Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set

Colin P. Morice[1], John J. Kennedy[1], Nick A. Rayner[1] and Phil D. Jones[2,3]

Abstract

Recent developments in observational near-surface air temperature and sea-surface temperature analyses are combined to produce HadCRUT4, a new data set of global and regional temperature evolution from 1850 to the present. This includes the addition of newly digitised measurement data, both over land and sea, new sea-surface temperature bias adjustments and a more comprehensive error model for describing uncertainties in sea-surface temperature measurements. An ensemble approach has been adopted to better describe complex temporal and spatial interdependencies of measurement and bias uncertainties and to allow these correlated uncertainties to be taken into account in studies that are based upon HadCRUT4. Climate diagnostics computed from the gridded data set broadly agree with those of other global near-surface temperature analyses. Fitted linear trends in temperature anomalies are approximately 0.07 degC/decade from 1901 to 2010 and 0.17 degC/decade from 1979 to 2010 globally. Northern/southern hemispheric trends are 0.08/0.07 degC/decade over 1901 to 2010 and 0.24/0.10 degC/decade over 1979 to 2010. Linear trends in other prominent near-surface temperature analyses agree well with the range of trends computed from the HadCRUT4 ensemble members.

[1] Met Office Hadley Centre, Exeter, UK

[2] Climatic Research Unit, University of East Anglia, Norwich, UK

[3] Center of Excellence for Climate Change Research / Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

1 Introduction

This paper reports on the development of HadCRUT4, the most recent update to the HadCRUT series of observational surface temperature data sets [*Jones*, 1994, *Jones and Moberg*, 2003, *Brohan et al*., 2006]. This new version of the HadCRUT data set has been developed to incorporate updates to the land air temperature [*Brohan et al*., 2006] and sea-surface temperature (SST) [*Rayner et al*., 2006] anomaly data sets that formed the land and sea portions of HadCRUT3 [*Brohan et al*., 2006]. The land record has now been updated to include many additional station records and re-homogenized station data. This new land air temperature data set is known as CRUTEM4 [*Jones et al*., 2012]. A major update to the sea-surface temperature (SST) component of the global record has also been completed. This is known as HadSST3 [*Kennedy et al*., 2011a, 2011b]. In addition to the inclusion of additional measurements, HadSST3 includes a more thorough assessment of SST uncertainty, incorporating a more comprehensive uncertainty model, new bias adjustments and analysis of bias adjustment uncertainty.

The surface temperature analyses used to monitor climate are largely based on a similar set of temperature measurements, augmented by additional data where available. Land station records are mostly obtained from national meteorological services through World Meteorological Organization (WMO) and Global Climate Observation System (GCOS) initiatives. These station data are typically updated through monthly CLIMAT message transmissions (co-ordinated by the WMO), Monthly Climatic Data for the World (MCDW) publications, and decadally produced World Weather Records (see *Jones et al*. [2012] for details). Current data sets of historical SSTs are largely based on the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, *Woodruff et al*., 2011]), a compilation of meteorological data collected by ships and drifting and tethered buoys. Operationally, these data sets are updated using data received over the Global

Telecommunication System (GTS). Additionally, some global surface temperature analyses incorporate SST retrieved from satellite measurements. Despite the data being largely drawn from the same sources, there are small but appreciable differences between prominent near-surface temperature data sets and their derived global and regional temperature records [*Kennedy et al*., 2010].

Differences between these data sets, and derived analyses of global and regional temperature, may result from: the inclusion of additional observational data to supplement the sources mentioned above; differences in data quality control methods; applied measurement bias adjustments and data set gridding methodologies. The land and sea components of HadCRUT4 are formed by gridding temperature anomalies calculated from observations made in each box of a regular latitude/longitude grid, without using interpolation. HadCRUT4 remains the only one of the four prominent combined land and SST data sets that does not employ any form of spatial infilling and, as a result, grid-box anomalies can readily be traced back to observational records. The global near-surface temperature anomaly data set of the Goddard Institute for Space Studies (GISS) [*Hansen et al*., 2010], is again a blend of land and SST data sets. The land component is presented as a gridded data set in which grid-box values are a weighted average of temperature anomalies for stations lying within 1200km of grid-box centres. The sea component is formed from a combination of the HadISST1 data set [*Rayner et al*., 2003] with the combined *in situ* and satellite SST data set of *Reynolds et al*. [2002]. The National Climatic Data Center (NCDC) analysis is a blend of land data from the Global Historical Climate Network (GHCN) with the ERSST3b [*Smith et al*., 2008] interpolated sea-surface temperature data set, with land data in unobserved regions reconstructed using a method known as empirical orthogonal teleconnections [*Smith et al*., 2008; *Menne and Williams*, 2009; *Lawrimore et al.*, 2011]. The analysis of the Japanese Meteorological Agency (JMA) [*Ishihara et al*., 2012] is a blend of temperatures over land principally derived from GHCN

and CLIMAT reports with the optimally interpolated COBE SST data set [*Ishii et al*., 2005]. In addition to the differences arising from data set construction methodologies, differences in computed climate diagnostics, such as regional average temperatures, can result from differing approaches to compensating for non-uniform observational coverage across the globe.

The differences in temperature analyses resulting from the various approaches is referred to as "structural uncertainty": the uncertainty in temperature analysis arising from the choice of methodology [*Thorne et al*., 2005]. It is because of this structural uncertainty that there is a requirement for multiple analyses of surface temperatures to be maintained so that the sensitivity of results to data set construction methodologies can be assessed. The requirement for any given analysis is to strive to both reduce uncertainty and to more completely describe possible uncertainty sources, propagating these uncertainties through the analysis methodology to characterize the resulting analysis uncertainty as fully as possible.

So, how certain can we be of the temperature evolution observed in a given observational analysis? A detailed measurement error and bias model was constructed for HadCRUT3 [*Brohan et al*., 2006]. This included descriptions of: land station homogenization uncertainty; bias related uncertainties arising from urbanization, sensor exposure and SST measurement methods; sampling errors arising from incomplete measurement sampling within grid–boxes; and uncertainties arising from limited global coverage. The uncertainty model of *Brohan et al*. [2006] allowed conservative bounds on monthly and annual temperature averages to be formed. However, it did not provide the means to easily place bounds on uncertainty in statistics that are sensitive to low frequency uncertainties, such as those arising from step changes in land station records or changes in the makeup of the SST observation network. This limitation arose because the uncertainty model did

not describe biases that persist over finite periods of time, nor complex spatial patterns of interdependent errors.

To allow sensitivity analyses of the effect of possible pervasive low frequency biases in the observational near-surface temperature record, the method used to present these uncertainties has been revised. HadCRUT4 is presented as an ensemble data set in which the 100 constituent ensemble members sample the distribution of likely surface temperature anomalies given our current understanding of these uncertainties. This approach follows the use of the ensemble method to represent observational uncertainty in the HadSST3 [*Kennedy et al*., 2011a, 2011b] ensemble data set. There has been similar use of ensembles in other studies, e.g. in Rayner et al. [2006] to quantify uncertainties in SST biases, and in *Mears et al*. [2011] in the study of uncertainties in Microwave Sounding Unit (MSU) based measures of temperature in the upper atmosphere. For HadCRUT4, the individual ensemble members will be made available to allow the sensitivity to slowly varying observational error components to be taken into account in studies based on the data set. It should be noted that the HadCRUT4 uncertainty model only takes into account uncertainties identified in the construction of HadCRUT4, and other as yet unidentified sources of uncertainty may exist. This model cannot take into account structural uncertainties arising from data set construction methodologies. It is clear that a full description of uncertainties in near-surface temperatures, including those uncertainties arising from differing methodologies, requires that independent studies of near-surface temperatures should be maintained. We recommend that, in addition to the use of HadCRUT4, data set users consider testing the robustness of their results by comparison to other available data sets.

This paper is structured as follows. Section 2 provides an overview of the updated land station record CRUTEM4 [*Jones et al*., 2012] and SST data set HadSST3 [*Kennedy et al*., 2011a, 2011b]

from which the global analysis HadCRUT4 is formed. Section 3 describes the production of an ensemble of CRUTEM4 realizations from the *Brohan et al*. [2006] uncertainty model and briefly describes the construction of the HadSST3 ensemble data set. Section 4 describes the method for combining land and marine components to form the HadCRUT4 data set. Section 5 describes the methods used to generate time-series and their related uncertainties from the gridded HadCRUT4 data. The improvements in global coverage achieved through the inclusion of additional data in HadCRUT4 are discussed in Section 6. In Section 7 global and regional time-series computed from HadCRUT4 are presented and compared to other analyses of near-surface temperature. Section 8 concludes and describes areas in which we believe further study is required.

The data set described in this paper, HadCRUT4, and derived time series can be obtained from http://www.metoffice.gov.uk/hadobs/ and http://www.cru.uea.ac.uk/cru/data/temperature/.

2 Updates to land and sea components

Both the CRUTEM4 [*Jones et al*., 2012] and HadSST3 [*Kennedy et al*., 2011a, 2011b] data sets that respectively form the land and sea components of HadCRUT4 have been updated substantially since *Brohan et al*. [2006]. In this section, an overview of the updates is presented.

2.1 The sea-surface temperature record - HadSST3

The marine component of the HadCRUT4 global near-surface temperature data set is HadSST3 [*Kennedy et al*., 2011a, 2011b], an updated sea-surface temperature anomaly data set. In this update, additional SST observations from a number of digitization efforts have been included, new

adjustments have been developed to address recently identified biases in SST [*Emery et al*., 2001; *Kent and Taylor*, 2006; *Thompson et al*., 2008; *Kennedy et al*., 2011c] and a new model of measurement and sampling uncertainty is used.

The HadSST3 data is based upon an updated version of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS). The SST data used in HadSST2 [*Rayner et al*., 2006] was sourced from ICOADS 2.0 [*Worley et al*., 2005]. HadSST3 is based upon ICOADS 2.5 [*Woodruff et al*., 2011]. This new version of the ICOADS databank has benefitted from many newly digitised SST data obtained through record digitization efforts, such as those of *Brohan et al*. [2009], and as a result the observational coverage in HadSST3 has improved.

Core aims in the development of HadSST3 were the development of improved bias adjustments for sea-surface temperature measurements and better understanding of the uncertainties in the data. Throughout the 19th century and early 20th century, SST measurements were typically obtained by drawing buckets of water onto a ship's deck. During the 20th century the make up of the measurement network shifted towards Engine Room Intake (ERI) water temperature measurements, special insulated buckets, and the use of hull contact sensors. The end of the 20th century saw the deployment of large networks of drifting buoys and other platforms, which continue to provide more comprehensive temperature measurement coverage than was previously possible using purely ship based measurements. The various techniques used to obtain SSTs each have their own bias characteristics. SSTs obtained using buckets tend to be cooled by evaporation and by heat exchange with the air to a degree that is dependent on the construction of the bucket used. ERI measurements tend to be biased warm because of heating of water while within the ship. Observations obtained by buoys have their own characteristics and tend to be obtained at different water depths than bucket or ERI measurements. *Kennedy et al*. [2011b] conducted an assessment of

these large-scale measurement biases and developed new bias adjustments to compensate for them, along with adjustment uncertainties. The resulting bias-adjusted data set is presented as an ensemble of 100 SST anomaly data sets, each generated with different feasible bias adjustments.

Additionally, in *Kennedy et al*. [2011a] a new measurement and sampling uncertainty model was developed to accompany the bias-adjusted data. This model includes uncertainty arising from uncorrected micro-biases unique to individual ships or buoys; the unknown residual biases in each platform after applying large-scale bias adjustments. These uncertain SST micro-biases form a significant component of uncertainty in time-series derived from the gridded data.

2.2 The land surface station record – CRUTEM4

The land-surface air temperature database that forms the land component of the HadCRUT data sets has recently been updated to include additional measurements from a range of sources [*Jones et al*., 2012]. US station data have been replaced with the newly homogenized US Historical Climate Network (USHCN) records [*Menne et al*., 2009]. Many new data have been added from Russia and countries of the former USSR, greatly increasing the representation of that region in the database. Updated versions of the Canadian data described in [*Vincent and Gullett*, 1999, *Vincent et al.*, 2002] have been included. Additional data from Greenland, the Faroes and Denmark have been added, obtained from the Danish Meterological Institute [*Cappeln et al*., 2010, 2011, *Vinther et al*., 2006]. An additional 107 stations have been included from a Greater Alpine Region (GAR) data set developed by the Austrian Meteorological Service [*Auer et al*., 2001], with bias adjustments accounting for thermometer exposure applied [*Böhm et al*., 2010]. In the Arctic, 125 new stations have been added from records described in *Bekryaev et al*. [2010]. These stations are mainly

situated in Alaska, Canada and Russia. See *Jones et al*. [2012] for a comprehensive list of updates to included station records.

The error model used in the CRUTEM4 data set [*Jones et al*., 2012] is the same as that used in CRUTEM3 [*Brohan et al*., 2006].

3 Ensemble data set generation

Uncertain systematic biases in observations can lead to complex, interrelated patterns of uncertainty in a gridded observational data set. For example, measurement platforms with uncertain systematic biases moving from one grid-box to another will produce uncertainties in the gridded data set which are correlated between grid-boxes and from one month to the next. The importance of this correlation is dependent on both the magnitude of the uncertainty and the number of platforms with differing or identical uncertain biases contributing to the grid-box averages. These uncertainties are important for two reasons: correlated uncertainties do not cancel in the computation of averages of the data; and gradual changes in an observational network in which systematic biases pervade can lead to low frequency components in time-series derived from the data. Accordingly, an understanding of systematic biases in the data can be important when studying the sensitivity of scientific analyses to observational uncertainty.

