**Earth System
Science
Data**

The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth

T. J. Osborn[1] **and P. D. Jones**[1,2]

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK
[2]Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence to: T. J. Osborn (t.osborn@uea.ac.uk)

Abstract. The CRUTEM4 (Climatic Research Unit Temperature, version 4) land-surface air temperature data set is one of the most widely used records of the climate system. Here we provide an important additional dissemination route for this data set: online access to monthly, seasonal and annual data values and time series graphs via Google Earth. This is achieved via an interface written in Keyhole Markup Language (KML) and also provides access to the underlying weather station data used to construct the CRUTEM4 data set. A mathematical description of the construction of the CRUTEM4 data set (and its predecessor versions) is also provided, together with an archive of some previous versions and a recommendation for identifying the precise version of the data set used in a particular study.

The CRUTEM4 data set used here is available from doi:10.5285/EECBA94F-62F9-4B7C-88D3-482F2C93C468.

1 Introduction

The Climatic Research Unit (CRU) temperature data set (CRUTEM) forms the land component of the global temperature data set developed jointly by the UK Met Office Hadley Centre and CRU (HadCRUT). The scientific basis for the latest versions of these data sets was published in 2012 (CRUTEM4: Jones et al., 2012; HadCRUT4: Morice et al., 2012). These are (and remain) the primary references for these data sets. The purpose of the present paper is three-fold. Firstly, to provide a more complete technical description of the construction of CRUTEM4 and comparison with the construction of its predecessor versions. This information, provided qualitatively across various previous publications, is drawn together here and presented in mathematical form. Reproducibility is a key determinant of confidence in scientific results, and because particular attention has been

paid to the CRUTEM data set (e.g. House of Commons Science and Technology Committee, 2010), it is valuable for the information presented here to be in the public domain. Secondly, we provide an archive of some previous versions of the CRUTEM data set so that the precise version used in a particular study can be identified.

Finally, to facilitate direct access to visualisations and the underlying data values, the current version is made available via an interface written in Google Earth's Keyhole Markup Language (KML). This is a standard of the Open Geospatial Consortium (http://www.opengeospatial.org/standards/kml) and allows the data set to be rendered in Earth browsers such as Google Earth (http://earth.google.com/). The KML interface overlays an online rendition of the data set – the weather station locations and monthly average temperatures, the grid-box monthly temperature anomalies, and seasonal and annual time series graphs of all these data onto the

three-dimensional representation of the Earth provided by the Google Earth software – enabling the data set to be explored and accessed interactively and graphically. This significantly enhances the accessibility of this key climate data set, whether for exploring the data and extracting regional information for research and teaching or for identifying errors and limitations, without the need to develop bespoke software to analyse the data. Given that the Google Earth software is freely available and has been downloaded more than one billion times (http://googleblog.blogspot.co.uk/2011/10/google-earth-downloaded-more-than-one.html), this represents an important additional dissemination route for the CRUTEM4 data set.

2 Construction of CRUTEM gridded, global and hemispheric temperature anomalies from the CRUTEM station temperature database

2.1 Introduction

A mathematical description of the stages in the construction of CRUTEM gridded temperature anomalies from the CRUTEM station temperature database is presented here. These steps have been identified from the published articles describing the construction of the CRUTEM4 data set (and its previous versions). The page numbers from which the information compiled here was taken are indicated for each of these articles, where CRUTEM1986NH is Jones et al. (1986a); CRUTEM1986SH is Jones et al. (1986b); CRUTEM1 is Jones (1994); CRUTEM2 is Jones and Moberg (2003); CRUTEM3 is Brohan et al. (2006) and CRUTEM4 is Jones et al. (2012). The current article does *not* outline how the station temperature database was assembled (i.e. sources of data and what precedence is given when multiple sources are available), which is described elsewhere (US DoE TR017, 1985; Jones and Moberg, 2003; Jones et al., 2012).

2.2 Homogeneity adjustments

Changes in the location or local environment of a weather station, in the way in which the thermometer is exposed, or in recording practices (such as time of observation or the way in which daily and monthly averages are calculated) can introduce inhomogeneities into the station time series of monthly temperatures. An estimated time series of temperatures that might have been recorded in the absence of such changes can be made by adjusting the mean values of the series prior to each change. For a change in station altitude, the adjustment factor might be estimated using a lapse rate, but for other cases it is estimated by differences from multiple neigh-

bouring series (note that in data-sparse regions, neighbours might be separated by several hundred kilometres). This approach was the basis for the extensive homogenisation exercise undertaken by CRU in the 1980s, documented in US DoE TR022 (1985) and US DoE TR027 (1986). Electronic copies of these reports are available on the CRU website (http://www.cru.uea.ac.uk/publications/crurp).

