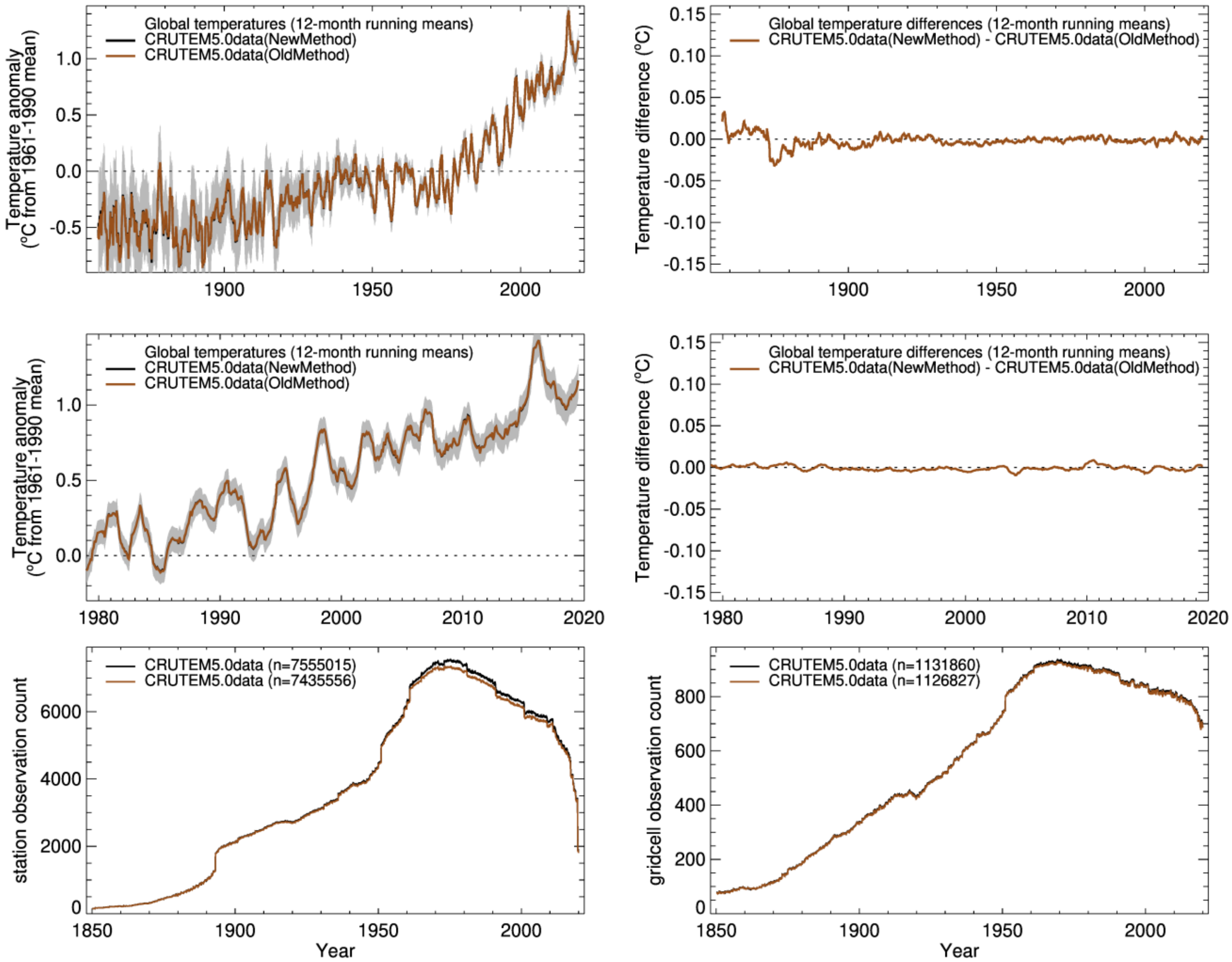


Figure 7. As **Figure 6** except comparing global-mean temperature series and observation counts from the CRUTEM5.0 station database using outlier checking and normal requirements from CRUTEM4 (brown: "OldMethod") or CRUTEM5 (black: "NewMethod"). The grey shading is the 95% confidence interval for CRUTEM5.0 data with CRUTEM5 methods. Note that the black lines are obscured by the brown lines where the values are close.