Given distributions of likely measurement biases, feasible biases can be drawn from the distribution and a gridded temperature data set can be created by applying the derived bias adjustments. By repeating this procedure multiple times, drawing different bias realizations each time, an ensemble of gridded data sets is created, which together capture the complex spatial and temporal structure of uncertainties that arose from uncertainties in the required bias adjustments.

In this way, uncertainties in HadCRUT4 are expressed by providing multiple realizations of the gridded temperature anomaly data set. These represent feasible realizations of the data set, given uncertainties in measurement biases and in the applied bias adjustments. These 100 realizations are formed as one-to-one combinations of each of the 100 HadSST3 realizations with 100 realizations of the CRUTEM4 data set, as shown in Figure 1. Section 3.1 provides an overview of ensemble generation in HadSST3. The generation of ensemble members from the CRUTEM4 uncertainty model is described in Section 3.2. The method by which the HadCRUT4 data set is generated by blending the land and sea ensembles is described later in Section 4.

3.1 The HadSST3 ensemble data set

A brief overview of the HadSST3 uncertainty model and ensemble generation is presented in this section. For a full description see *Kennedy et al*. [2011a, 2011b].

3.1.1 SST bias adjustment realizations

Differences in techniques for measuring sea-surface temperature, such as the use of engine room intake (ERI) measurements, measurements from various forms of buckets or the use of drifting or tethered buoys, lead to large-scale biases in SST measurements. In HadSST3, large scale bias adjustments are applied to gridded SST anomalies to compensate for differences in measurement technique. Large-scale adjustments applied to gridded SST anomalies are derived from a number of sources: for engine room intake (ERI) measurements they are inferred from literature on the subject; bucket measurement adjustments are drawn from the model of *Rayner et al*. [2006]; and adjustments for drifting buoys are derived from matchups of coincident ship and buoy

measurements (see *Kennedy et al*. [2011b] for details). Uncertainties in bias adjustments applied to the gridded anomalies have complicated spatial and temporal correlations caused both by geographic variations in relative fractions of measurements obtained using each technique and changes in measurement network composition over time. The interdependencies of uncertainty in the HadSST3 data set are represented by creating multiple realizations of the data set, each using different realizations of bucket, ERI and drifting buoy bias adjustments. These bias adjustment realizations are created through a combination of adjustments for each measurement type, weighted by the fractions of measurements in each grid-box (which are uncertain) obtained using each of the observation techniques. These realizations are then added to the gridded temperature anomalies to create multiple realizations of the SST data set representing uncertainty in the required bias adjustments. Together these realizations span the distribution of uncertainties in the bias adjustments, encoding spatial and temporal interdependencies resulting from differing geographic distributions of measurement methods and changes in the makeup of the measurement network over time.

3.1.2 SST measurement and sampling error

In addition to large-scale bias adjustments and related uncertainties, the HadSST3 uncertainty model also incorporates uncertainties in individual measurements, inter-platform biases or micro-biases, and sampling uncertainty arising from the formation of grid-box averages from a limited number of discrete measurements. The inclusion of uncorrected micro-biases (systematic biases in individual measurement platforms around the mean bias of a specific platform type) in the uncertainty model results in uncertainties that are correlated between grid-boxes and in time. These are not explicitly included in the ensemble members and are instead provided in HadSST3 as monthly error covariance matrices describing these correlated uncertainty components.

3.2 The CRUTEM4 ensemble data set

The CRUTEM4 data set [*Jones et al*., 2012] is not an ensemble data set. However, the CRUTEM4 uncertainty model [*Brohan et al*., 2006] contains various sources of uncertainty that can be well represented through the use of the ensemble approach. Specifically, homogenization adjustment uncertainties, uncertainties in the calculation of long-term averages over the 1961-1990 climatological normal reference period, and uncertain biases arising from urbanization and sensor exposure have correlation structures that complicate the computation of uncertainties in diagnostics such as time-series computed from the grid-box anomalies. Rather than directly combine the non-ensemble CRUTEM4 data set with the ensemble HadSST3 data set, the method adopted here is to first construct an ensemble version of CRUTEM4 by drawing possible error realizations from the [*Brohan et al*., 2006] uncertainty model and combining them with station records. This allows easy calculation of uncertainty ranges in averages of grid-box anomalies arising from correlated uncertainties. It also allows straightforward blending of the land near-surface air temperature measurements with the ensemble HadSST3 data set, as is described later in Section 4.

3.2.1 Combining realizations of possible errors with station records

This section describes the manner in which error components are combined with station records to produce the ensemble members of CRUTEM4. The ensemble realizations of CRUTEM4 are drawn by perturbing the station time series and gridded anomaly values with plausible realizations of known uncertainties described by the CRUTEM4 uncertainty model that have spatial or temporal correlation structures. These are the station homogenization error ε_H, the station climatological normal error ε_N and large scale urbanization and exposure biases, ε_u, and ε_e. A schematic of the

procedure used to combine these components with land station records is shown in Figure 2. Example realizations of the uncertainty components sampled in generation of ensemble members are shown in Figure 3. Each of these uncertainty components is discussed in turn in the following sections, along with descriptions of the methods used to draw plausible realizations of each component.

In this study, the true monthly average temperature at a meteorological station T_{true} for a given month is considered to be related to the observed monthly average temperature through the following relationship:

$$T_{obs} + C_H = T_{true} + \varepsilon_{obs} + \varepsilon_H \tag{2}$$

Where T_{obs} is the observed temperature, C_H is a homogenization correction applied to remove inhomogeneities in the station record, ε_{obs} is a random measurement error and ε_H is the error in the applied homogenization correction. Each of these error components, and any temporal correlation structures they may have are discussed in detail in the following sections. Ideally, the above equation would also include the effects of urbanization and changing sensor exposure, arising from changes in enclosures used to shield thermometers from the elements. These terms are omitted at this stage as the urbanization and sensor exposure models used here are based on studies of the influence of these factors on regional averages, and the derived biases may not be representative of the influence of these factors on individual station records. These factors are instead applied to gridded temperature anomalies.

To calculate a temperature anomaly from a station's monthly average temperature, a climatological station normal is calculated. The true climatological station normal T_N is defined as follows:

$$T_N = \text{avg}_{1961\text{–}1990}(T_{obs} + C_H - \varepsilon_{obs} - \varepsilon_H) - \varepsilon_N$$
$$= \tilde{T}_N - \varepsilon_N$$

(1)

where $\tilde{T}_N$ is an estimate of the climatological station normal and ε_N is the error in this estimate arising from measurement error and the computation of normal temperatures from a finite number of years of data.

Here the climatological station normal for each month is computed over all instances of a calendar month over the 1961 to 1990 period. The true station temperature anomaly T_a is therefore given by:

$$T_a = T_{obs} + C_H - \varepsilon_{obs} - \varepsilon_H - \tilde{T}_N + \varepsilon_N .$$

(3)

As ensemble members of CRUTEM4 only include realizations of uncertainties that have temporal or spatial correlations, the random observational error ε_{obs} in a monthly station average is not included in the ensemble members. This observational uncertainty component is uncorrelated between different observations and stations and can be readily added to the ensemble if required. A realization of the true monthly station anomaly is then produced, by perturbing the observed value, as follows:

$$T_a = T_{obs} + C_H - \varepsilon_H - \tilde{T}_N + \varepsilon_N .$$

(4)

Here $T_{obs} + C_H$ represents the homogenized station temperature series provided in the CRUTEM4 database (having performed an outlier check as described by *Jones et al*. [2012]). Realizations of

ε_H are drawn as described in Section 3.2.1.1 and realizations of ε_N are drawn as described in Section 3.2.1.2.

Grid-box anomaly realizations are computed as an average of all perturbed station records for stations lying within the same 5˚ latitude by 5˚ longitude grid-box. As described in *Brohan et al.* [2006], this average is subject to a sampling error ε_s which is the error in computing a grid-box average temperature from measurements at a finite number of positions. Like the measurement error ε_{obs}, the sampling error is uncorrelated between grid-boxes and in time and so realizations of this uncertainty component are not encoded into the ensemble members. As realizations of urbanization and exposure biases represent the possible influence of these factors on regional averages, rather than on individual stations, realizations of the large-scale urbanization, ε_u, and exposure, ε_e, biases are removed from the grid-box anomalies. For an individual ensemble member, each monthly grid-box anomaly in the gridded data set is therefore computed as:

$$A^{land} = \left(\frac{1}{K} \sum_{n=1}^{K} T_{a[n]} \right) - \varepsilon_u - \varepsilon_e , \tag{5}$$

where A^{land} is a realization of the CRUTEM4 monthly grid-box temperature anomaly computed from K perturbed station anomalies $T_{a[n]}$ located within the grid-box.

The following sections describe each of the (possible) error realizations, ε_H, ε_N, ε_u, and ε_e, and each of their correlation structures in detail.

3.2.1.1 Station homogenization adjustment error

Homogenization is the process of identification and removal of artefacts in station records such as those caused by changes in measurement equipment, relocation of stations within their local area, changes in time of day of measurements, and changes in methods used to compute monthly mean temperatures. Homogenization adjustments have been applied to the land station data included in HadCRUT4 [*Jones et al*., 2012]. *Brohan et al*. [2006] compared adjusted time-series in the CRU archive to unadjusted records where unadjusted records were available. Through this comparison it was concluded that small discontinuities in station records were difficult to detect in the homogenization process and that a residual error in the homogenization process exists. This error was modelled as a zero mean Gaussian distribution with a standard deviation of $\sigma_H = 0.4^{\circ}\mathrm{C}$. Recent studies of homogenization uncertainty report broadly similar magnitudes of homogenization uncertainty [*DeGaetano*, 2006; *Menne and Williams*, 2009, *Menne et al*., 2009] and so the model of $\sigma_H = 0.4^{\circ}\mathrm{C}$ is maintained in CRUTEM4 and HadCRUT4. The assessment of the [*Brohan et al*., 2006] analysis was that homogenization step changes occurred on average every 40 years, which is the average occurrence rate used in this study. It is worth mentioning that in a study of US stations *Menne et al*. [2009] detected a more frequent average step change rate of 15 – 20 years. This difference may have arisen because of different methods for detecting required adjustments, regional differences in changing measurement practice (such as the documented large scale movement towards the use of automated stations in the US in the 1980s) or improved detection of changes owing to the density of the US network.

The ensemble approach allows the correlation structure resulting from uncertainties in the homogenization process to be encoded into the ensemble members. To generate an ensemble member, a series of possible errors in the homogenization process was created by first selecting a set of randomly chosen step change points in the station record, with each point indicating a time at which the value of the homogenization adjustment error changes. These change points are drawn

from a Poisson distribution with a 40 year repeat rate. For each period of constant homogenization adjustment error, a value of the adjustment error ε_H is then drawn from a zero mean Gaussian distribution with a standard deviation of $\sigma_H = 0.4^{\circ}\mathrm{C}$. This mimics the behaviour of undetected or residual inhomogeneities in station records, as described by *Brohan et al*. [2006]. Example realizations of plausible homogenization error for a single station are shown in Figure 3.

Note that the formulation of the homogenization model used to generate ensemble members is designed only to allow a description of the magnitude and temporal behaviour of possible homogenization errors to contribute to the calculation of uncertainties in regional averages. Change times are unknown and chosen at random, so realizations of change time will be different for a given station in each member of the ensemble. Additionally, the model used here does not describe uncertainty in adjustment of coincident one-way step changes associated with countrywide changes in measurement practice, such as those discussed in Menne et al. [2009] for US data.

3.2.1.2 Station climatological normal uncertainty

The climatological normal uncertainty represents the uncertainty in forming the calendar monthly climatological average temperatures over the 1961 to 1990 reference period used to convert temperatures into anomalies. As in [*Brohan et al*., 2006], the station climatological normal uncertainty is modelled as being totally temporally correlated for a given calendar month in all years, and uncorrelated between different calendar months. This uncertainty component is totally uncorrelated between differing stations.

In the ensemble version of CRUTEM4, a single sample of the possible climatological normal error is drawn for each station for each of the 12 calendar months of the year. These 12 realizations are

held constant for all years in the station record. Samples are drawn from a Gaussian distribution with zero mean and a standard deviation that is dependent on the number of years of data that are available at a station in the 1961-1990 reference period. For stations with at least 14 years of data in the 1961-1990 reference period, the sampling distribution of the station climatological normal error ε_N has a standard deviation of $\sigma_w / \sqrt{P}$, where σ_w is the standard deviation of observed monthly temperatures at a station (here computed over a period of 1941 – 1990) for a given calendar month, and P is the number of years of station data in the 1961-1990 reference period for that calendar month. For some stations without 14 years of data available in the normal reference period, station normals are available from the World Meteorological Organization (WMO) [*WMO*, 1996]. Where climatological station normals were obtained from the WMO, the analysis of *Brohan, et al*., [2006] found that uncertainties in climatological station normals were equivalent to about $0.3\sigma_w$. As in *Brohan et al*. [2006], uncertainties are attributed to climatological station normals obtained from the WMO by scaling σ_w by this factor.

3.2.1.3 Urbanization bias

The urbanization bias model used here is that of the CRUTEM4 data set [*Jones et al*., 2012], as described in *Brohan et al*. [2006]. It is based upon studies of the effect of urbanization on large-scale temperature anomaly averages, rather than on urbanization at specific stations. Since the review of urbanization presented in *Brohan et al*. [2006], further studies have been conducted to assess both large-scale and regional urbanization effects, many of which are summarised in a review by *Parker* [2010]. Comparisons of observations over the eastern US to dynamic reanalysis reconstructions by *Kalnay et al*. [2006] indicated an urbanization effect of 0.09˚C per decade. In China, *Jones et al*. [2008] found warming trends over the period of 1951 to 2004 of 0.08 to 0.1 ˚C per decade. A study of Japanese station records estimated an effect of approximately 0.1˚C per

decade over the 20th century [*Fujibe*, 2009]. To study the effect on global trends, *Efthymiadis and Jones* [2010] studied differences between gridded land station observations and SST in coastal regions, coming to the conclusion that the average effect of urbanization is between zero and 0.02˚C per decade across the globe, with the caveat that the upper value is a conservative estimate as temperatures over land are known to warm at a greater rate than SST. Because regional studies of urbanization have only been conducted for a limited number of regions, and because results of recent studies are compatible with the *Brohan et al*. [2006] assessment, the urbanization model used here is based upon *Brohan et al*. [2006].