Since the 1980s, CRU recommended that further homogeneity efforts should be instead undertaken by National Meteorological Services (e.g. Jones and Moberg, 2003), and the results of such projects have been incorporated into the CRUTEM station temperature database (Jones et al., 2012). The only homogeneity adjustment made by CRU since those made in the 1980s (recorded as explained above) is a recent adjustment to the St. Helena station (ID 619010), which was incorporated into CRUTEM.4.2.0.0 in May 2013 (see Sect. 3 for an explanation of version numbering). This was a standard lapse-rate adjustment for a known reduction in altitude of 166 m that occurred when the station moved in September 1976 (D. Lister, personal communication, 23 November 2012); all values prior to 1977 were increased by 1.1 °C.

2.3 Some definitions

The nomenclature used to identify the station data, gridded data, and the space and time dimensions are defined first. The definitions are for a grid with spatial resolution 5° latitude by 5° longitude, as used in CRUTEM1 (p. 1795), CRUTEM2 (p. 213), CRUTEM3 (p. 2; but note that results could be calculated for other grid sizes), and CRUTEM4 (p. 6). In earlier versions, a 5° latitude by 10° longitude grid was used (CRUTEM1986NH, p. 167; CRUTEM1986SH, p. 1216) and the definitions provided here would need to be modified appropriately.

2.3.1 Dimensions

Year index is $t = 1 \ldots n_t$.
Month index is $m = 1 \ldots 12$ (1 = January, 12 = December).
Grid box index in longitude direction is $i = 1 \ldots n_i$ (for 5° grid, $n_i = 72$).
Grid box index in latitude direction is $j = 1 \ldots n_j$ (for 5° grid, $n_j = 36$).
Weather station index is $s = 1 \ldots n_s$.

2.3.2 Time variables

Year value is yr_t (currently running from $\mathrm{yr}_1 = 1850$ to $\mathrm{yr}_{nt} = 2013$).

2.3.3 Gridded variables

Longitude[1] of centre of grid box i is x_i ($x_1 = -177.5°$, $x_2 = -172.5°$, $x_3 = -167.5°$, etc., $x_{ni} = 177.5°$). Note that $x_i < 0$ are west, and $x_i > 0$ are east, of the Greenwich meridian; $\Delta x = 5°$ is the east–west size of all grid boxes; the longitude of the western edge of grid box i is $x_i - \Delta x/2$ and its eastern edge is $x_i + \Delta x/2$.

Latitude of centre of grid box j is y_j ($y_1 = -87.5°$, $y_2 = -82.5°$, $y_3 = -77.5°$, etc., $y_{nj} = 87.5°$). Note that $y_j < 0$ are south, and $y_j > 0$ are north, of the Equator; $\Delta y = 5°$ is the north–south size of all grid boxes; the latitude of the southern edge of grid box j is $y_j - \Delta y/2$ and its northern edge is $y_j + \Delta y/2$.

Grid-box temperature anomaly (°C) for grid box i, j in year t and month m is $G'_{i,j,t,m}$.

A "mask" to indicate which grid-box temperature anomalies are missing and which are available is $\Delta_{i,j,t,m}$, where $\Delta_{i,j,t,m} = 0$ means that there is no value in year t and month m for grid box i, j while $\Delta_{i,j,t,m} = 1$ means that there is a value.

2.3.4 Station data

Longitude[2] of weather station s is lon_s.
Latitude of weather station s is lat_s.

"Raw" monthly-mean station temperature observation (°C) for station s in year t and month m is $T_{s,t,m}$. Note that "raw" means as received by CRU (after conversion of units or application of homogeneity adjustments in some cases), and previous processing (not least the calculation of the monthly means from the daily or sub-daily measurement values) will have been done before we receive the data.

A "mask" to indicate which station temperature observations are missing (or are not to be used) and which are available (and should be used) is $\delta_{s,t,m}$. Some values are available but are considered unreliable or fail the outlier check; for these values, this "mask" is set to zero to prevent them being used. $\delta_{s,t,m} = 0$ means that there is no observational value

[1]The conventions for the ordering of grid boxes (west to east, south to north) are consistent throughout this article, but the algorithms can be implemented using other conventions provided that they remain consistent throughout. Indeed, some versions of the CRUTEM programs begin the grid at the North Pole rather than the South Pole, or begin the grid at the Greenwich Meridian rather than at the International Dateline. A particular aspect to be aware of is that the longitudes of the weather stations stored in the station data files are recorded under the convention that positive values are *west* of the Greenwich Meridian, not east (which are recorded with negative values). The programs that implement the algorithms described here take this into account.