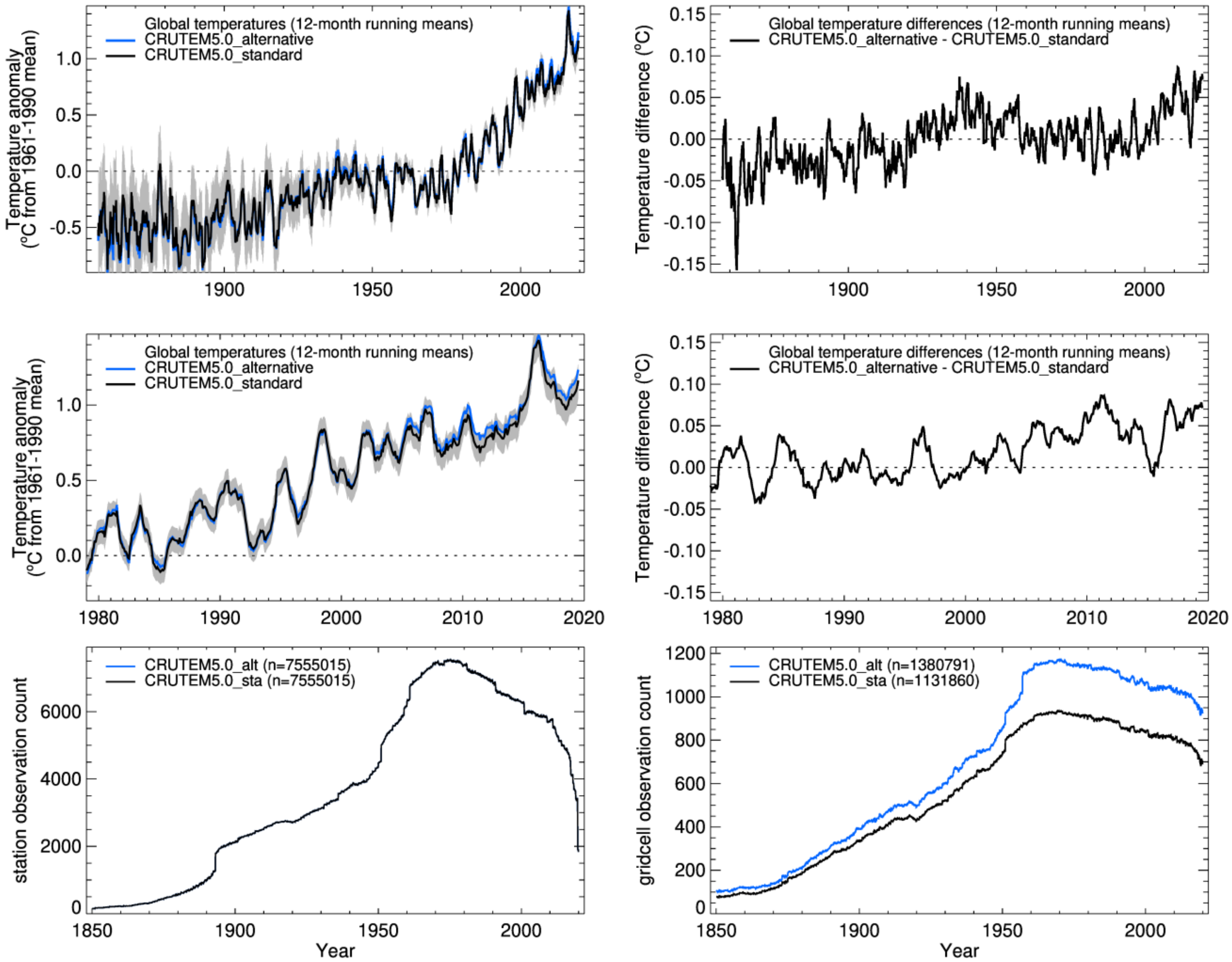


Figure 8. As **Figure 6** except comparing global-mean temperature series and observation counts from the CRUTEM5.0 station database using outlier checking and normal requirements from CRUTEM5 for standard gridding (black) and alternative gridding (blue). Alternative gridding allows high latitude stations to contribute to multiple grid cells that lie within a similar longitudinal distance as an equatorial grid cell. The grey shading is the 95% confidence interval for CRUTEM5.0 data with CRUTEM5 methods and standard gridding.

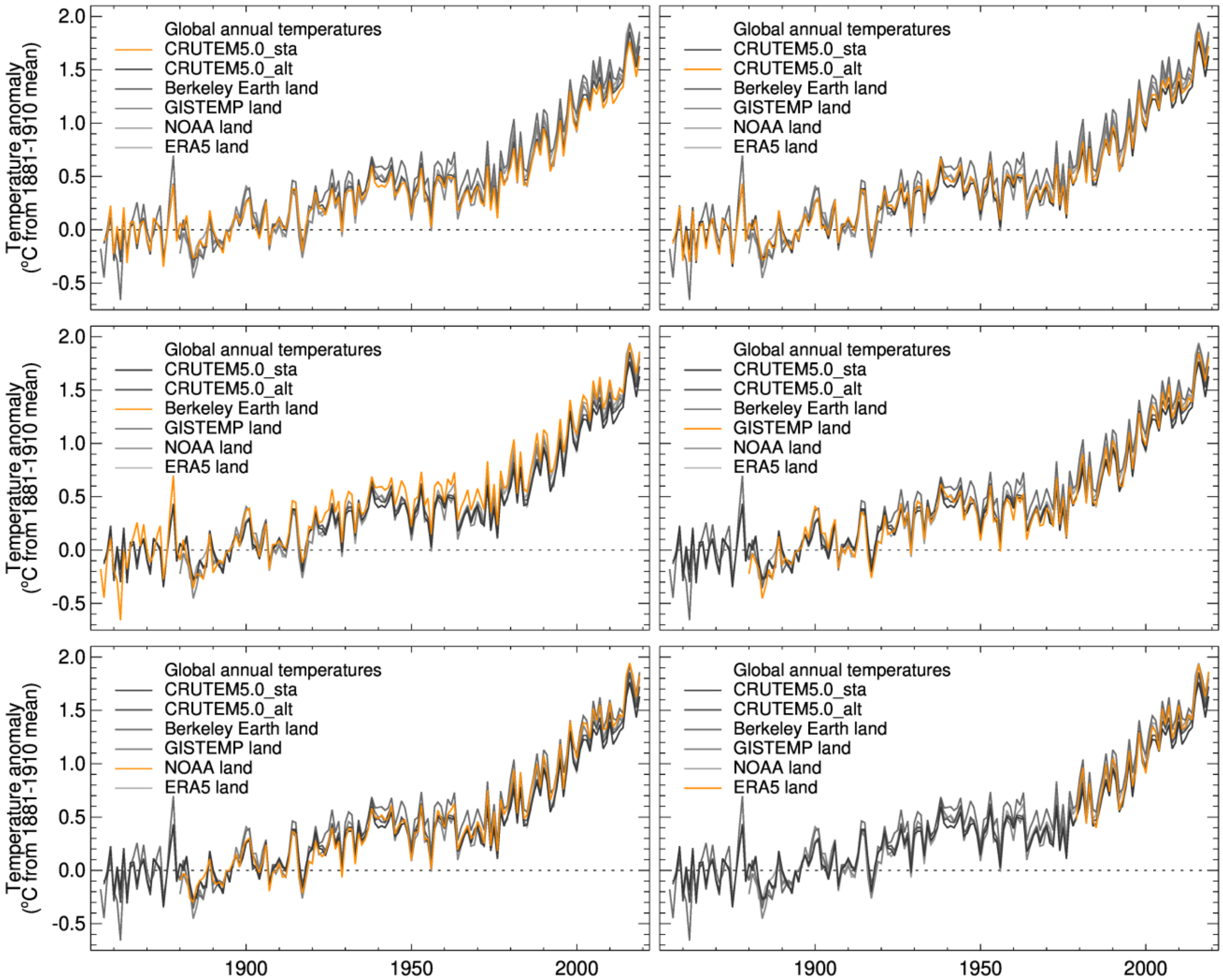


Figure 9. Comparison of global, annual-mean land temperature series from CRUTEM5.0 with standard gridded, CRUTEM5.0 with alternative gridding, Berkeley Earth, GISTEMP, NOAA V5 and ERA5, as anomalies from the 1881-1910 mean (dotted horizontal lines), the first 30-year mean for which five of the six series have data. The ERA5 series (which begins in 1979) is shifted so that it's mean matches the mean of the other five series over their overlap period. All panels show the same data, but each series is highlighted in orange in one panel each, so that the position of that series compared with the multi-dataset ensemble can clearly be seen.

Table 1. Main sources of regular (approximately annual) updates for CRUTEM4.0 and CRUTEM5.0.