The influence of urbanization on global and regional averages is modelled as a one sided uncertainty in temperature measurements; urbanization may lead to temperature measurements that are on average warmer, but not cooler than regionally representative temperatures. The value of the urbanization bias, ε_u, is assumed to have a value of 0.0˚C prior to 1900 and then increase linearly at a constant rate. This warming rate is sampled from a truncated Gaussian distribution. A realization of the warming rate is drawn from a Gaussian distribution with a standard deviation of 0.0055˚C per decade. If a negative warming rate is drawn, the warming rate is set to 0.0˚C per decade, representing the findings of a number of studies that indicate no statistically significant effect of urbanization on regionally averaged temperatures.

3.2.1.4 Exposure bias

The exposure bias component of the uncertainty model represents the uncertainty in measurement bias on a regional to global scale arising from the introduction of new varieties of measurement sensor enclosures throughout history. Examples of this are the changes in biases in hemispheric

averages arising from the transition from thatched enclosures and north wall (in the NH) facing exposures to Stevenson-type shelters.

As in *Brohan et al*. [2006], the exposure bias model followed is that of [*Folland et al*., 2001], which is derived from the results of [*Parker*, 1994]. For grid-boxes in the latitude range of 20˚S-20˚N a 1σ uncertainty of 0.2˚C is assumed prior to 1930. This then decreases linearly towards a value of zero in 1950. For stations that lie outside of 20˚S-20˚N the exposure bias uncertainty takes a value of 0.1˚C prior to 1900, decreasing linearly to zero by 1930. Ensemble members are generated from this model by drawing a single random number from a standard normal distribution for each ensemble member, which is then scaled for each grid-box by the appropriate 1σ uncertainty range based on latitude and time as described above to produce an exposure bias realization, ε_e.

There is scope for construction of a more detailed exposure bias model in future. Seasonal cycles in exposure biases were identified by *Parker* [1994] for various enclosure types. *Moberg et al*. [2003] found evidence of bias seasonality in Swedish station records. *Böhm et al*. [2010] derived bias adjustments for the Greater Alpine Region prior to 1870. These adjusted data for the Greater Alpine Region are incorporated into HadCRUT4, although uncertainties in the applied bias adjustments are not explicitly accounted for in the HadCRUT4 error model. If a more detailed exposure bias model is to be constructed for the global data set then further study of seasonality in regional exposure biases is required.

3.2.2 Measurement and sampling error

The models of random measurement error, ε_{obs}, and sampling error, ε_s, used in this analysis are exactly as described in *Brohan et al*. [2006]. Sampling error has been recomputed using the *Jones et al*. [1997] method, with inter-grid-box correlations recomputed from the CRUTEM4 station data. As the model of measurement and sampling error used here for land stations has no temporal or spatial correlation structure, there is no need for the use of an ensemble approach to describe these error components. These components are instead attributed to each ensemble member when time series are computed.

4 Blending land and sea components

CRUTEM4 and HadSST3 overlap in grid-boxes which are partially land and partially sea. Here we combine land air temperature and SST anomalies and their measurement and sampling uncertainties.

4.1 Fractional area weighting

The blending approach adopted here differs from that used in the *Brohan et al*. [2006] data set. Here, land and sea components are combined at a grid-box level by weighting each component by fractional areas of land and sea within each grid-box, rather than weighting in inverse proportion to error variance. This approach has been adopted to avoid over representation of sea temperatures in regions where SST measurements dominate the total number of measurements in a grid-box. The grid-box average temperature $A_{[i]}$ for grid-box i is formed from the grid-box average temperature anomalies $A_{[i]}^{land}$ for the land component, $A_{[i]}^{SST}$ for the SST component, and the fractional area of land in the grid-box $f_{[i]}$ as follows:

$$A_{[i]} = f_{[i]} A_{[i]}^{land} + (1 - f_{[i]}) A_{[i]}^{SST}$$

(6)

Coastal grid-boxes for which the land fraction is less than 25% of the total grid-box area are assigned a land fraction weighting of 25%. Here, we are making the assumption that land near-surface air temperature anomalies measured in grid-boxes that are predominantly sea covered are more representative of near-surface air temperature anomalies over the surrounding sea than sea-surface temperature anomalies. These fractions ensure that islands with long land records are not swamped by possibly sparse SST data in open ocean areas (where the island is only a small fraction of the total grid-box area).

4.2 Blending ensemble members

To produce the gridded temperature anomaly ensemble, the 100 land near-surface air temperature anomaly ensemble members have been blended with 100 SST anomaly ensemble members on a one-to-one basis. This results in a set of 100 realizations of the global temperature anomalies with respect to a 1961 to 1990 reference period on a monthly grid of 5 degrees latitude by 5 degrees longitude. Example fields for nine ensemble members of HadCRUT4 are shown in Figure 4.

In addition to the 100 ensemble members, there are two additional uncertainty components: the contributions to grid-box uncertainty from the uncorrelated measurement and sampling uncertainties of the land component, CRUTEM4, and those from the partially correlated measurement and sampling uncertainties of the sea component, HadSST3. For a grid-box *i*, the combined uncertainty arising from these two measurement and sampling error components, $\sigma_{[i]}^{land}$ and $\sigma_{[i]}^{SST}$, is:

$$\sigma_{[i]}^{combined} = \sqrt{\left(f_{[i]}\right)^2 \left(\sigma_{[i]}^{land}\right)^2 + \left(1 - f_{[i]}\right)^2 \left(\sigma_{[i]}^{SST}\right)^2} .$$

(7)

The first of the terms under the square root, is the contribution of land measurement and sampling uncertainty to the grid-box error variance and is totally uncorrelated between grid-boxes. The second term is the contribution from SST measurement and sampling uncertainty, which is correlated between grid-boxes. For these uncertainties in the SST component to be propagated into regional averages, it is necessary to compute global error covariance matrices of SST uncertainty contributions, weighted by fractional areas of SST. In this weighting scheme, cross-covariances $C_{[i,j]}$ between grid-boxes i and j of the HadCRUT4 grid-box measurement and sampling uncertainty are computed from the HadSST3 cross-covariances $V_{[i,j]}$ as follows:

$$C_{[i,j]} = (1 - f_{[i]})(1 - f_{[j]})V_{[i,j]},$$

(8)

which is equal to the grid-box error variance arising from SST measurement and sampling uncertainty for grid-box i when $i = j$. The above equation defines the elements of HadCRUT4 error covariance matrices describing grid-box uncertainty arising from SST measurement, sampling and micro-bias uncertainty. The construction of uncertainties in time series derived from the gridded data is described in Section 5.

5 Calculation of global and regional time-series

5.1 Anomaly time series

Monthly regional average temperature anomaly time series for each ensemble member are computed as weighted averages of the gridded temperature anomalies in the region of interest. Grid-box weights are chosen to be proportional to grid-box area. Using the grid-box temperature

anomalies $A_{[i]}$ and weights $w_{[i]}$, a monthly regional average temperature anomaly $\overline{A}$ is computed over N grid-boxes with non-missing data as:

$$\overline{A} = \sum_{i=1}^{N} w_{[i]} A_{[i]} \,,$$

(9)

where the weights $w_{[i]}$ of data filled grid-boxes are normalised to sum to one. To compensate for different sampling of the northern and southern hemispheres, global averages $\overline{A}_G$ are computed from the northern and southern hemisphere averages, $\overline{A}_{NH}$ and $\overline{A}_{SH}$, as:

$$\overline{A}_G = \frac{\overline{A}_{NH} + \overline{A}_{SH}}{2} \,.$$

(10)

Annual, seasonal or other multi-month time series are computed as a simple average of the monthly time-series. Annual averages are computed over $M = 12$ months as:

$$\overline{A}_{annual} = \frac{\mathrm{sum}(\overline{A}_{monthly})}{M} \,.$$

(11)

Note that the order of averaging in this method is different from the method of *Brohan et al.*, [2006], in which annual anomalies were calculated by first computing annual averages of temperatures in each grid-box and then computing a grid-box area weighted average of the annual temperature field. The two methods place different weight on anomalies in grid-boxes in which observations are not available for all months. Resulting differences in annual averages are small in comparison to the computed uncertainties.

5.2 Uncertainties in calculated time-series

Measurement and sampling uncertainties are not included in the individual ensemble members and are instead handled analytically in computation of temporal and spatial averages. Note that here, sampling uncertainty is the error due to under-sampling of individual grid boxes and is distinct from coverage uncertainty, which relates to under-sampling of regions by grid-boxes containing measurements.

5.2.1 Land station measurement and sampling uncertainties

As the land component of grid-box measurement and sampling uncertainties is completely uncorrelated between grid-boxes, the resultant uncertainty in monthly regional averages, σ^u, is computed from N grid-boxes, with grid-box measurement and sampling errors of $\sigma^u_{[i]}$, as follows:

$$\sigma^u = \sqrt{\sum_{i=1}^{N} \left(w_{[i]} \sigma^u_{[i]} \right)^2}$$

(12)

To compute land measurement and sampling uncertainty in global averages, σ^u_G, uncertainties in northern hemisphere and southern hemisphere regional averages are first computed using the above equation. These uncertainties are denoted σ^u_{NH} and σ^u_{SH}. The uncertainty in global averages due to land measurement and sampling error is then computed as:

$$\sigma^u_G = \sqrt{\left(0.5\sigma^u_{NH}\right)^2 + \left(0.5\sigma^u_{SH}\right)^2}$$

(13)

As measurement and sampling uncertainties in land station data also have no temporal correlation structure, land contributions to measurement and sampling uncertainties in annual averages are computed from uncorrelated uncertainties in monthly regional averages as follows:

$$\sigma^{u}_{annual} = \frac{\sqrt{\mathrm{sum}\left((\sigma^{u}_{monthly})^{2}\right)}}{M}.$$

(14)

5.2.2 Correlated SST measurement and sampling uncertainties

When measurement and sampling uncertainties in the monthly gridded temperatures have a complicated pattern of grid-box to grid-box correlations, the uncertainties are represented by error covariance matrices. This applies to uncertainties in SSTs from 1981 onwards. Although error covariance matrices for HadSST3 are available prior to 1981, the entries for these covariance matrices are incomplete, owing to incomplete metadata describing individual historical ships' call signs, which are required to construct spatial correlation patterns. Here, we include an update to HadSST3 which contains information on modern ships' callsigns after November 2007. Thus we are able to extend the method of *Kennedy et al*. [2011a] to calculate error covariance matrices after 2006.

Uncertainty in a monthly regional average temperature anomaly, arising from correlated measurement and sampling uncertainties σ^{c}, is computed from the error covariance matrices as follows:

$$\sigma^{c} = \sqrt{\mathbf{b}^{T}\mathbf{C}\mathbf{b}},$$

(15)

where $\mathbf{b}$ is a vector of normalised weights with $\mathbf{b}^{T} = \left[w_{[1]}, \cdots, w_{[i]}, \ldots w_{[n]}\right]$, and $\mathbf{C}$ is the error covariance matrix with elements equal to the cross-covariances between grid-boxes arising from

SST measurement and sampling uncertainty, weighted by their fractional grid-box areas of sea. Here the elements of $\mathbf{b}$ are zero where there is neither a land nor sea measurement contributing to the grid-box. Otherwise they are proportional to grid-box area, such that the weights are normalised to sum to one.

For calculation of uncertainties in global averages from the error covariance matrix, the weights are stored in a matrix $\mathbf{B}$ which is formed as:

$$\mathbf{B} = \begin{bmatrix} \mathbf{w}_{NH} & \mathbf{0} \\ \mathbf{0} & \mathbf{w}_{SH} \end{bmatrix}, \tag{16}$$

where $\mathbf{w}_{NH}$ is a normalised vector of weights for northern hemisphere grid-points and $\mathbf{w}_{SH}$ is a normalised vector of weights for southern hemisphere grid points, and $\mathbf{0}$ is a vector of $N/2$ zeros. The weight vectors should contain zero entries at locations relating to grid-points in the hemisphere with missing data and non-zero values at entries relating to any grid-point in the hemisphere at which there is temperature data. An error covariance matrix $\mathbf{C}_{NS}$ for the hemispheric averages is then computed as:

$$\mathbf{C}_{NS} = \mathbf{B}^T \mathbf{C} \mathbf{B} \tag{17}$$

This covariance matrix contains the error variances of the northern and southern hemisphere averages on its diagonal and the cross-covariance of the hemispheric values in the off diagonal entries. The global uncertainty σ_G^c is then computed as:

$$\sigma_G^c = \left(\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}^T \mathbf{C}_{NS} \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \right)^{\frac{1}{2}} \tag{18}$$

Prior to 1982 insufficient metadata is available to adequately account for the full correlation structure of measurement and sampling uncertainty for SST. Our approach for handling correlation in computation of uncertainties in regional averages follows the method of [*Kennedy et al*., 2011a]. For years following 1981, for which the number of ships with unique call-signs is large compared to the number of unidentifiable ships, *Kennedy et al*. [2011a] calculated the ratios of uncertainties obtained using the full error model to those calculated from just the diagonal entries of the monthly error covariance matrices. *Kennedy et al*. [2011a] found that prior to 1982 the SST contribution to uncertainties in regional averages can be well approximated by scaling the uncertainties calculated from the diagonal of the covariance matrices by set scale factors: global, 2.2; northern hemisphere, 1.9; southern hemisphere, 2.2; tropics, 2.2. These scale factors are used here to compute the contribution of this uncertainty component to uncertainty in regional averages prior to 1982.