[2]The convention used within this article is that the longitudes of the weather stations follow the same sign convention as the longitudes of the grid boxes, but note that the station data files currently follow the opposite sign convention and the programs that implement the algorithms described here take this into account (see footnote [1]).

in year t and month m for station s, while $\delta_{s,t,m} = 1$ means that there is an observational value and it is used.

2.4 Calculation of "normals" and standard deviations

A "normal" is the name used in climatology for the mean value over a reference period (also known as the "base" period or the "normal" period). For each station s and each month of the year m, the number of values within the reference period is determined:

$$N_{s,m} = \sum_{t=\mathrm{REF1}}^{\mathrm{REF2}} \delta_{s,t,m}, \tag{1}$$

where REF1 and REF2 define the reference period (currently we use $\mathrm{yr}_{\mathrm{REF1}} = 1961$ and $\mathrm{yr}_{\mathrm{REF2}} = 1990$; it was 1951–1970 for CRUTEM1986NH p. 167 and CRUTEM1986SH p. 1217; 1961–1990 for CRUTEM1 p. 1795; 1961–1990 for CRUTEM2 p. 212; 1961–1990 for CRUTEM3 p. 2; 1961–1990 for CRUTEM4 p. 6). If there are sufficient values to obtain a reasonable estimate of the mean value (currently we require $N_{s,m} \geq 14$, though we have previously used different criteria, including ones that require a minimum number of values during each decade of the reference period: ≥ 15 for CRUTEM1986NH p. 167 and CRUTEM1986SH p. 1217; ≥ 21 for CRUTEM1 p. 1795; ≥ 20 with ≥ 4 in each decade for CRUTEM2 p. 213; ≥ 15 for CRUTEM3 p. 2; ≥ 14 for CRUTEM4 p. 6), then we estimate the mean value by

$$\bar{T}_{s,m} = \frac{\sum\limits_{t=\mathrm{REF1}}^{\mathrm{REF2}} \delta_{s,t,m} T_{s,t,m}}{N_{s,m}}. \tag{2}$$

Note that we take into account the uncertainty (sampling error) associated with $\bar{T}_{s,m}$, including the greater error when $N_{s,m} < 30$ (i.e. incomplete data), in the estimation of CRUTEM3 and CRUTEM4 uncertainty ranges (as described in Sect. 2.3.1.3 of Brohan et al., 2006).

If we have insufficient values to determine a normal (i.e. $N_{s,m} < 14$, or the different criteria noted above for the earlier versions), then we follow one of four options:

1. If the station has sufficient data to estimate a normal for the 1951–1970 period and the grid box from an earlier version of CRUTEM contains data across the longer 1951–1990 period, then we estimate the 1961–1990 normal for the station using its 1951–1970 normal adjusted by the difference between the grid-box averages (in the earlier version) for 1961–1990 and 1951–1970.

2. If a neighbouring station does have sufficient values to determine its normal, then we calculate the mean difference between the temperatures recorded at this neighbouring station and the temperature recorded at the current station over a different period when they both have

data (e.g. 1951–1970), and assume that this mean difference still holds during the reference period. The normal for the current station is then calculated as the sum of the normal for the neighbouring station plus the mean difference between the two stations' temperatures.

3. If the World Meteorological Organisation (WMO) have published a 1961–1990 normal for the station and the reference period (perhaps because the National Meteorological Service had calculated it from additional data not available to us), then we use that. We rely on a WMO normal for about 150 ($\sim 2.5\%$) of the stations.

4. Otherwise we omit all temperatures recorded at this station for all months of the year and set all $\delta_{s,t,m} = 0$.

The options used, and their order of priority, have been as follows: (i) estimated from neighbours for CRUTEM1986NH (p. 167) and CRUTEM1986SH (p. 1217); (ii) estimated from gridded data or from neighbours for CRUTEM1 (p. 1795); (iii) WMO normals or estimated from gridded data or from neighbours for CRUTEM2 (p. 213); and (iv) WMO normals or estimated from neighbours for CRUTEM3 (p. 2) and CRUTEM4.