Region	Provider	Current source	Publication	Comments
Australia	Bureau of Meteorology (BoM)	http://www.bom.gov.au/climate/data/acorn-sat/	Jovanovic et al. (2012); Trewin (2018)	ACORN-SAT: 112 series plus 8 from remote islands and Antarctic coastal stations
Canada	Environment and Climate Change Canada	http://data.ec.gc.ca/data/climate/scientificknowledge/adjusted-and-homogenized-canadian-climate-data-ahccd/homogenized-surface-air-temperature-ahccd/	Vincent et al. (2012)	338 series
China	China Meteorological Agency (CMA)	Provided through personal contacts at CMA or Qingxiang Li at Sun Yat-Sen University	Xu et al. (2013)	380 series (but 420 were received in 2018)
Russian Federation	All-Russia Institute of Hydrometeorological Information – World Data Centre (RIHMI-WDC)	http://meteo.ru/english/climate/d_temp.php ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/	Bulygina & Razuvaev (2012); Menne et al. (2012)	518 series until 2017; subsequent updates were obtained from GHCN-Daily
Contiguous USA	National Oceanic and Aeronautical Administration (NOAA)	https://www.ncdc.noaa.gov/ushcn/data-access	Menne et al. (2009)	USHCN 1218 series

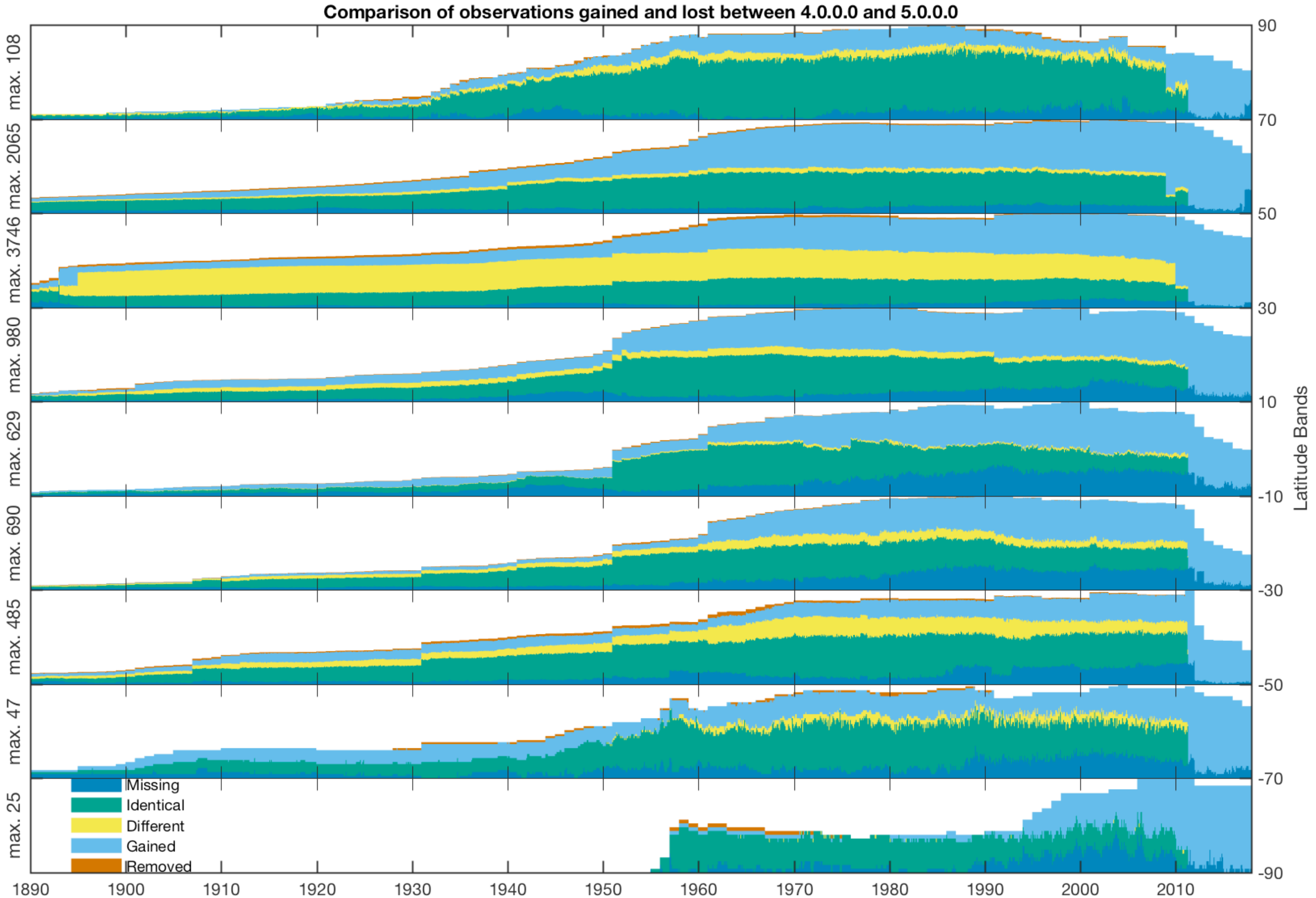
Table 2. Listing of significant new acquisitions since CRUTEM4.0 (see Supplementary Tables for more details).

Region	Provider, Source and/or Comments	Relevant publication	Number of series
Global	ISTI / ex-colonial archive		264
Global	NOAA GHCN v4	Menne et al. (2018)	141
Alaska	NOAA GHCN v4	Menne et al. (2018)	~50 new, ~50 augmented
Caribbean	CARIWIG project	Jones et al. (2016)	~50 new
South America	Latin American Climate Assessment & Dataset LACA&D		21
South America	Regional Climate Centres Network in southern South America (RCC-SSA)		~92 new, ~88 augmented
Andean region	Personal contact with colleagues in Chile and Argentina		~20 new, ~100 augmented
Bolivia, Peru	DECADE project / homogenized	Hunziker et al. (2017)	8 new, 1 augmented
Chile	CLARIS project	Penalba et al. (2014)	9
Chile	Center for Climate and Resilience Research (CR2) / CRU archives	Boisier et al. (2018)	~27
Uruguay	ISTI / Institute of Meteorology, Uruguay		11
Europe	ECA&D project / KNMI / not homogenized	van der Schrier et al. (2013)	1357
Denmark, Faroes, Greenland	Danish Meteorological Institute (DMI) / CRU / most are homogenized		11 new, 29 augmented
Germany, Poland	ISTI	Rennie et al. (2014)	58 new, 33 augmented
Iceland	Iceland Met. Office / homogenized		8 new, 10 augmented
Netherlands	KNMI / homogenized	van der Schrier et al. (2011)	10 augmented
Norway	Norwegian Meteorological Institute (NMI) / homogenized		186 augmented
Pyrenees	Servei Meteorològic de Catalunya (SMC) / homogenized		38 new
Spain	Universitat Rovira i Virgili (URV) / SDATS / homogenized	Brunet et al. (2006)	10 new, 12 augmented
Sweden	Swedish Meteorological and Hydrological Institute (SMHI) / ECA&D project		37
Southern Africa	SASSCAL project / CSAG / CRU archives		94
ASEAN region	Malaysia Meteorological Department for ASEAN		324
Indonesia	Meteorological, Climatological and Geophysical Agency (BMKG)		80+
Israel	Israel Meteorological Service (IMS)		4
Japan	Japan Meteorological Agency (JMA) / ISTI / NOAA GHCN v3		294
SE Asia, Australia	Southeast Asian Climate Assessment & Dataset (SACA&D)	van den Besselaar et al. (2017)	50 new, 4 augmented
Taiwan	Central Weather Bureau		32

NE Tibet	Key Laboratory of Desert and Desertification		8 new, 18 augmented
New Zealand	National Institute of Water & Atmospheric Research (NIWA) / homogenized	Mullan et al. (2018)	7

ISTI=International Surface Temperature Initiative; NOAA=National Oceanic and Aeronautical Administration; KNMI=Royal Netherlands Meteorological Institute; see Supplementary Material tables for definition of other acronyms.