Following *Kennedy et al*. [2011a], the measurement and sampling uncertainty in regional averages incorporating SSTs is modelled as having a temporal correlation structure, arising from uncorrected biases persisting in measurements from individual measurement platforms. The computation of this error component for annual averages is based on an assumed effective number of independent monthly averages in a year, n_{eff}. Using this methodology, the contribution of SST measurement and sampling errors to uncertainties in annual average anomalies is computed from those of monthly regional averages as:

$$\sigma_{annual}^c = \sqrt{\frac{\mathrm{sum}\left(\left(\sigma_{monthly}^c\right)^2\right)}{12 n_{eff}}}.$$

(19)

The value of n_{eff} used here is that computed by *Kennedy et al*. [2011a] for annual averages of HadSST3 data, $n_{eff} = 2.25$.

This model of uncertainty in SST measurements assumes that the effects of micro-biases on SST anomaly time-series are autocorrelated. If realizations of this autocorrelated uncertainty component are required, for example if the influence of temporally correlated uncertainties is to be taken into account in fitting trends to time series, realizations should be drawn taking this autocorrelation into account. Time series of possible measurement and sampling error in SSTs should be drawn from a zero mean distribution with an error covariance matrix with elements:

$$\mathrm{cov}(\overline{A}_{[k]}, \overline{A}_{[l]}) = \sigma^{c}_{[k]}\sigma^{c}_{[l]}\mathrm{corr}(\overline{A}_{[k]}, \overline{A}_{[l]}),$$

(20)

where $\overline{A}_{[k]}$ is a regional average for month k, $\overline{A}_{[l]}$ is a regional average for month l, $\mathrm{cov}(\overline{A}_{[k]}, \overline{A}_{[l]})$ is the autocovariance between them and $\mathrm{corr}(\overline{A}_{[k]}, \overline{A}_{[l]})$ is their autocorrelation. Here, autocorrelations between months take a value of $\mathrm{corr}(\overline{A}_{[k]}, \overline{A}_{[l]}) = \varphi^{|k-l|}$, where the correlation parameter φ is equal to $\varphi = 0.77$.

5.3 Coverage uncertainty

An additional component of uncertainty arises from the computation of spatial and temporal averages using gridded anomaly fields in which not all grid-boxes are populated with measurements. For HadCRUT4, the coverage uncertainty calculation follows the same method as that described in *Brohan et al*. [2006]. To compute coverage uncertainty in HadCRUT4 time-

series, NCEP reanalysis [*Kalnay et al*., 1996] near-surface temperatures are sub-sampled to HadCRUT4 coverage and differences are computed between averages calculated using NCEP reanalysis temperature anomalies with global coverage and with reduced (sub-sampled) coverage. For annual/monthly series, the coverage uncertainty for a given year/month is estimated by firstly applying the observational coverage for that year/month to every year/equivalent calendar month in the reanalysis. The required average is then computed for each year/equivalent calendar month in the reanalysis for both the sub-sampled and complete data, and residuals between these averages are computed. For any given observational coverage, the coverage uncertainty is estimated as the standard deviation of these residuals.

6 Improvements to global coverage

Both the land and sea components of HadCRUT4 have benefited from additional historical temperature data, as described in Section 2. Many of these additional measurements are from regions of the globe that were poorly represented in *Brohan et al*. [2006]. The resulting improvement in global coverage can be seen in Figure 5. Much of the improvement in coverage in the early record is due to the digitization of additional SST data. The new land station data sourced for CRUTEM4 has greatly improved observational coverage across Russia. Arctic coverage has improved notably (particularly in Russia and Canada) throughout the record. Measurement coverage in the Southern Ocean and the Antarctic remains sparse.

7 Discussion of global and regional time-series

7.1 Global time-series

Monthly, annual and decadally-smoothed global-average temperature anomaly time-series from HadCRUT4 are shown in Figure 6, along with uncertainties in the time-series arising from measurement and sampling error, bias uncertainties (uncertainty in homogenization error, sensor exposure, urbanization, and SST bias adjustments), and incomplete observational coverage. The relative magnitudes of the various uncertainty components depend on the area and time scale considered, as discussed in this section.

Measurement and sampling uncertainties in HadCRUT4 are a combination of totally uncorrelated measurement and sampling uncertainties in the land station record and measurement and sampling uncertainties in sea-surface temperatures that have both spatial and temporal correlation resulting from micro-biases in individual ships and buoys. These micro-biases in marine observations produce uncertainties in gridded sea-surface temperatures that may be dependent across multiple grid-boxes locations and times (see Section 5.2.2). These correlated uncertainties now form a large contribution to the uncertainty in both spatially and temporally-averaged temperature anomaly time series. However, as autocorrelation lengths in SST measurement uncertainty are relatively short in comparison to those of large-scale bias adjustment uncertainty, the measurement uncertainty in SSTs tends to reduce in the computation of annual averages and decadally smoothed series.

Bias-related uncertainties include contributions from both CRUTEM4 and HadSST3. From CRUTEM4, this includes uncertainties in homogenization adjustments applied to the station records, in the calculation of long-term averages for the 1961 to 1990 reference period and the influence of urbanization and changes in sensor exposure. From HadSST3, this comprises uncertainties in sea-surface temperature bias adjustments. Because the uncertainties in the effect of

urbanization, land sensor exposure uncertainty and SST bias adjustment uncertainty are strongly related over large spatial scales, these reduce little when producing regional averages. For this reason, the bias uncertainty is a large fraction of the total uncertainty in the global average temperature anomaly time-series. All components of bias uncertainty are also strongly correlated over long time scales, and so tend to reduce little in the computation of annual and decadally smoothed averages when compared to measurement and sampling uncertainties.

Coverage uncertainty represents the range of likely errors in regional averages computed from data with incomplete spatial coverage. Autocorrelation exists in the coverage uncertainty because of the persistence of weather patterns in unobserved regions and because measurement coverage does not change dramatically from month to month. As the coverage uncertainty is computed by sub-sampling reanalysis data, i.e. using a measurement-assimilating dynamical model (see Section 5.3), this autocorrelation is captured in the coverage uncertainty so long as typical persistent weather patterns in the reanalysis data used are representative of the real world in unobserved regions. Coverage uncertainty is a large component of uncertainty at monthly time-scales and continues to be a large component of uncertainty in annual and decadally smoothed series, despite the improved observational coverage in HadCRUT4.

7.2 Comparison to HadCRUT3 global time series

The improvements in HadCRUT4, including the greater number of observations, the new sea-surface temperature bias adjustments and the updated sea-surface temperature uncertainty model have resulted in a refined time-series of global average temperatures. Figure 7 shows the annual time series, with 95% confidence intervals, for HadCRUT4 compared to the equivalent series for HadCRUT3.

Refinements to the bias adjustments have altered the time series most significantly in the period from the mid 1940s to the end of the 1960s. During this period, the HadCRUT4 median lies close to, or just outside of, the upper confidence limit of the HadCRUT3 time series. The period from the mid-1940s to the 1960s is warmer in HadCRUT4 than in HadCRUT3, largely as an effect of the new bias adjustments that have been applied to the sea-surface temperature data. These account for a large number of uninsulated bucket observations in the International Comprehensive Ocean-Atmosphere Data Set between 1945 and 1970 (see *Kennedy et al*. [2011b] for details).

Further differences between the HadCRUT4 and HadCRUT3 time series can be seen in recent years. Both CRUTEM4 and the HadSST3 median indicate warmer temperatures in the last 10 years than in the previous version of each data set. This results from the improved measurement coverage in CRUTEM4, particularly in Asia and at high latitudes in the northern hemisphere, and from the new bias adjustments applied in HadSST3 to account for the effect of the shift from ship based measurements to the use of buoys. However, the difference in these recent temperatures between HadCRUT4 and HadCRUT3 is small in comparison to the uncertainties in global annual temperature estimates.

The size of the uncertainty range in HadCRUT4 is typically similar to or slightly larger than that of HadCRUT3, despite the increased number of stations included in the data set. This largely stems from the inclusion of interdependencies of sea-surface temperature measurement uncertainties arising from SST micro-biases. Because of the inclusion of these interdependencies, when temperature anomalies are averaged globally their uncertainties do not reduce to the same degree as they were considered to in HadCRUT3. This offsets the reduction in coverage uncertainty achieved through the inclusion of additional records in HadCRUT4.

7.3 Regional time-series

Monthly and annual average and decadally smoothed time-series and associated uncertainties have been computed from HadCRUT4 for the northern hemisphere (Figure 8), southern hemisphere (Figure 9) and the tropics (30˚S – 30˚N) (Figure 10). In all three regions, the contributions of the measurement and sampling error are greater than was the case in *Brohan et al*. [2006], owing to the inclusion of correlated sea-surface temperature measurement error in the uncertainty model. Measurement and sampling error in the southern hemisphere and the tropics are a larger fraction of the total uncertainty than in the northern hemisphere. This arises from historical sea-surface temperature measurements in the southern hemisphere and in the tropics being obtained by relatively few ships in comparison to the northern hemisphere [*Kennedy et al*., 2011a]. Because fewer measurements contribute to regional averages than to global averages, and because interdependence of errors in sea-surface temperature measurements tends to be strongest for measurements that are locally close, measurement and bias-related uncertainties in regional averages tend to be larger than for global averages.

Uncertainties arising from limited coverage remain a major component of uncertainty in regional averages. Large coverage uncertainties in the northern hemisphere monthly averages likely arise from the scarcity of measurements at the highest latitudes, i.e. in the Arctic Ocean. Measurement coverage of the southern hemisphere has not improved significantly in HadCRUT4. Coverage uncertainty remains the largest component of uncertainty here, due to poor coverage of the Antarctic and Southern Ocean, as well as only sporadic coverage in parts of South America and Africa. In the tropics, temperature anomalies tend to vary little over large distances. Measurement coverage over the ocean is generally good in this region in HadCRUT4. As a result, coverage uncertainties tend to be small for the tropics.

7.4 Comparisons to other global temperature analyses

Figure 11 shows a comparison of HadCRUT4 time series with three other analyses of global temperatures: that of NASA's Goddard Institutes of Space Studies (GISS) [*Hansen et al*., 2010], that of NOAA's National Climatic Data Center (NCDC) [*Smith et al*., 2008; *Menne and Williams*, 2009; *NCDC*, 2010] and that of the Japanese Meteorological Agency (JMA) [*Ishihara et al*., 2012]. The depicted series largely rely on the same core set of measurements, with the addition of some supplementary records in each analysis (see references for details). Although the bulk of the measurement records in each data set are the same, there are differences in data set construction methodologies and time series calculation methods, as summarised in *Kennedy et al*. [2010].

Despite these differences, the data sets are in broad agreement about large scale surface temperature development. Temperatures in HadCRUT4 are typically warmer than other analyses from the mid 1940s through to around 1960 in global, hemispheric and tropical time series, with NCDC and JMA analyses lying outside of the uncertainty range of HadCRUT4 for much of this period. The difference between HadCRUT4 and other data sets in this period is largely due to the bias adjustments applied in HadSST3 to account for a shift from ERI based SST measurements to the use of uninsulated buckets in this period [*Thompson et al*., 2008], the effects of which can be seen in comparisons of HadSST3 with other SST data sets in *Kennedy et al*. [2011b]. The GISS, NCDC and JMA data sets do not include such bias adjustments in this period.

In Figure 11, least squares linear trends in time series are shown for the periods of 1901 to 2010 and 1979 to 2010. Trends in HadCRUT4 global average temperatures are 0.074˚C per decade over 1901 to 2010 and 0.169˚C per decade over 1979 to 2010. Northern hemisphere/southern

hemisphere trends for HadCRUT4 are 0.077/0.071˚C per decade over 1901 to 2010 and 0.241/0.096˚C per decade over 1979 to 2010. The uncertainty ranges shown for HadCRUT4 trends are 95% confidence intervals in the trends calculated from the 100 ensemble members of the HadCRUT4 series. These do not include uncertainty in trends computed from any auto-regressive component of residual departures from the computed trends. We neglect these here, since these uncertainties are common to all data sets and tell us nothing about differences between them. The HadCRUT4 error bars indicate that autocorrelated uncertainty components in the measurement data (which are bias related) result in uncertainties in linear trends that are small in comparison to observed trends for all four regions shown. Over this long time period, computed trends are most sensitive to autocorrelated uncertainties with long correlation lengths. Uncertainties in trends over the 1901-2010 period are therefore most likely to arise from uncertainties in the influence of land station sensor exposure biases in the early 20th century, before the introduction of Stevenson screens, uncertainties in the impact of urbanization on regional temperature averages and uncertainties in bias adjustment for each type of measurement platform in the slowly changing SST measurement network. The similarity between trends in the four data sets over this period indicates that, although the different analyses produce differing representations of temperature in individual years, the observed trends are robust to the choice of data set over timescales of about a century.

Uncertainties in short term trends from 1979 to 2010 are larger than in the 1901 to 2010 trends. The influence of land station homogenization and the contribution of SST micro-biases towards measurement uncertainty in SSTs are likely to be more important over this shorter time scale, particularly in the southern hemisphere, where fewer independent measurements are used to compute time-series than for global and northern hemisphere series. Differences in trends in each temperature data set are larger for 1979 to 2010 than for 1901 to 2010 for all series. This may be related to the different observational coverage and methods used to represent temperatures in

unobserved regions in each data set. In the northern hemisphere and global series, the differences in trends are greatest, which is likely to be related to different coverage of the Arctic, a region in which temperature change is believed to be more rapid than the global average [*Bekryaev et al*., 2010]. The cause of differences between JMA trends and trends in other data sets in this time period may be related to the reduced spatial coverage of the JMA data set over land in comparison to the other data sets and the use of optimal interpolation in the SST portion of the JMA data set, a method that is known to suppress temperature anomalies and so underestimate climate change [*Hurrell and Trenberth*, 1999].