The calculation of the station standard deviations is quite similar, though a longer reference period is used because (i) the sampling error associated with standard deviations tends to be relatively larger than that associated with means, and thus a bigger sample is preferred; and (ii) we do not use other sources (e.g. WMO) in cases where we have insufficient data. For each station s and each month of the year m, the temperature standard deviation is calculated as

$$\sigma_{s,m} = \sqrt{\frac{\sum\limits_{t=\text{REF3}}^{\text{REF4}} \delta_{s,t,m}\left(T_{s,t,m} - \bar{T}_{s,m}\right)^2}{\left(\sum\limits_{t=\text{REF3}}^{\text{REF4}} \delta_{s,t,m}\right) - 1}}, \quad (3)$$

where $\bar{T}_{s,m}$ is the mean computed over the longer reference period (currently we use $\text{yr}_{\text{REF3}} = 1941$ and $\text{yr}_{\text{REF4}} = 1990$; it was 1941–1990 for CRUTEM1 p. 1795; 1921–1990 for CRUTEM2 p. 213; 1941–1990 for CRUTEM4 p. 6). If the station does not have at least 15 yr of data during this longer reference period, then a standard deviation was not estimated and that station was not used (CRUTEM4 p. 6).

2.5 Conversion to anomalies (i.e. deviations from a reference value)

Given a "normal" for station s and month m, $\bar{T}_{s,m}$, all observations are converted to anomalies according to

$$\text{where} \quad \delta_{s,t,m} = 1, \quad T'_{s,t,m} = T_{s,t,m} - \bar{T}_{s,m}. \quad (4)$$

2.6 Removal of outliers (quality control)

Given an estimate of the standard deviation of monthly temperatures for station s and month m, $\sigma_{s,m}$, any anomalies

exceeding 5 standard deviations are removed (it was > 6 for CRUTEM1 p. 1795; > 5 for CRUTEM2 p. 213; > 5 for CRUTEM3 p. 2; > 5 for CRUTEM4 p. 6):

$$\text{where} \quad \left|T'_{s,t,m}\right| > 5\sigma_{s,m} \quad \text{set } \delta_{s,t,m} = 0. \quad (5)$$

In fact, this outlier check resulted in some cases where an obvious error could be corrected (e.g. the measurement value was recorded or digitised as 10 or 20 °C too warm or cold, or the sign was incorrect); in these cases the station temperature files have been corrected. After removal or correction of outliers, the mean (i.e. normal) and standard deviation of the monthly temperatures is recalculated and the outlier check is repeated (CRUTEM1 pp. 1795–1796; CRUTEM2 p. 213; CRUTEM4 p. 6).

2.7 Gridding of station anomalies

This is done separately for each year t and each month m, and for each grid box i, j. The grid box mean anomaly is the arithmetic mean of all the available values (a different method was used for CRUTEM1986NH p. 167 and CRUTEM1986SH p. 1216; the method described here was used for CRUTEM1 pp. 1795–1796; CRUTEM2 p. 213; CRUTEM3 p. 2; CRUTEM4 p. 6), i.e. for all s where $x_i - \Delta x/2 \leq \text{lon}_s < x_i + \Delta x/2$ and $y_j - \Delta y/2 \leq \text{lat}_s < y_j + \Delta y/2$

$$G'_{i,j,t,m} = \frac{\sum\limits_s \delta_{s,t,m} T'_{s,t,m}}{\sum\limits_s \delta_{s,t,m}} \quad \text{and} \quad \Delta_{i,j,t,m} = 1. \quad (6)$$

If no s match the above criteria, then $\Delta_{i,j,t,m} = 0$.

2.8 Calculation of hemisphere and global means

The hemispheric means are averages of the grid boxes for which temperature anomalies are available (this is, of course, time dependent), with weighting according to the area of the grid boxes (proportional to the cosine of their central latitude). Details were given for CRUTEM1986NH p. 167; CRUTEM1986SH p. 1217; CRUTEM2 p. 216; CRUTEM3 p. 10; CRUTEM4 p. 8.

$$\text{NH}_{t,m} = \frac{\sum\limits_{j=1+nj/2}^{nj} \sum\limits_{i=1}^{ni} \cos(y_j)\Delta_{i,j,t,m} G'_{i,j,t,m}}{\sum\limits_{j=1+nj/2}^{nj} \sum\limits_{i=1}^{ni} \cos(y_j)\Delta_{i,j,t,m}} \quad (7)$$

$$\text{SH}_{t,m} = \frac{\sum\limits_{j=1}^{nj/2} \sum\limits_{i=1}^{ni} \cos(y_j)\Delta_{i,j,t,m} G'_{i,j,t,m}}{\sum\limits_{j=1}^{nj/2} \sum\limits_{i=1}^{ni} \cos(y_j)\Delta_{i,j,t,m}} \quad (8)$$