Figure 1.



Comparison of observations gained and lost between 4.0.0.0 and 5.0.0.0
max. 108
max. 2065
max. 3746
max. 980
max. 629
max. 690
max. 485
max. 47
max. 25
Missing
Identical
Different
Gained
Removed
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
90
70
50
30
10
-10
-30
-50
-70
-90
Latitude Bands

Figure 2.

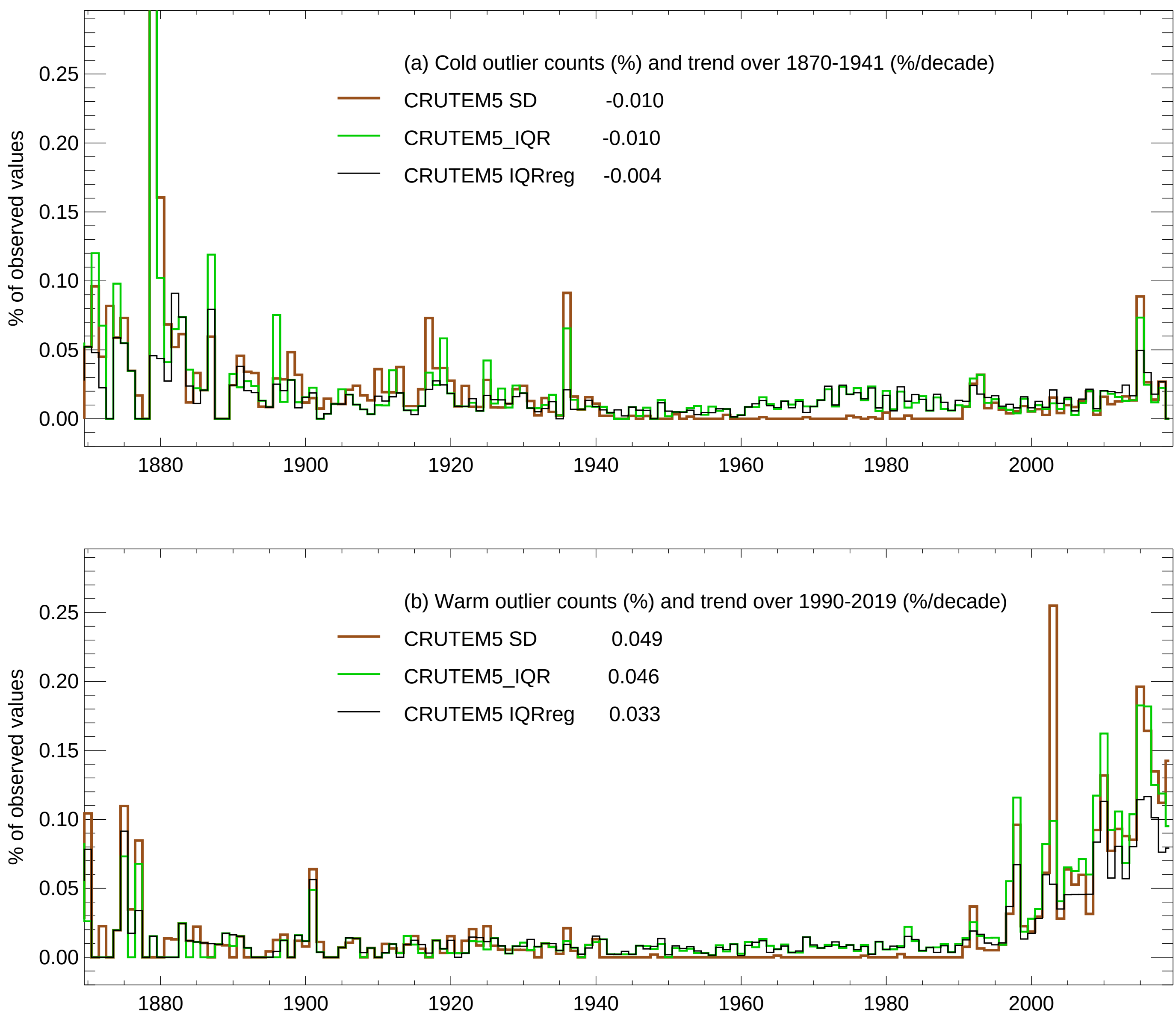
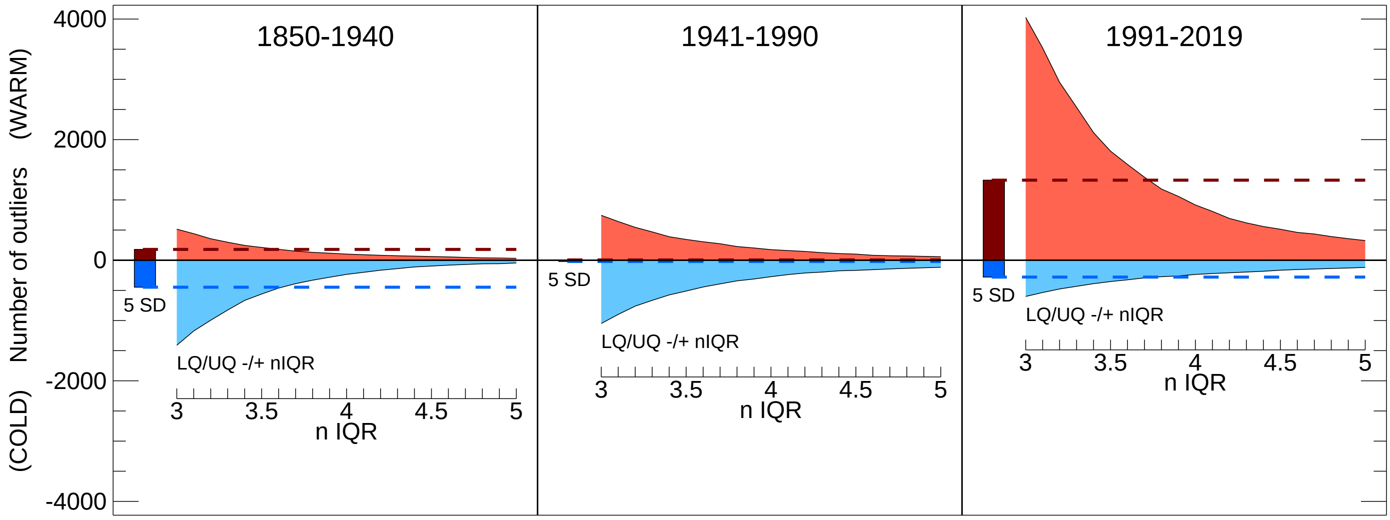
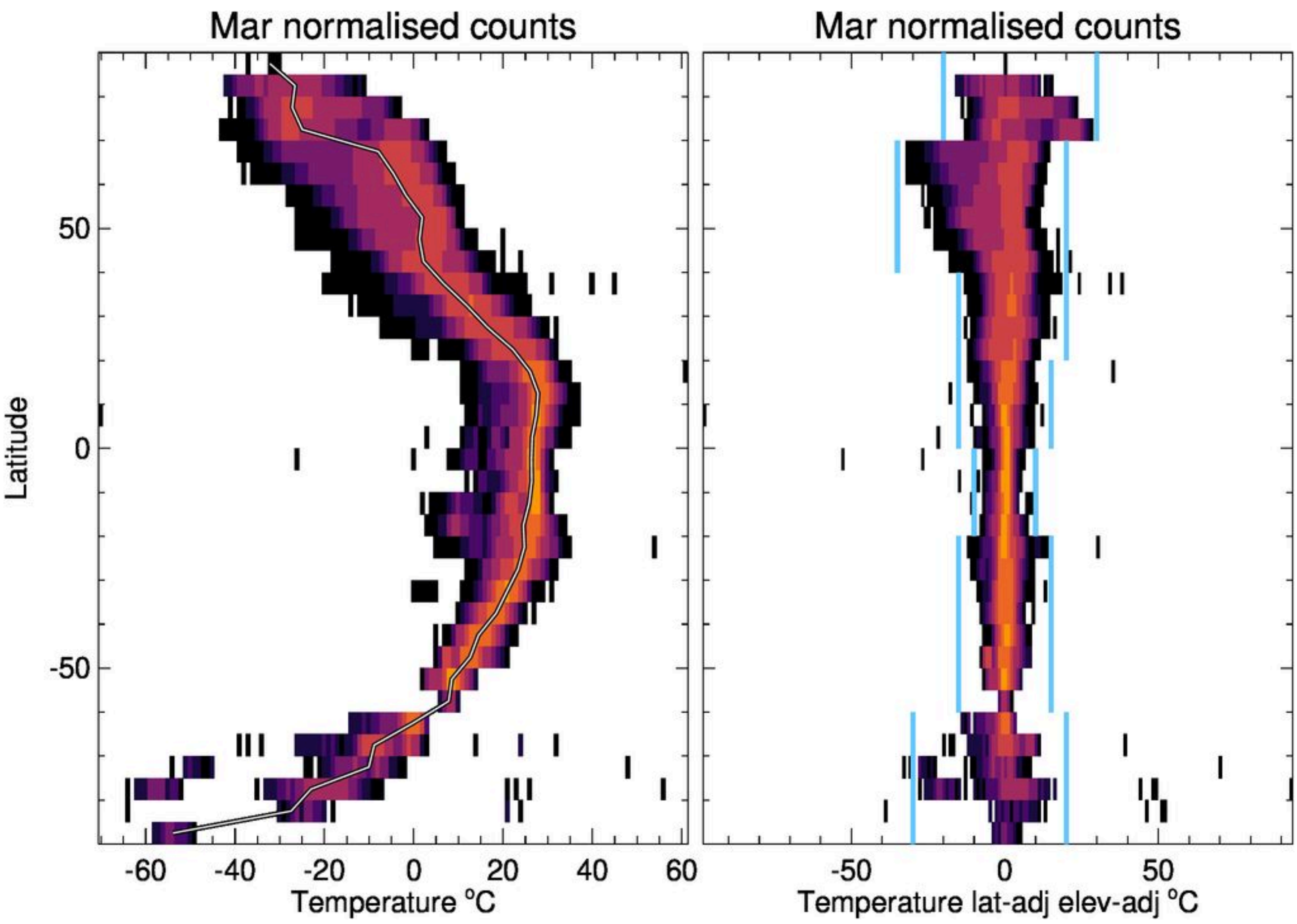