To remove the influence of different global coverage from the series, Figure 12 shows time series for each of the four observational data sets with observational coverage reduced to the minimum coverage that exists in all four data sets (co-locating). Additionally, to remove the influence of differing time series calculation methodologies, each series is computed using the methods described in Section 5. Co-locating the data sets has a most prominent affect on the GISS series, indicating that a large proportion of the difference between GISS and the other data sets results from differences in measurement coverage and the extrapolation of data into unobserved regions in the GISS data set. The reduction of measurement coverage has the most profound influence on 1979 to 2010 trends in the GISS data set in the northern hemisphere. In each data set, trends over the 1901 to 2010 period are largely in agreement. In trends for 1979 to 2010 there is less agreement between data sets. Although co-location reduces the spread in linear trends in the 1979 to 2010 period, JMA trends in the 1979 to 2010 period remain suppressed in comparison to other data sets over this time period, and lie outside of the uncertainty range of HadCRUT4. This implies that trend differences can result from differences in data set construction methodologies.

8 Conclusions

The updated analysis has refined, but not significantly altered, our understanding of the evolution of the climate since 1850. The inclusion of new bias adjustments for marine data has resulted in warmer temperatures in the mid 20th century in comparison to previous studies of historical temperature observations. The inclusion of new land station data at high latitudes and the inclusion of improved SST bias adjustments have resulted in a warming of years in the late 20th century/early 21st century.

Studies of uncertainties in near-surface temperature measurements have identified correlation structures in measurement uncertainties that translate into correlated uncertainties in derived data sets. Because an ensemble of HadCRUT4 data sets has been constructed based upon analysis of correlation structures in uncertainties, it is possible to assess the sensitivity of scientific analyses to these uncertainties by applying the analysis to each individual ensemble member. This kind of analysis has not previously been possible for global surface temperature data sets because spatially and temporally correlated uncertainties were not well enough defined and uncertainties in gridded data were not expressed in a manner that allowed the description of uncertainties with complex interdependencies.

The ensemble technique allows scientific analyses based on HadCRUT4 to explicitly explore sensitivities to observational uncertainties that have a complex spatial correlation structures and low frequency biases. The Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) [*IPCC*, 2007, Chapter 9] recognised the need to assess the influence of systematic observational error on climate change detection, noting that few detection studies have explicitly considered the influence of observation uncertainties and that these uncertainties may be important

for the detection of temperature changes averaged over small regions. Since this assessment, further study of observation biases in SST measurements [*Kennedy et al*., 2011a, 2011b] has made the need to assess sensitivity to these biases all the more necessary. To this date, there remain few detection studies that consider observational uncertainty, but examples are that of *Hegerl et al*. [2001] and a recent study into detection and attribution sensitivities to the choice of near-surface temperature set used [*Jones and Stott*, 2011]. Through use of ensemble data sets, more detailed studies of the sensitivity of climate change detection and attribution studies to observational uncertainty should be possible.

Improvements to the characterization of uncertainties in the land portion of the HadCRUT4 data set would be greatly assisted by greater access to station metadata, full knowledge of applied homogenization methods or access to uncorrected station records. With access to uncorrected station records and metadata describing station histories, *Menne et al*. [2009] found a higher average occurrence rate of step changes in US station records than is represented by the global parameter value used in this study. Additional research at a regional level, with supporting station metadata, would allow the assessment of whether these results are indicative of station record characteristics in other regions, providing more information on which to base choices of uncertainty model structure and parameters. At present, sufficient metadata are not available and studies over small regions are too few for uncertainties in land station homogenization, urbanization and exposure biases to be adequately described on an individual grid-box level. In a similar fashion, the characterization of spatial and temporal correlations in SSTs is limited by missing ship call-sign information prior to 1981. Without this information, the SST uncertainties cannot be constructed in a manner that fully represents the relationships between intra-platform micro-biases in gridded SST observations. Instead, uncertainty scaling parameters are derived to accommodate spatial correlations in regional averages in periods in which there are insufficient metadata to produce

complete measurement and sampling covariance matrices. Further digitization efforts are needed to rescue relevant information.

The effect of limited observational coverage remains uncertain, particularly with regard to the role of Arctic amplification and the capability to sample any potentially large variability in polar temperatures with available measurements. The additional high latitude temperature series sourced for CRUTEM4 have allowed improved coverage in historical land data. However, future monthly data set updates will have reduced coverage because updates to these station records will not be available in near real time. This will result in a reduced capability to monitor polar temperatures, and a possible cool bias in northern hemisphere temperatures, until updates to these series become available.

The assessment of uncertainties in HadCRUT4 is based upon the assessment of uncertainties in the choice of parameters used in forming the data set, such as the scale of random measurement errors or uncertainties in large-scale bias adjustments applied to measurements. This model cannot take into account structural uncertainties arising from fundamental choices made in constructing the data set. These choices are many and varied, including: data quality control methods; methods of homogenization of measurement data; the choice of whether or not to use *in situ* measurements or to include satellite based measurements; the use of sea-surface temperature anomalies as a proxy for near-surface air temperature anomalies over water; choices of whether to interpolate data into data sparse regions of the world; or the exclusion of any as yet unidentified processing steps that may improve the measurement record. That the reduction of the four data sets compared in Section 7.4 to the same observational coverage does not resolve discrepancies between time series and linear trends is evidence that choices in analysis techniques result in small but appreciable differences in derived analyses of surface temperature development, particularly over short time scales. As these

differences are not captured by the HadCRUT4 uncertainty model, it is important that multiple temperature data sets are maintained so that the sensitivity of studies based on historical temperature records to data set construction methodologies can be explored. This requirement is recognised in the upper air observation community ("No matter how august the responsible research ground, one version of a data sets cannot give a measure of the structural uncertainty inherent in the information", [*Thorne et al*., 2011]) and applies no less to near-surface temperature records.

Acknowledgements

This work was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). P.D. Jones has been supported by the USDoE (Grant DE-SC0005689) and by the University of East Anglia. We thank Philip Brohan, Gareth Jones, David Parker and Peter Stott for helpful discussions and suggestions contributing to this project.

References

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

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

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

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

Brohan, P., R. Allan, J.E. Freeman,, A.M. Waple, D. Wheeler, C. Wilkinson and S. Woodruff (2009), Marine observations of old weather, *Bulletin of the American Meteorological Society*, 90(2), pg 219-230

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

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

DeGaetano, A.T. (2006), Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate*, 19, 838-853

Efthymiadis, D. and P.D. Jones, (2010), Assessment of maximum possible urbanization influences on land temperature data by comparison of land and marine data around coasts. *Atmosphere* 2010, 1(1), 51-61; doi:10.3390/atmos1010051.

Emery, W.J., D.J. Baldwin, P. Schlussel R.W. Reynolds (2001), Accuracy of in situ sea surface temperature used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, 106(C2), 2387-2405

Folland, C.K., N.A. Rayner, S.J. Brown, T.M. Smith, S.S.P. Shen, D.E. Parker, I. Macadam, P.D. Jones, R.N. Jones, N. Nichols and D.M.H. Sexton (2001), Global temperature change and its uncertainties since 1861, *Geophysical Research Letters*, 28(13), pg 2621-2624

Fujibe, F. (2009), Detection of urban warming in recent temperature trends in Japan, *Int. J. Climatol*., 29, 1811-1822, doi: 10.1002/joc.1822

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

Hegerl, G. C., P.D. Jones and T.P. Barnett (2001), Effect of observational sampling error on the detection of anthropogenic climate change, *Journal of Clim*ate, 14, 198-207.

Hurrell, J.W. and K.E. Trenberth (1999), Global sea surface temperature analyses: Multiple problems and their implications for climate analysis, modelling and reanalysis, *American Meteorological Society*, 80, 2661-2678

Ishihara, K., T. Umeda,, Y. Oikawa, H. Saitou, H. Murai, S. Hirahara, S. Hagiya, and S. Kobayashi (2012), Long-term Change of Global Average Surface Temperature with the JMA Combined Land and Ocean Temperature Data Set (JMATMP1), *Journal of the Meteorological Society of Japan (Submitted)*

Ishii, M. and A. Shouji, and S. Sugimoto, and T. Matsumoto (2005). Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection. *International Journal of Climatology*, 25(7), 865-879

IPCC, (2007), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Jones, G.S. and P.A. Stott (2011) Sensitivity of attribution of near surface temperature warming to the choice of observational data set, Geophysical Research Letters, 38, L21702, doi:10.1029/2011GL049324

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

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

Jones P.D. and A. Moberg (2003), Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001, *Journal of Climate*, 16, 206-223

Jones, P. D., D. H. Lister, and Q. Li (2008), Urbanization effects in large-scale temperature records, with an emphasis on China. *Journal of Geophysical. Research*, 113, D16122, doi:10.1029/2008JD009916.

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

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne and D. Joseph (1996), The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 437-471

Kalnay, E., M. Cai, H. Li and J. Tobin (2006), Estimation of the impact of land-surface forcings on temperature trends in eastern United States, *Journal of Geophysical Research*, 111, D06106, doi:10.1029/2005JD006555

Kennedy J.J., P.W. Thorne, T.C. Peterson, R. Ruedy, P.A. Stott, D.E. Parker., S.A. Good, H. Titchner and K. Willett (2010), How do we know the world has warmed? *in State of the Climate 2009*, *Bull. Am. Meteorol. Soc.*, 91(7) S26-27.

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

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

Kennedy, J.J., R.O. Smith and N.A. Rayner (2011c), Using AATSR data to assess the quality of in situ sea surface temperature observations for climate studies. *Remote Sensing of Environment*, doi: 10.1016/j.rse.2010.11.021

Kent, E.C. and P.K. Taylor (2006), Toward estimating climatic trends in SST part 1: methods of measurement, *Journal of Atmospheric and Oceanic Technology*, 23, 464-475

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

Mears, C.A., F.J. Wentz, P. Thorne and D. Bernie (2011). Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique, Journal of Geophysical Research, 116, D08112, doi:10.1029/2010JD014954

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

Menne, M.J., and C.N. Williams, Jr., (2009) Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22, 1700-1717

Moberg, A., H. Alexandersson, H. Bergstrom and P.D. Jones (2003), Were southern Swedish summer temperatures before 1960 as warm as measured?, *International Journal of Climatology.* 23: 1495–1521. doi: 10.1002/joc.945

Parker, D.E. (1994), Effects of changing exposure of thermometers at land stations, *International Journal of Climatology*, 14, pg 1-31

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

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108, No. D14, 4407

Rayner, N.A., P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett (2006), Improved analyses of changes and uncertainties in sea-surface temperature measured *in situ* since the mid-nineteenth century: the HadSST2 data set, *Journal of Climate*, 19, 446-469.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *Journal of Climate*, 15(13), pg 1609-1625

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

Thompson, D.W.J., J.J. Kennedy, J.M. Wallace and P.D. Jones (2008), A large discontinuity in the mid-twentieth century in observed global mean surface temperature, *Nature*, 453, 646-649, doi:10.1038/nature06982

Thorne, P.W., D.E. Parker, J.R. Christy and C.A. Mears (2005), Uncertainties in climate trends: Lessons from upper-air temperature records. *Bulletin of the American Meteorological Society*, 86, 1437-1442, doi: 10.1175/BAMS-86-10-1437

Thorne, P.W., J.R. Lanzante, T.C. Peterson, D.J. Seidel and K.P. Shine (2011), Tropospheric temperature trends: history of an ongoing controversy, *Wiley Interdisciplinary Reviews: Climate Change*, 2(1), pg 66-88

Vincent, L.A. and D.W. Gullet (1999), Canadian historical and homogeneous temperature data sets for climate change analyses. *International Journal of Climatology*, 19, 1375-1388.

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

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

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

Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker and N. Lott (2005), ICOADS release 2.1 data and products, *International Journal of Climatology*, 25, 823-842

Woodruff, S.D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith and C. Wilkinson (2011), ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive, *International Journal of Climatology*, 31: 951–967. doi: 10.1002/joc.2103

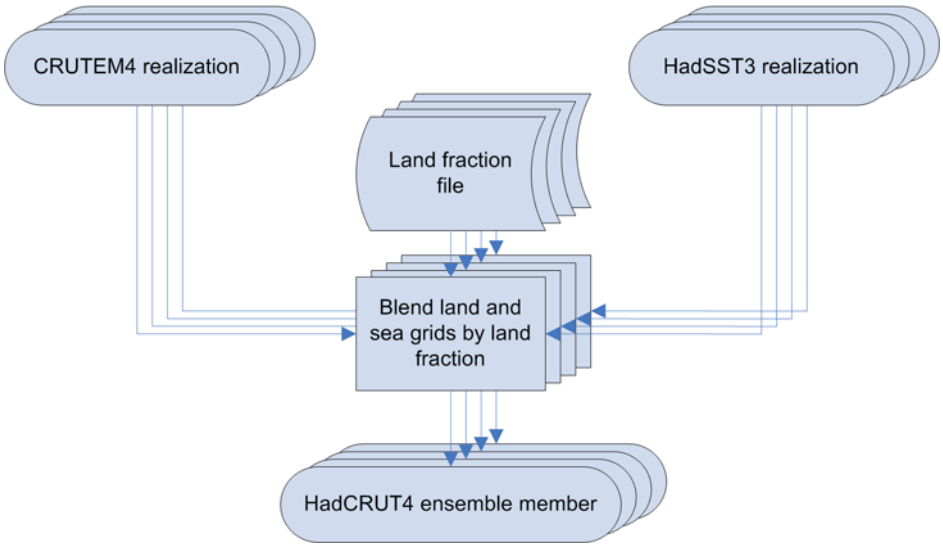

Figure 1 – The generation of the HadCRUT4 ensemble by land-fraction-weighted one-to-one blends of the 100 HadSST3 ensemble members with 100 realizations of the CRUTEM4 data set.