The global mean can be calculated in a similar way, or as the simple arithmetic mean of the two hemisphere means, or as a

weighted mean of the two hemisphere means. Global-mean land-surface air temperature time series were not presented in CRUTEM1986NH, CRUTEM1986SH or CRUTEM1. For CRUTEM2 (p. 217) and CRUTEM3 (p. 10), global-mean time series were calculated using the simple arithmetic mean of the two hemispheres. However, this can result in a biased estimate of the global land temperature because the land areas in each hemisphere are not the same, and so for CRUTEM4 (p. 8) the global series is computed by weighting the two hemispheres approximately in proportion to the areas of their landmasses:

$$\text{Global}_{t,m} = \frac{2\text{NH}_{t,m} + \text{SH}_{t,m}}{3}. \tag{9}$$

When calculating annual-mean hemisphere-means, or annual-mean global-means, there are two options available: average in time then space or average in space then time. In other words, either (i) calculate annual-mean anomalies for each grid box first, and then calculate hemispheric or global means of those; or (ii) calculate the hemispheric or global means of the monthly anomalies, and then calculate the annual means of those monthly hemispheric or global means. The Met Office Hadley Centre uses method (i), while CRU uses method (ii).

There are various advantages and disadvantages to the choices in calculating global means and in calculating annual and hemispheric/global means, which we do not list here; we just note that small differences in the results can arise. These differences are illustrated in Fig. 1, together with the simple arithmetic (i.e. unweighted) mean of the two hemispheres used in previous versions.

2.9 Adjusting the high frequency temporal variance

We also create "variance adjusted" versions of CRUTEMn (called CRUTEMnv, e.g. CRUTEM3v, with equivalent land and marine versions HadCRUTnv, e.g. HadCRUT3v) because the variance of the grid-box mean anomalies, $G'_{i,j,t,m}$, varies according to how many stations match the Eq. (6) criteria. This can cause artificial changes in the variance of $G'_{i,j,t,m}$ between different grid boxes, and also between different time periods for the same grid box when the number of stations varies through time. These artificial changes in the variance can prevent the use of CRUTEM for monitoring changes in variability and in the occurrence of extreme values. The "variance adjusted" version, CRUTEM4v, is obtained after the construction of CRUTEM4 is complete by subsequently following these steps:

1. apply a filter to separate the grid box temperature anomaly time series into "low" and "high" frequency components;

2. scale the "high" frequency component with a time-varying scaling factor described in Osborn et al. (1997) and Jones et al. (2001) to remove the expected artificial
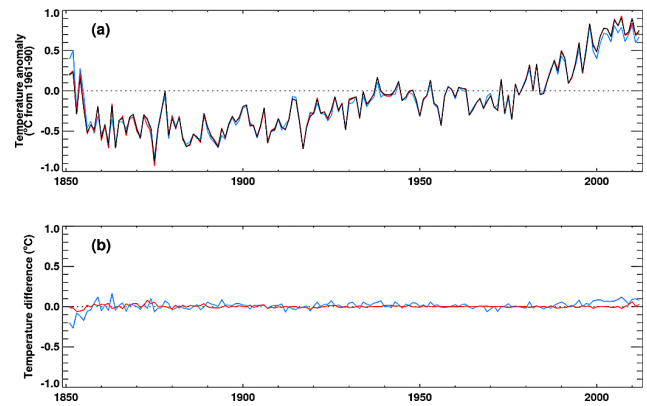

Figure 1. Annual and global mean land air temperatures anomalies calculated using three approaches: (1) form monthly hemispheric means, then annual means, then global means weighted by land area in each hemisphere (black); (2) form annual grid-box means, then hemispheric means, then global means weighted by land area in each hemisphere (red); (3) form hemispheric means, then annual means, then global means of equally-weighted hemispheres (blue). The three results are compared in **(a)**, and differences [red = (1)–(2), blue = (1)–(3)] are shown in **(b)** with the same vertical scale as in **(a)**.

time variations in the variance of the grid-box temperature anomalies;

3. combine the adjusted "high" frequency component with the original "low" frequency component to obtain the CRUTEM4v grid-box mean temperature anomaly time series.

Further details of this process are given in Jones et al. (2001) and Appendix A of Brohan et al. (2006).

3 An archive of previous versions of the CRUTEM data set

After each major version of CRUTEM has been adopted as the current operational version (e.g. CRUTEM3 was adopted in 2006), it is updated each month. Each update adds an extra month's data to the end of the series, but also includes any late reported station data for previous months and even previous years. Very occasionally, changes affecting earlier values have also been made to correct errors that have been identified or to improve the homogeneity of the series. Each month, therefore, a new version of CRUTEM is generated (usually 3 to 4 weeks after the end of the month).