(a) Cold outlier counts (%) and trend over 1870-1941 (%/decade)
CRUTEM5 SD -0.010
CRUTEM5_IQR -0.010
CRUTEM5 IQRreg -0.004
% of observed values
0.25
0.20
0.15
0.10
0.05
0.00
1880
1900
1920
1940
1960
1980
2000
(b) Warm outlier counts (%) and trend over 1990-2019 (%/decade)
CRUTEM5 SD 0.049
CRUTEM5_IQR 0.046
CRUTEM5 IQRreg 0.033
% of observed values
0.25
0.20
0.15
0.10
0.05
0.00
1880
1900
1920
1940
1960
1980
2000

Figure 3.



1850-1940
1941-1990
1991-2019
4000
2000
0
-2000
-4000
Number of outliers
(WARM)
(COLD)
5 SD
LQ/UQ -/+ nIQR
3
3.5
4
4.5
5
n IQR

Figure 4.



Mar normalised counts
Mar normalised counts
Latitude
50
0
-50
-60
-40
-20
0
20
40
60
Temperature °C
-50
0
50
Temperature lat-adj elev-adj °C

Figure 5.

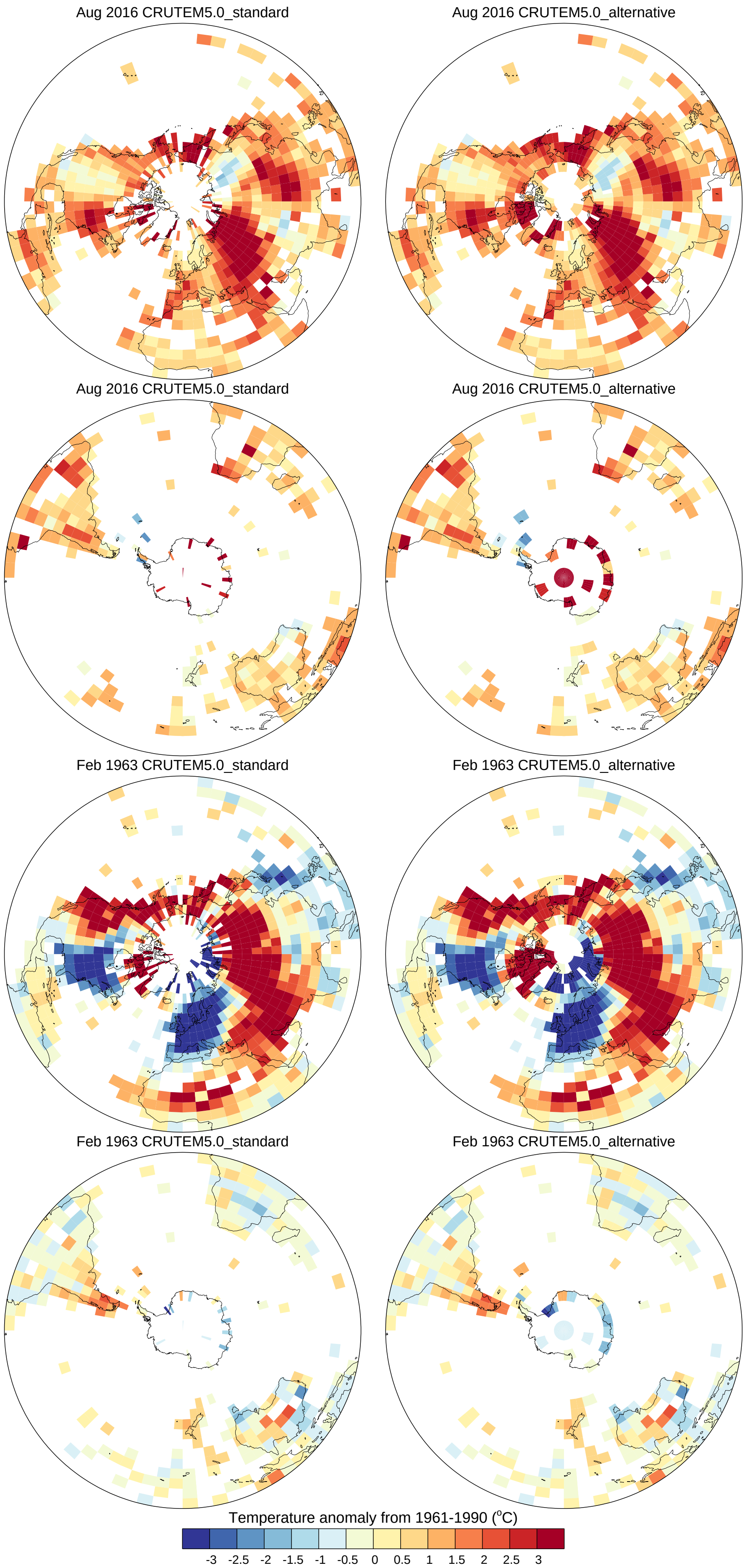

Aug 2016 CRUTEM5.0_standard
Aug 2016 CRUTEM5.0_alternative
Aug 2016 CRUTEM5.0_standard
Aug 2016 CRUTEM5.0_alternative
Feb 1963 CRUTEM5.0_standard
Feb 1963 CRUTEM5.0_alternative
Feb 1963 CRUTEM5.0_standard
Feb 1963 CRUTEM5.0_alternative
Temperature anomaly from 1961-1990 (°C)
-3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5 3

Figure 6.

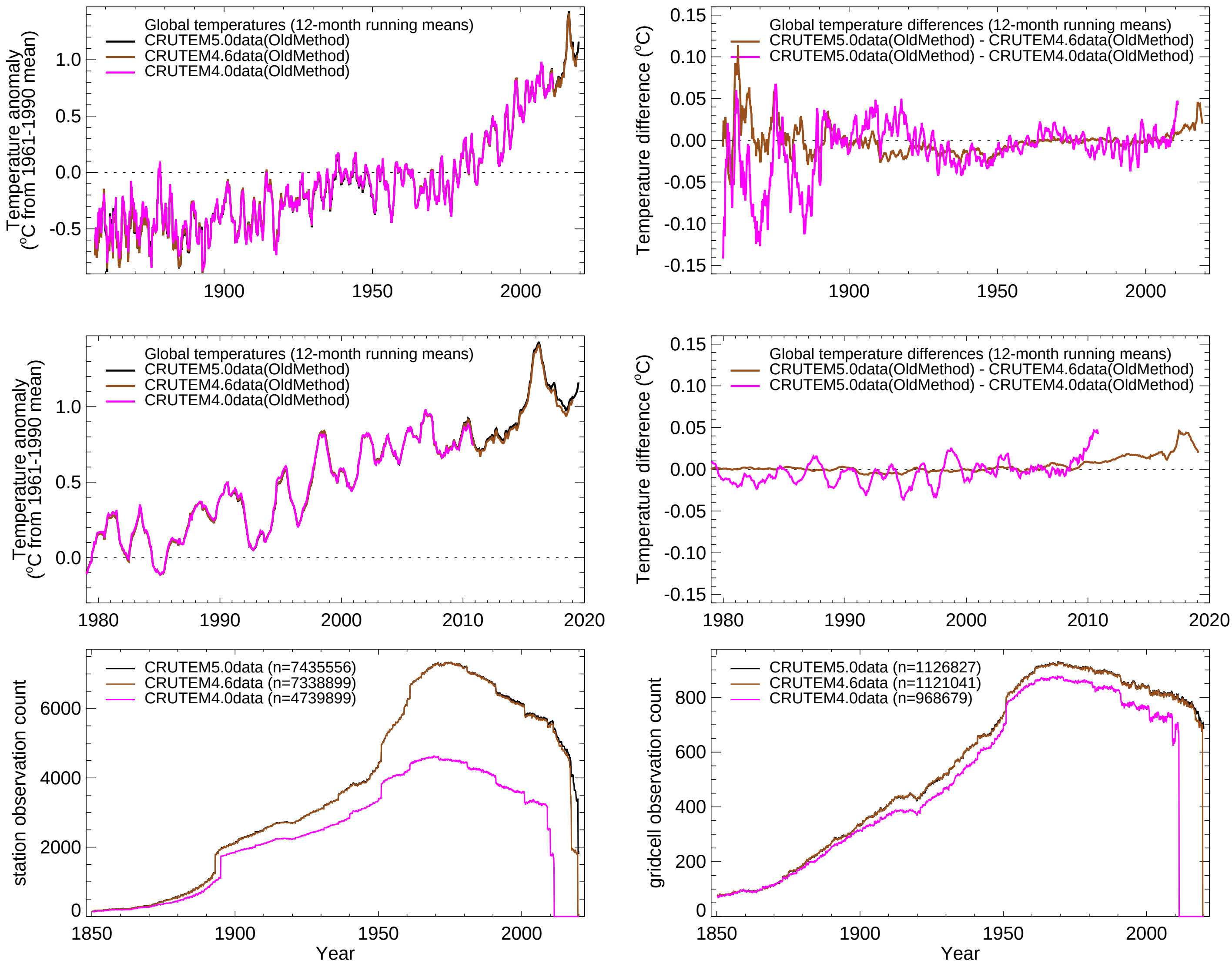

Global temperatures (12-month running means)
CRUTEM5.0data(OldMethod)
CRUTEM4.6data(OldMethod)
CRUTEM4.0data(OldMethod)
Temperature anomaly (°C from 1961-1990 mean)
Global temperature differences (12-month running means)
CRUTEM5.0data(OldMethod) - CRUTEM4.6data(OldMethod)
CRUTEM5.0data(OldMethod) - CRUTEM4.0data(OldMethod)
Temperature difference (°C)
CRUTEM5.0data (n=7435556)
CRUTEM4.6data (n=7338899)
CRUTEM4.0data (n=4739899)
station observation count
CRUTEM5.0data (n=1126827)
CRUTEM4.6data (n=1121041)
CRUTEM4.0data (n=968679)
gridcell observation count
Year

Figure 7.