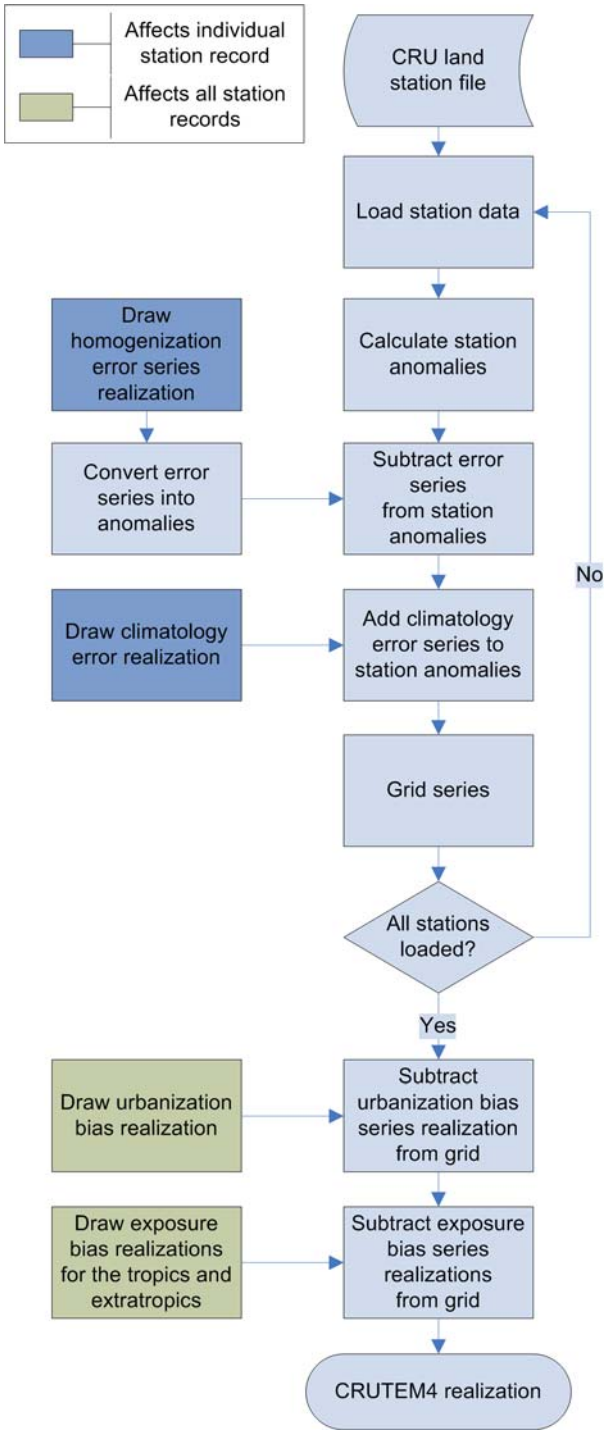

Figure 2 – Flow chart of the ensemble CRUTEM4 data set generation process, with processes that are allowed to vary in each ensemble member indicated.

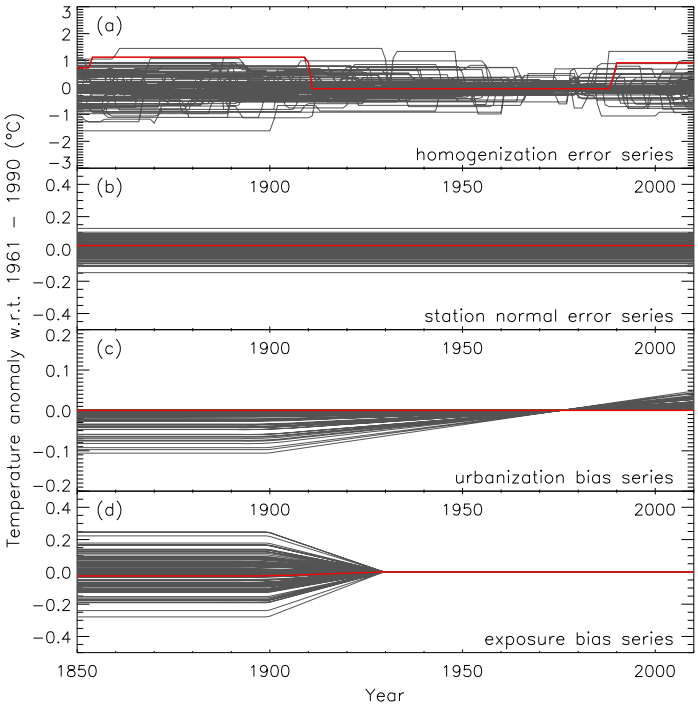

Figure 3 – 100 realizations of each uncertainty component contributing to the CRUTEM4 ensemble realizations, with an example realization of each component highlighted in red. Homogenization and station normal error realizations are drawn for each individual station record. Urbanization bias instances apply globally. Different values of exposure bias are applied uniformly across the extratropics (shown) and the tropics.

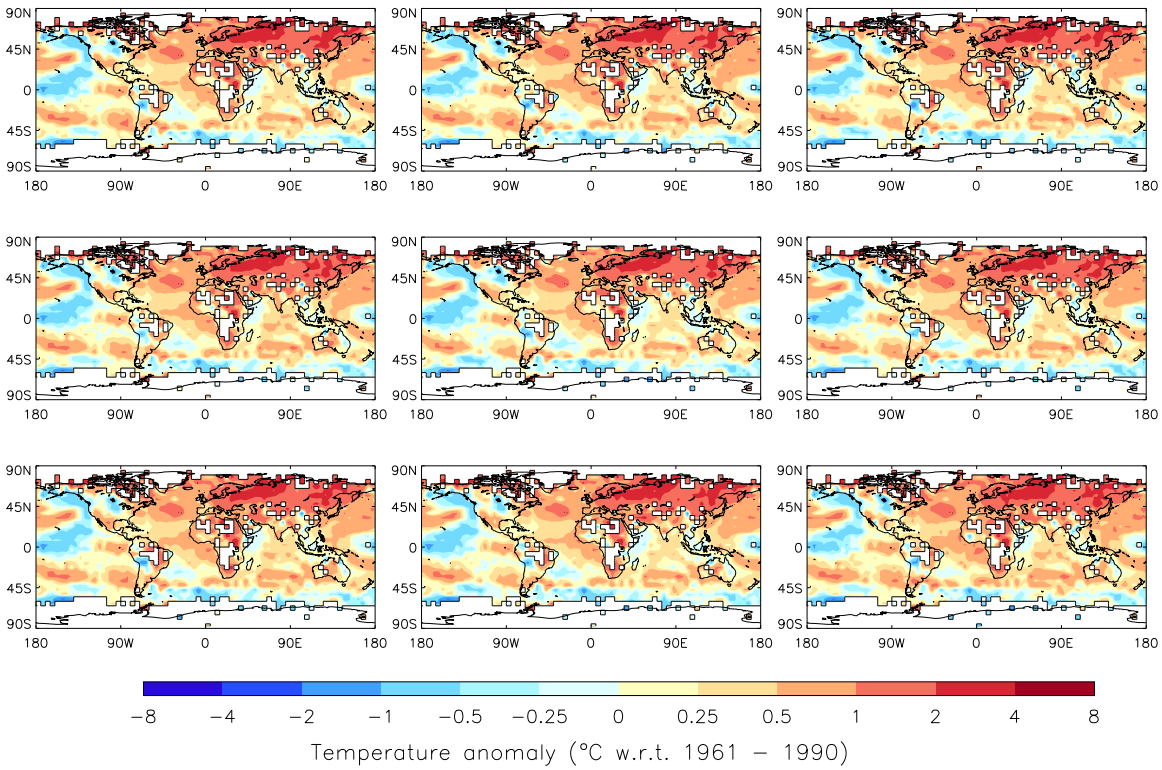

Figure 4 – Annual average surface temperature anomalies for 2008 (˚C with respect to 1961-1990) for 9 ensemble members of HadCRUT4. Anomalies are shown only for grid-boxes in which at least 6 months of data are available.

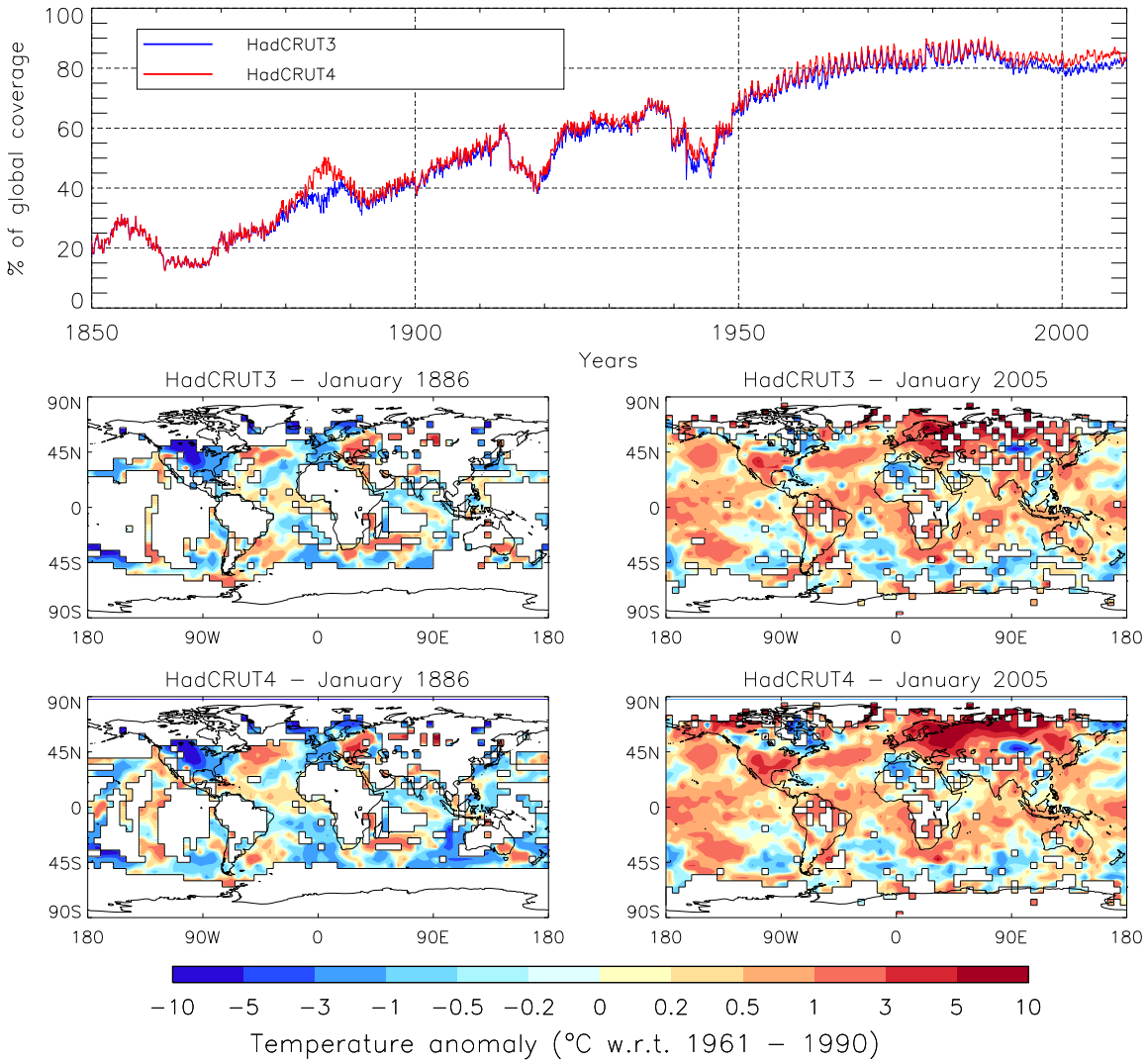

Figure 5 – Improvements in global coverage in HadCRUT4. Top: the percentage of global area observed. Bottom: anomaly maps for HadCRUT3 and HadCRUT4 for months of notable improvement in observational coverage. Maps show gridded temperature anomalies (˚C) with respect to grid-box average temperatures in the period of 1961-1990.