For most purposes, it has been sufficient to identify the version of the data set that has been used in a particular study by the major version number (e.g. CRUTEM3) together with the appropriate reference. In some cases, it might be valuable to be able to identify a specific monthly update of the data set – for instance, if it is necessary to identify the precise version
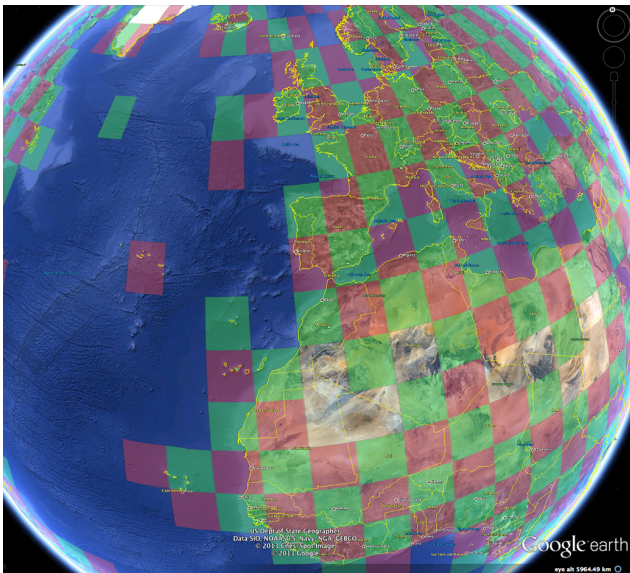
Figure 2. Image from Google Earth showing the grid boxes that contain some gridded temperature anomaly data in CRUTEM4, highlighted by translucent green and red boxes. Land areas that are not marked by a green or red box do not contain any gridded temperature anomaly data in the CRUTEM4 data set.
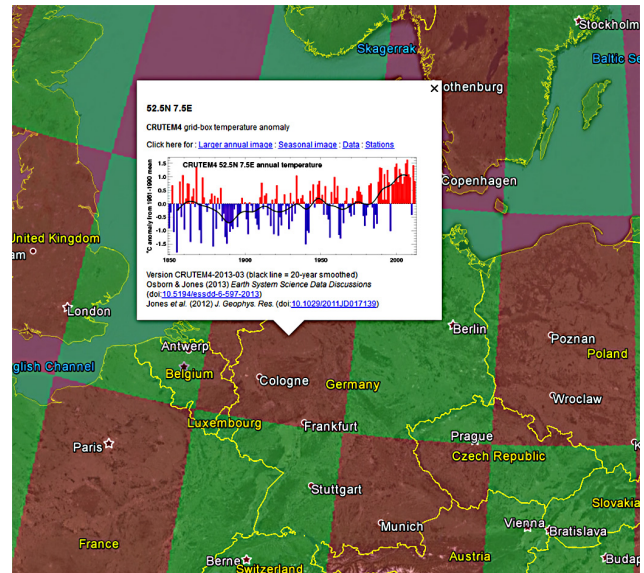

Figure 3. Image from Google Earth showing a box that appears after clicking on one of the highlighted grid boxes. The box shows the annual-mean temperature anomaly time series as bars (blue for values below the 1961–1990 reference period mean, red for values above) with a black line for the 20 yr smoothed variations. The box also shows the location of the centre of the grid box and provides links to a larger image of the annual time series, to an image of the seasonal-mean temperature anomaly time series, to a text file containing the monthly, seasonal and annual temperature anomaly values for this grid box, and a link to load into Google Earth a KML file for the CRUTEM4 weather stations that lie within this grid box.

used so that a particular published analysis can be repeated exactly.

Until now, individual monthly updates of CRUTEM have not been uniquely identified. Here we recommend that if a precise monthly version needs to be identified then an identifier that combines the major "version" with the date corresponding to the latest data contained in the data file can be constructed. Hence, a file containing data that run through to include values to the end of May 2011 would be labelled CRUTEM3-2011-05. Note that this is not the version that was created in May 2011 because that will likely have data that run through to only the end of April 2011 (and would be identified as CRUTEM3-2011-04). Updated versions are usually created 3 to 4 weeks after the end of the month. We note that the Met Office Hadley Centre have recently introduced a version numbering scheme (http://www.metoffice. gov.uk/hadobs/crutem4/data/version_numbering.html) to assist in identifying those updates that are accompanied by changes in the algorithm or changes in station data that are more significant than the routine monthly updating. The year-month identifier recommended here can be combined with the Met Office Hadley Centre version numbering scheme when both are needed.

An archive of some past versions of CRUTEM has been assembled and is available here: http://www.cru.uea.ac.uk/ cru/projects/acrid/crutem/version.htm. These past versions have been identified using the identifiers recommended above.