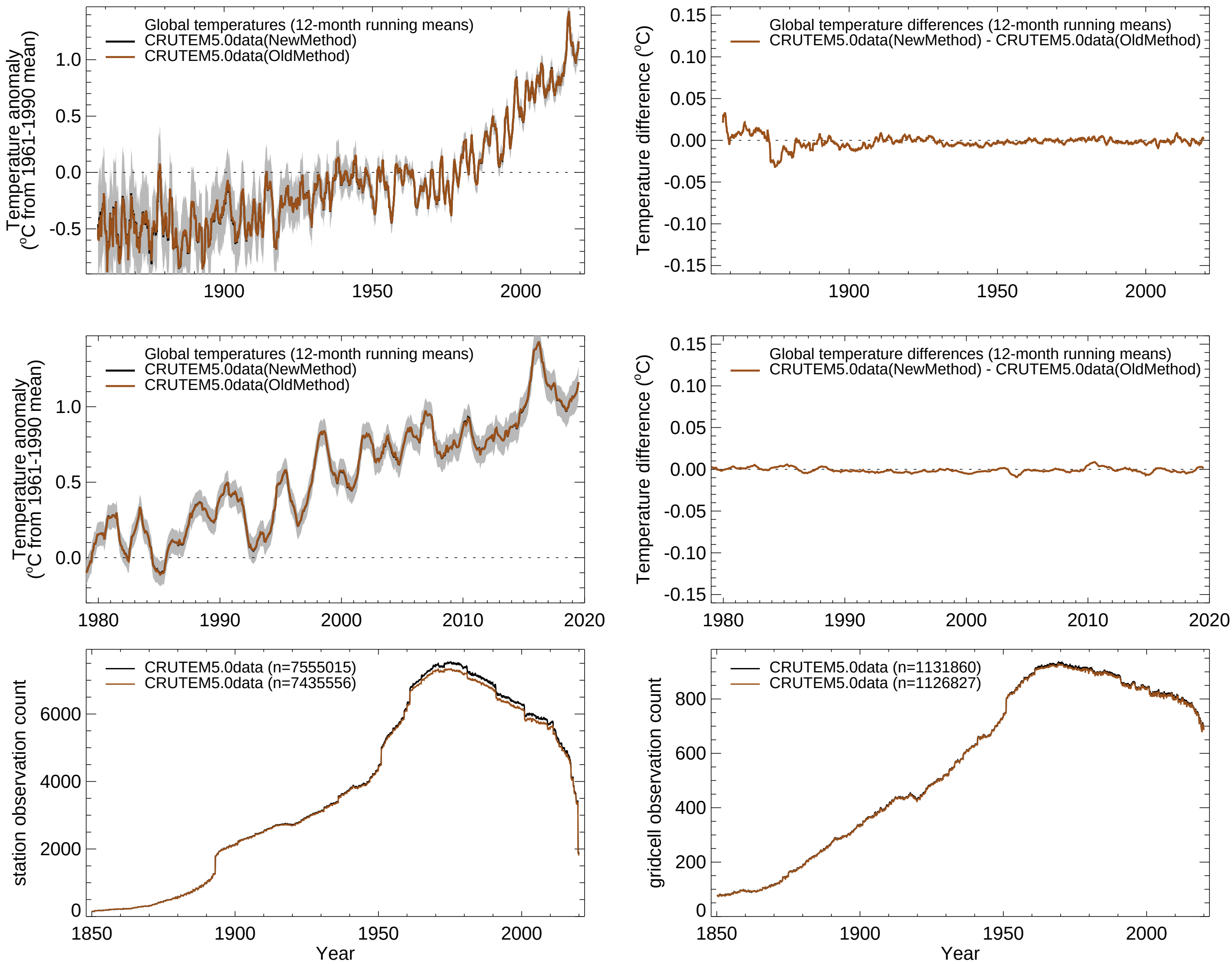

Global temperatures (12-month running means)
CRUTEM5.0data(NewMethod)
CRUTEM5.0data(OldMethod)
Temperature anomaly (°C from 1961-1990 mean)
Global temperature differences (12-month running means)
CRUTEM5.0data(NewMethod) - CRUTEM5.0data(OldMethod)
Temperature difference (°C)
CRUTEM5.0data (n=7555015)
CRUTEM5.0data (n=7435556)
station observation count
CRUTEM5.0data (n=1131860)
CRUTEM5.0data (n=1126827)
gridcell observation count
Year

Figure 8.