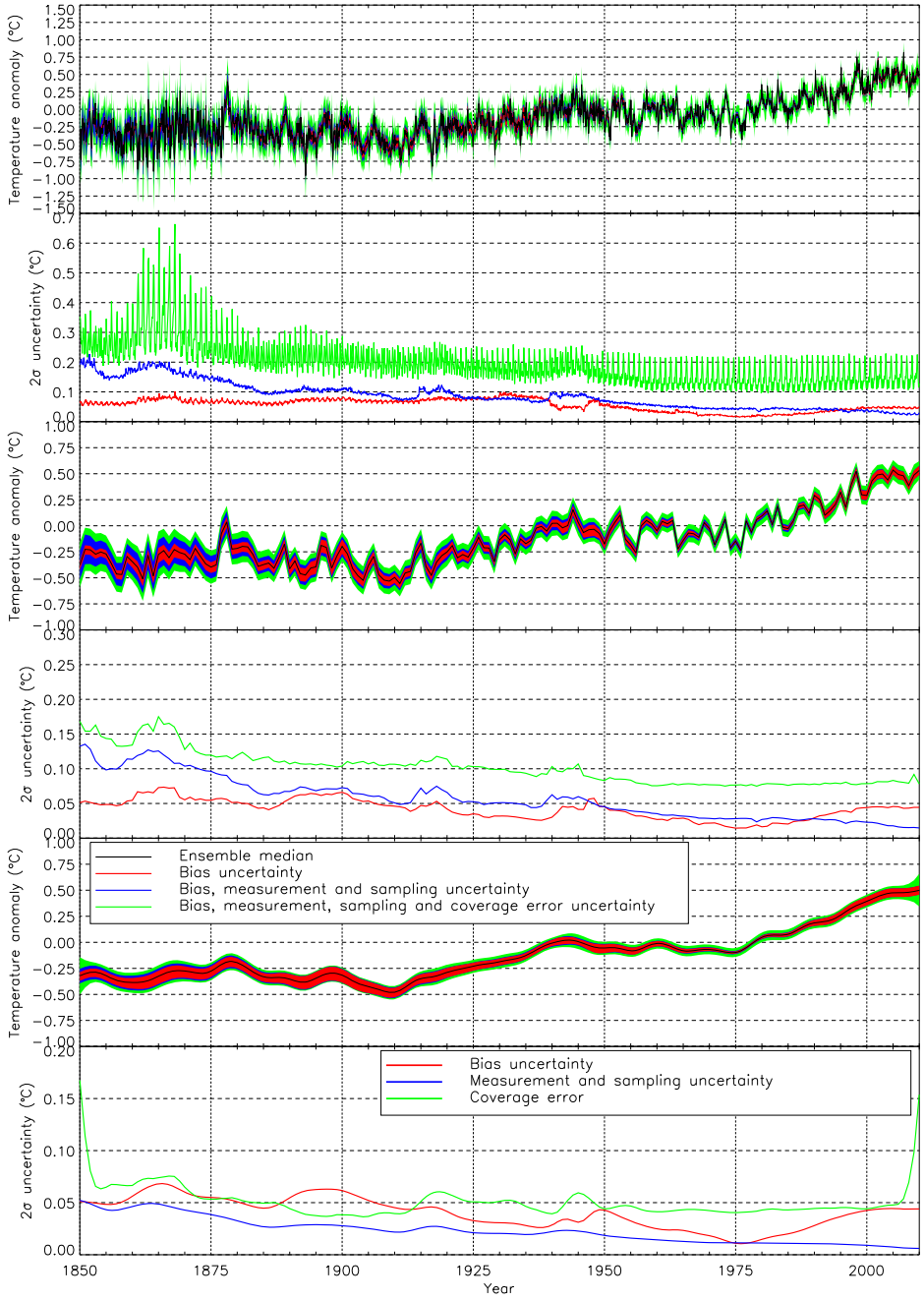

Figure 6 – Global average HadCRUT4 temperature anomaly time series 1850-2010 (°C, relative to the long-term average for 1961-90). (top) monthly time series and components of uncertainty in monthly averages. (middle) annual time series and components of uncertainty in annual series. (bottom) decadally-smoothed series and components of uncertainty in the decadally smoothed series

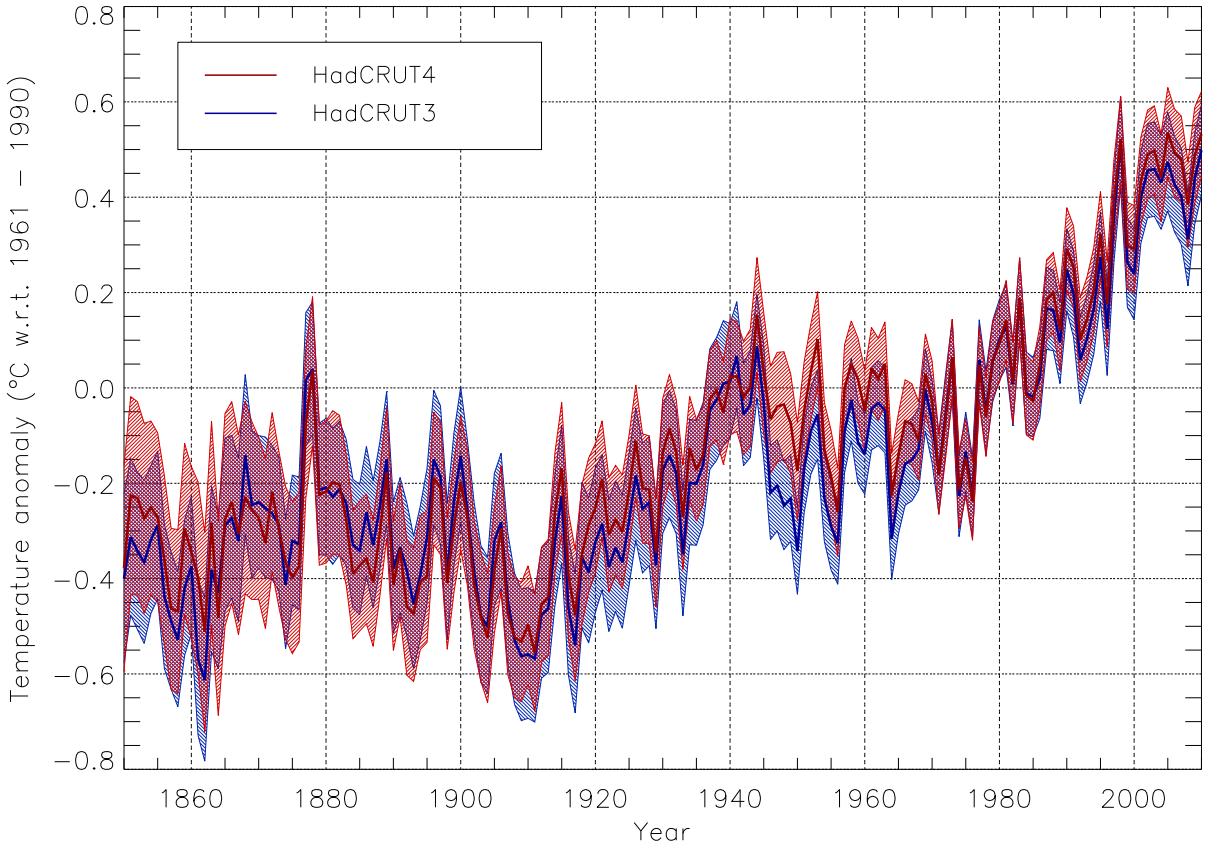

Figure 7 - Comparison of annual, global average temperature anomalies 1850-2010 (˚C, relative to the long-term average for 1961-90) for the HadCRUT4 median (red) and HadCRUT3 (blue). 95% confidence intervals are shown by the shaded areas.

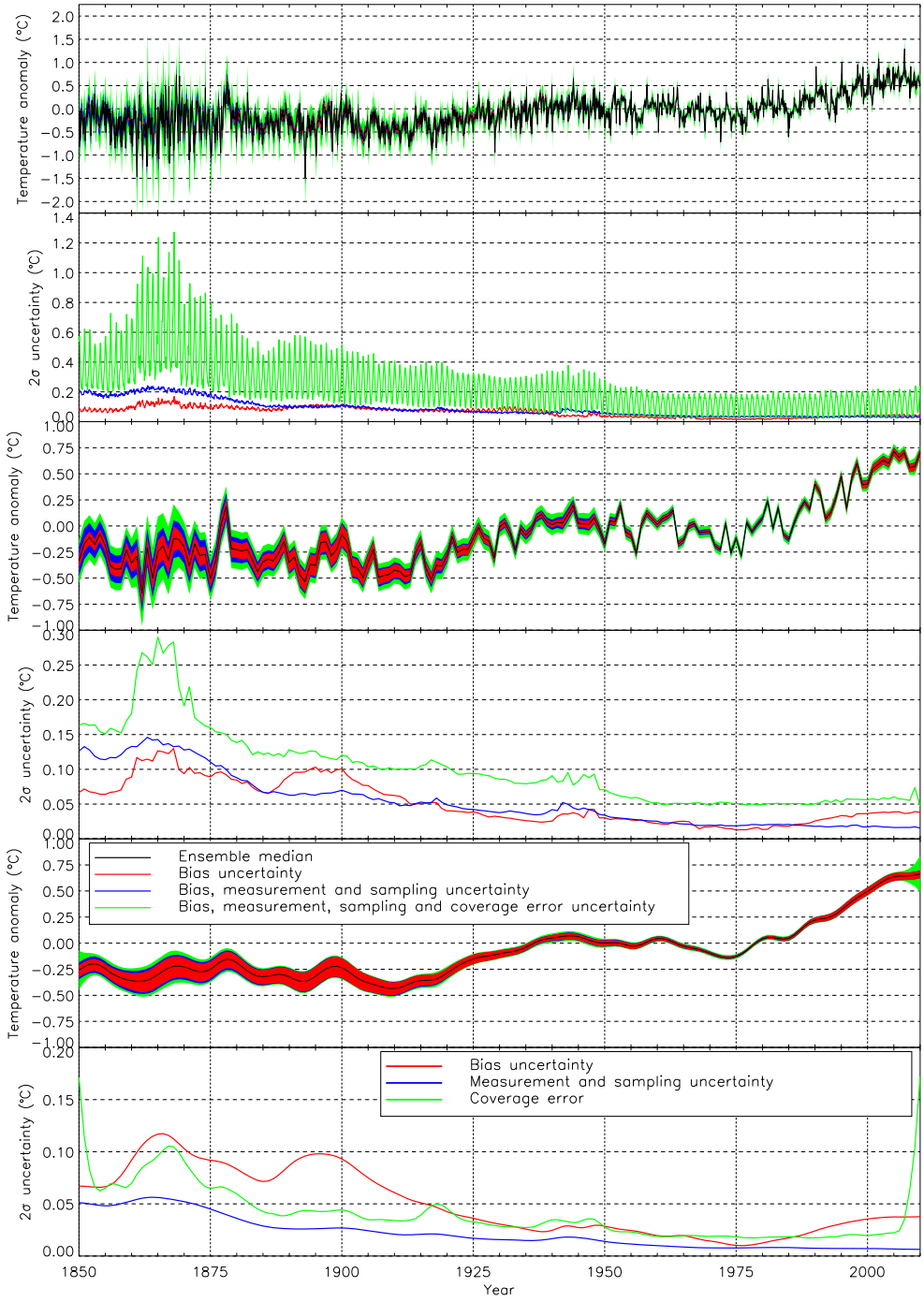

Figure 8 – Average HadCRUT4 temperature anomaly time series 1850-2010 (°C, relative to the long-term average for 1961-90) for the northern hemisphere. (top) monthly time series and components of uncertainty in monthly averages. (middle) annual time series and components of uncertainty in annual series. (bottom) decadally-smoothed series and components of uncertainty in the decadally smoothed series.

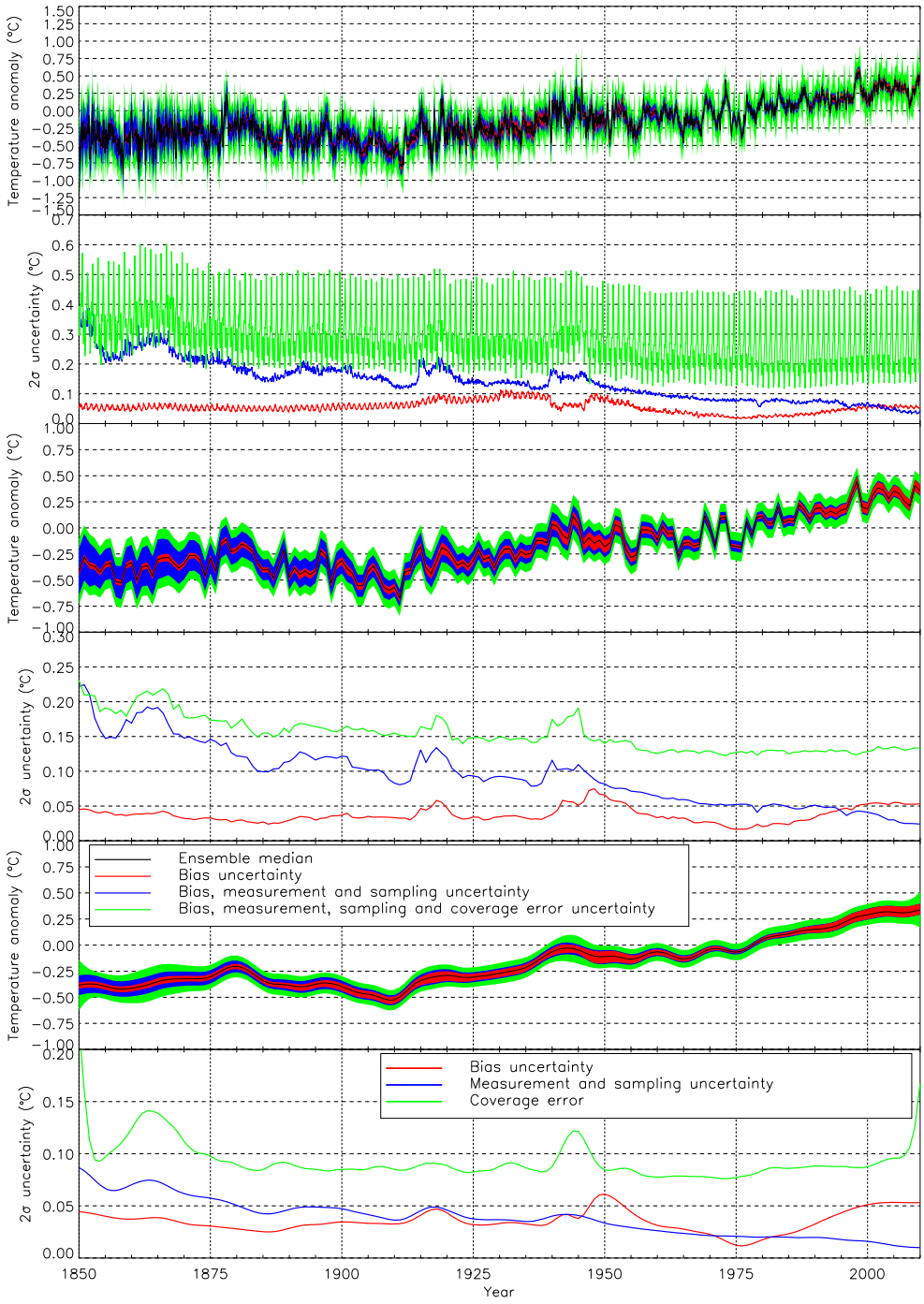

Figure 9 – Average HadCRUT4 temperature anomaly time series 1850-2010 (°C, relative to the long-term average for 1961-90) for the southern hemisphere. (top) monthly time series and components of uncertainty in monthly averages. (middle) annual time series and components of uncertainty in annual series. (bottom) decadally-smoothed series and components of uncertainty in the decadally smoothed series.