4   Disseminating the CRUTEM4 data set through Google Earth via a Keyhole Markup Language interface

The CRUTEM4 land-surface air temperature gridded data set is available in text and netCDF formats from the CRU and Met Office Hadley Centre websites, and the weather station monthly temperature data are available in text format at the Met Office Hadley Centre website (http://www.cru.uea.ac. uk/cru/data/temperature/ and http://www.metoffice.gov.uk/ hadobs/). We have now developed a KML interface that enables both the gridded temperature anomalies and the weather station temperatures to be visualised and accessed within Google Earth or other Earth browsers.

The version of CRUTEM4 currently available via this Google Earth KML interface is CRUTEM4-2013-03 (downloaded from http://www.metoffice.gov.uk/hadobs/crutem4/ data/download.html, best estimate temperature anomalies version 4.2.0.0), though only complete years are included so the data end in December 2012. We plan to provide a new version as data for each calendar year become available.

The main KML file for CRUTEM4 can be downloaded from http://www.cru.uea.ac.uk/cru/data/crutem/ge/. Once opened within the Google Earth software, it renders
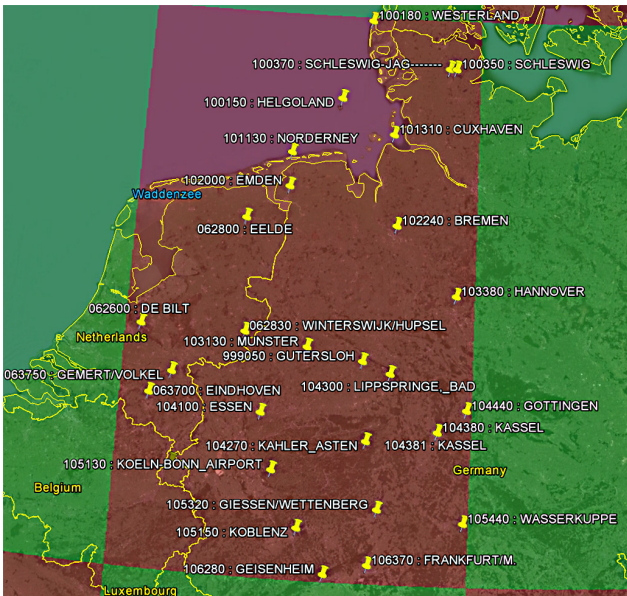
Figure 4. Image from Google Earth showing the CRUTEM4 weather station locations within a particular grid box (the yellow marker pins). The CRUTEM4 identifier number (ID) and name of each weather station is shown with white text next to each yellow marker pin. The ID is usually the WMO ID, made up of two or three digits to represent the country and the remaining digits to represent the station (in this example, the stations beginning with 06 are in the Netherlands, and with 10 are in Germany). If the station has not been allocated a WMO ID, or is from a source that has not used the WMO ID system, then it usually begins with 99 (e.g. 999050 Gutersloh in this example).

Figure 5. Image from Google Earth showing the box that appears after clicking on one of the weather station marker pins. The box shows the annual-mean temperature time series as bars (blue for values below a reference mean, red for values above; the reference mean is from the 1961–1990 period if there are at least 18 annual values within that period, otherwise it is the overall mean of the data) with a black line for the 20 yr smoothed variations. The box also shows the location of the weather station, its CRUTEM4 identifier (usually the WMO identifier) and name, and provides links to a larger image of the annual time series, to an image of the seasonal-mean temperature time series, and to a text file containing the monthly, seasonal and annual temperature values for this weather station.

all the 5° latitude by 5° longitude grid boxes that contain CRUTEM4 temperature anomaly values as a chequerboard of green and red boxes (Fig. 2). Grid boxes that do not contain any temperature anomaly values are left unshaded.

Clicking on one of the shaded grid boxes causes a "balloon" to appear that contains an image of the annual-mean temperature anomaly time series for the grid box (Fig. 3). The images are all pre-created and the KML file simply contains the uniform resource locator (URL) where the required image can be accessed on the CRU website, enabling the image to be downloaded and displayed. The balloon also includes links to a larger image of the annual time series, to an image of seasonal-mean temperature anomaly time series, and to a text file containing the monthly, seasonal and annual temperature anomaly values for this grid box. Again, these files are all pre-created and the KML file contains the URLs where these resources can be found.