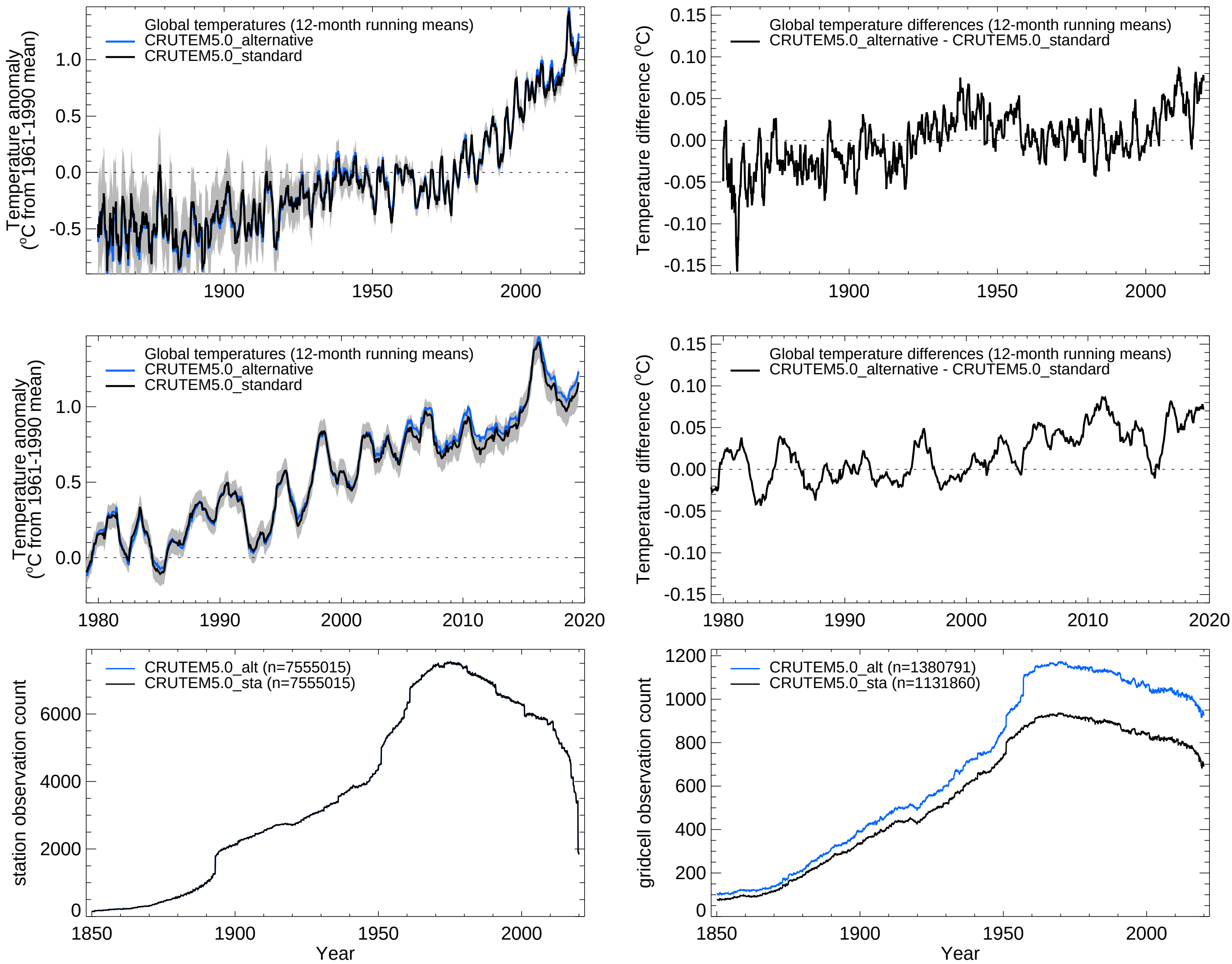
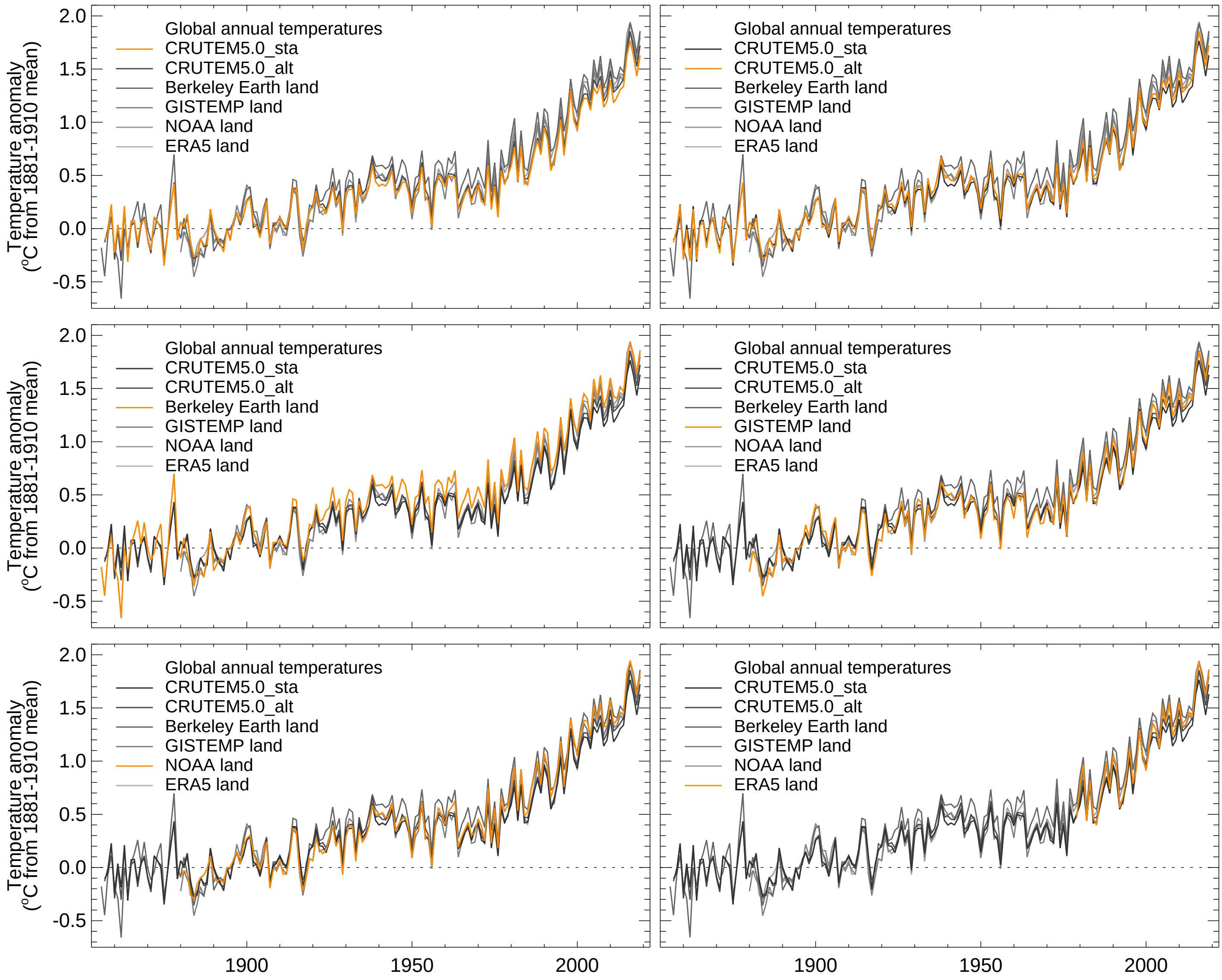

Global temperatures (12-month running means)
CRUTEM5.0_alternative
CRUTEM5.0_standard
Temperature anomaly (°C from 1961-1990 mean)
Global temperature differences (12-month running means)
CRUTEM5.0_alternative - CRUTEM5.0_standard
Temperature difference (°C)
CRUTEM5.0_alt (n=7555015)
CRUTEM5.0_sta (n=7555015)
station observation count
CRUTEM5.0_alt (n=1380791)
CRUTEM5.0_sta (n=1131860)
gridcell observation count
Year

Figure 9.


Global annual temperatures
CRUTEM5.0_sta
CRUTEM5.0_alt
Berkeley Earth land
GISTEMP land
NOAA land
ERA5 land
Temperature anomaly
(ºC from 1881-1910 mean)
2.0
1.5
1.0
0.5
0.0
-0.5
1900
1950
2000