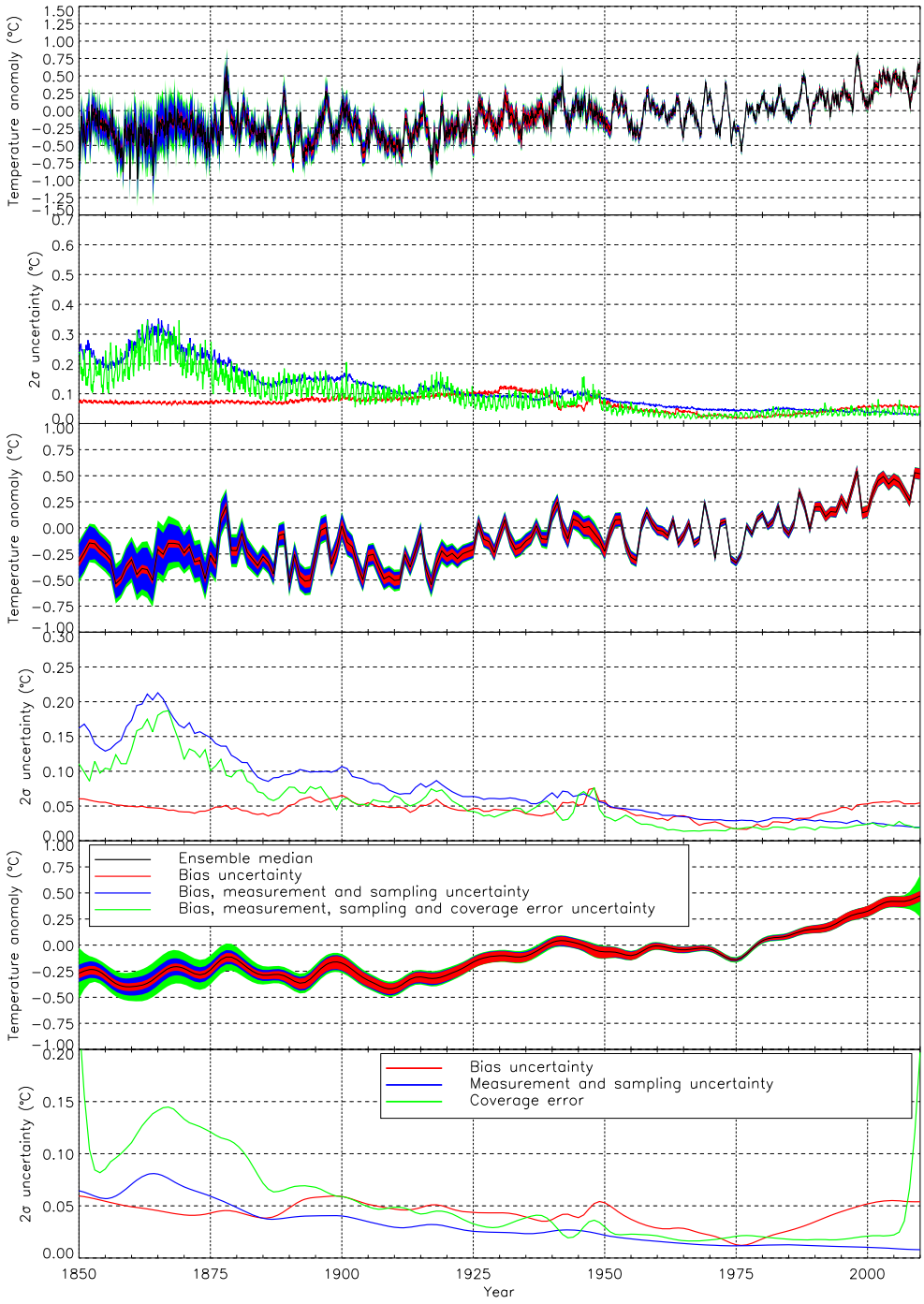

Figure 10 – Average HadCRUT4 temperature anomaly time series 1850-2010 (°C, relative to the long-term average for 1961-90) for the tropics (30˚S to 30˚N). (top) monthly time series and components of uncertainty in monthly averages. (middle) annual time series and components of uncertainty in annual series. (bottom) decadally-smoothed series and components of uncertainty in the decadally smoothed series.

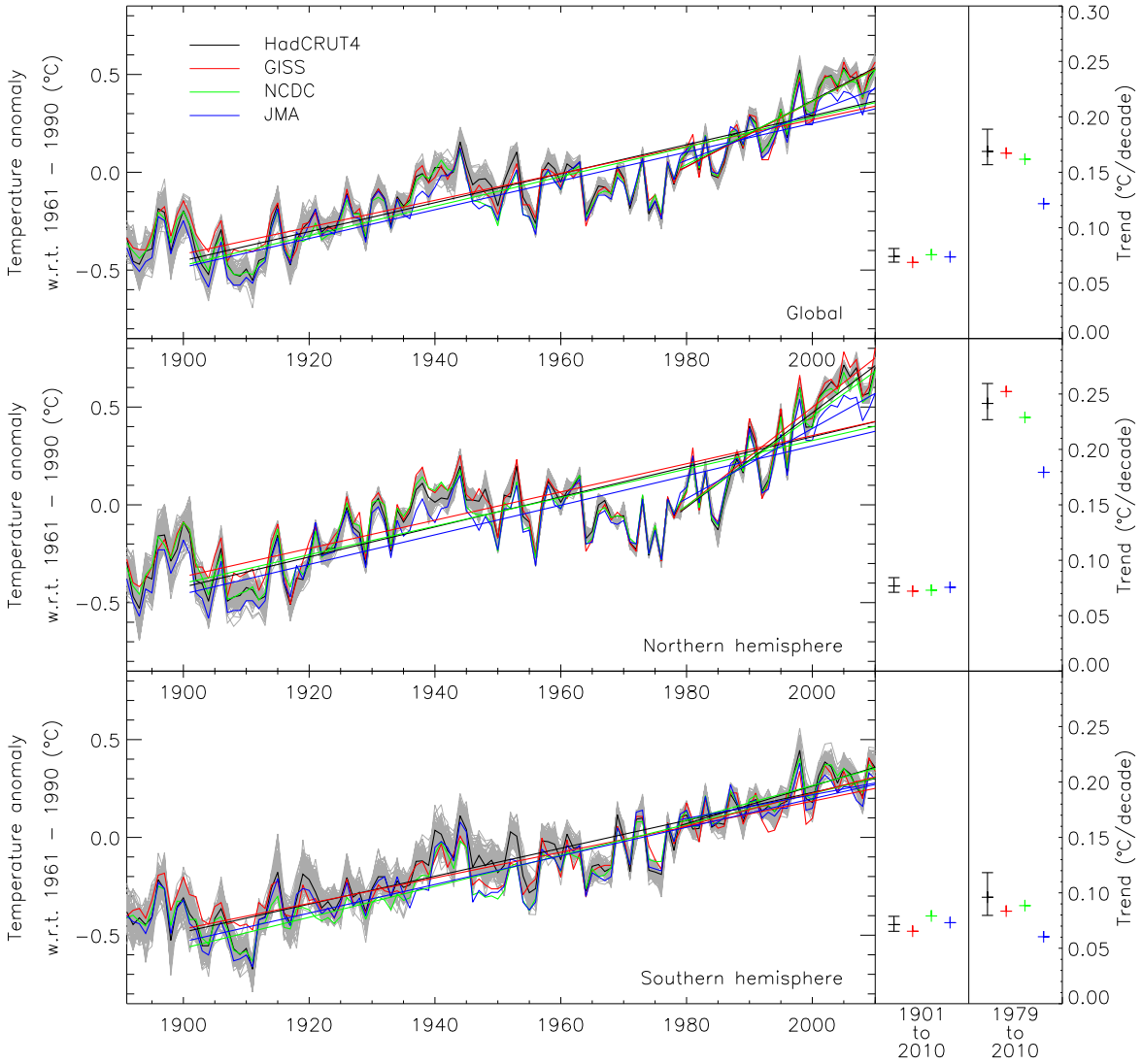

Figure 11 – Annual temperature anomaly development in the HadCRUT4, GISS, NCDC and JMA surface temperature analyses. Least squares linear trends are shown on the right for the periods of 1901to 2010 and of 1979 to 2010. Individual ensemble member realizations of HadCRUT4 are shown in grey. Uncertainty ranges in linear trends for HadCRUT4 data are computed as the 2.5% and 97.5% ranges in linear trends observed in the HadCRUT4 ensemble.

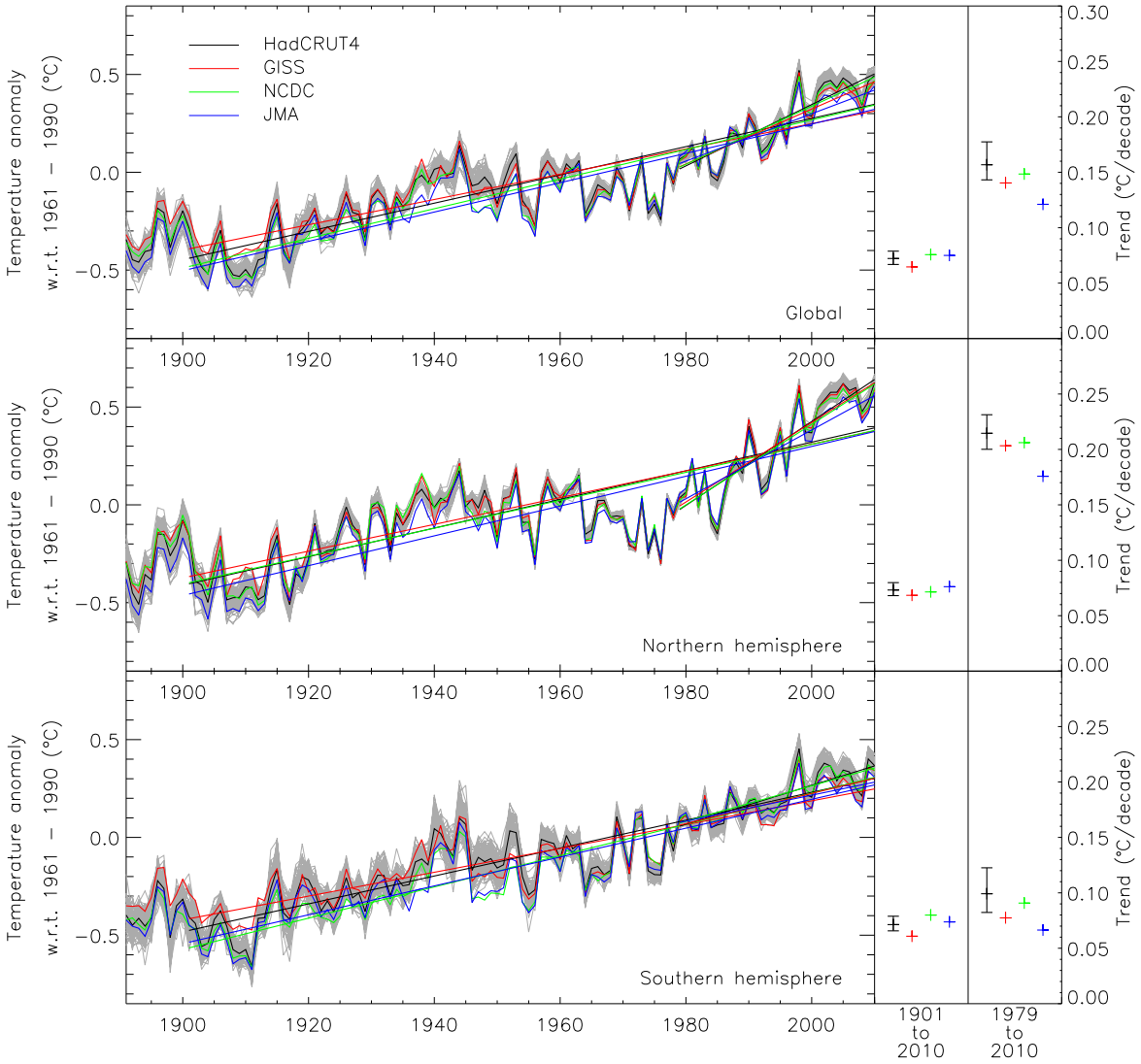

Figure 12 - Annual temperature anomaly development in the HadCRUT4, GISS, NCDC and JMA surface temperature analyses, with data set coverage reduced to the minimum coverage existing in all four data sets. Least squares linear trends are shown on the right for the periods of 1901 to 2010 and of 1979 to 2010. Individual ensemble member realizations of HadCRUT4 are shown in grey. Uncertainty ranges in HadCRUT4 data are computed as the 2.5% and 97.5% ranges in linear trends observed the HadCRUT4 ensemble.