The final element of the balloon is the URL for a KML file that contains details of the weather stations in the CRUTEM4 database that lie within the selected 5° latitude by 5° longitude grid box. Clicking the "stations" link causes this KML file to be downloaded and opened in Google Earth, rendering the location of these stations as a set of marker pins in the se-
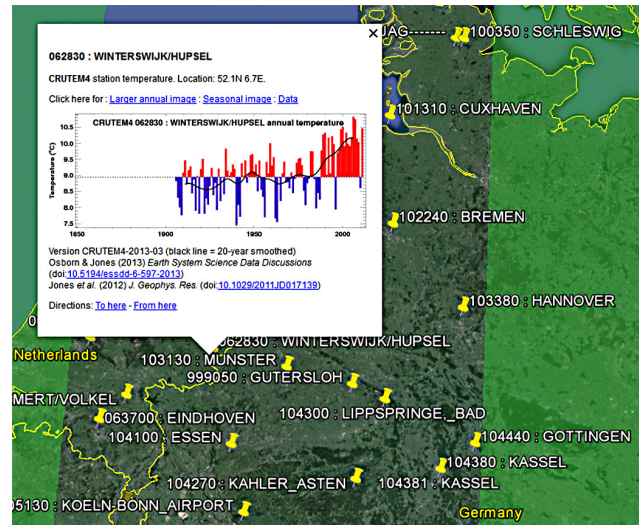
lected grid box (Fig. 4). All stations within the selected grid box are shown, even those whose data are not used in the construction of the grid-box temperature anomaly data because there is no 1961–1990 normal for the station (see Sect. 2.4).

Clicking on one of the station marker pins causes a new balloon to appear, containing some details about the station and an image of the annual-mean temperature time series for the station (Fig. 5). Links are included to a larger image of the annual time series, an image of seasonal-mean temperature time series, and to a text file containing the monthly, seasonal and annual temperature values for this station.

The KML interface allows users to locate grid boxes of interest, obtain pre-created images of the data, but also obtain the data file for the chosen grid box. All weather stations that contribute to that grid box can be located and their data values visualised or downloaded for further analysis. The grid-box and station data files use a comma-separated format to allow easy import into a spreadsheet for further analysis. We believe that this will significantly enhance the accessibility of this key climate data set, so that users can explore and extract subsets of data without the need to develop bespoke software to analyse the existing text or netCDF data files.

References

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

House of Commons Science and Technology Committee. The disclosure of climate data from the Climatic Research Unit at the University of East Anglia, Eighth Report of Session 2009–10, available from http://www.publications.parliament.uk/pa/cm/cmsctech.htm, 2010.

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

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

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., and Wigley, T. M. L.: Northern Hemisphere surface air temperature variations: 1851–1984, J. Clim. Appl. Meteorol., 25, 161–179, 1986a.

Jones, P. D., Raper, S. C. B., and Wigley, T. M. L.: Southern Hemisphere surface air temperature variations: 1851–1984, J. Clim. Appl. Meteorol., 25, 1213–1230, 1986b.

Jones, P. D., Osborn, T. J., Briffa, K. R., Folland, C. K., Horton, E. B., Alexander, L. V., Parker, D. E., and Rayner, N. A.: Adjusting for sampling density in grid-box land and ocean surface temperature time series, J. Geophys. Res., 106, 3371–3380, doi:10.1029/2000JD900564, 2001.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010, J. Geophys. Res., 117, D05127, doi:10.1029/2011JD017139, 2012.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set, J. Geophys. Res., 117, D08101, doi:10.1029/2011JD017187, 2012.

Osborn, T. J., Briffa, K. R., and Jones, P. D.: Adjusting variance for sample-size in tree-ring chronologies and other regional-mean timeseries, Dendrochronologia, 15, 89–99, 1997.

US DoE TR017 – Bradley, R. S., Kelly, P. M., Jones, P. D., Goodess, C. M., and Diaz, H. F.: A climatic data bank for Northern Hemisphere land areas, 1851–1980, U.S. Dept. of Energy, Carbon Dioxide Research Division, Technical Report TR017, 335 pp., available from http://www.cru.uea.ac.uk/publications/crurp/, 1985.

US DoE TR022 – Jones, P. D., Raper, S. C. B., Santer, B., Cherry, B. S. G., Goodess, C., Kelly, P. M., Wigley, T. M. L., Bradley, R. S., and Diaz, H. F.: A grid point surface air temperature data set for the Northern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, Technical Report TRO22, 251 pp., available from http://www.cru.uea.ac.uk/publications/crurp/, 1985.

US DoE TR027 – Jones, P. D., Raper, S. C. B., Goodess, C. M., Cherry, B. S. G., and Wigley, T. M. L.: A grid point surface air temperature data set for the Southern Hemisphere, U.S. Dept. of Energy, Carbon Dioxide Research Division, Technical Report TR027, 73 pp., available from http://www.cru.uea.ac.uk/publications/crurp/, 1986.