# Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century: the HadSST2 data set.

N.A. Rayner, P. Brohan, D.E. Parker, C.K. Folland, J.J. Kennedy, M. Vanicek, T. Ansell and S.F.B. Tett

Hadley Centre for Climate Prediction and Research, Met Office, FitzRoy Road, Exeter, EX1 3PB, U.K.

Corresponding author email address: **nick.rayner@metoffice.gov.uk**

*16th September 2005. Resubmitted to Journal of Climate.*

**Abstract**

A new flexible gridded dataset of sea surface temperature (SST) since 1850 is presented and its uncertainties quantified. This analysis (HadSST2) is based on data contained within the recently created ICOADS data base and so is superior in geographical coverage to previous datasets and has smaller uncertainties. Issues arising when analysing a data base of observations measured from very different platforms and drawn from many different countries with different measurement practices are introduced. Improved bias corrections are applied to the data to account for changes in measurement conditions through time. A detailed analysis of uncertainties in these corrections is included by exploring assumptions made in their construction and producing multiple versions using a Monte Carlo method. An assessment of total uncertainty in each gridded average is obtained by combining these bias correction related uncertainties with those arising from measurement errors and under-sampling of intra-grid box variability. These are calculated by partitioning the variance in grid box averages between real and spurious variability. From month to month in individual grid boxes, sampling uncertainties tend to be most important (except in certain regions), but on large scale averages bias correction uncertainties are more dominant owing to their correlation between grid boxes. Changes in large-scale SST through time are assessed by two methods. The linear warming between 1850 and 2004 was 0.52±0.19ºC (95% confidence interval) for the Globe, 0.59±0.20ºC for the Northern Hemisphere and 0.46±0.29ºC for the Southern Hemisphere. Decadally filtered differences for these regions over this period were 0.67±0.04°C, 0.71±0.06°C and 0.64±0.07°C.

## 1. Introduction

The now greater than 150-year record of global sea surface temperature (SST) provides (in combination with air temperature measured over the land) a well-established means to quantify and monitor changes in the surface temperature of the globe (e.g. Nicholls et al. 1996; Folland et al. 2001b). In addition, the long record of SST data is used routinely to verify coupled GCMs (e.g. McAveney et al. 2001) and to force atmospheric and oceanic GCMs (e.g. Sexton et al. 2003 and Smith et al. 2005, manuscript submitted to *J. Geophys. Res.*) and Reanalyses (e.g. Fiorino, 2004).

The opportunity to draw on these data back to the 1850s owes much to the Brussels Maritime Conference of 1853 when representatives from several sea-faring nations agreed the standardisation of meteorological and oceanographic observations from ships at sea (Maury, 1858; 1859). The useable data from before this time are few for SST and are generally less coherent. However, not every detail of the method of taking measurements was standardised in 1853, which led to different countries using, for example, different types of buckets to collect sea water samples. In time, new standards were adopted by individual countries and this led to a changing mixture of water collection methods. The types of ships providing measurements and hence their speeds have also diversified in time. Both sets of changes have affected the measurements, introducing temporally and geographically varying relative biases into the data.

The way that the data base itself has been constructed introduces similar changes into the historical record. The International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Worley et al. 2005), upon which the analyses presented here are based, is assembled from "decks" of observations. These decks

were originally decks of punched cards on which the digitised ships' records were exchanged and stored. The introduction of a new deck into the data base can cause sudden changes from one data source, with a certain observational practice, to another, with some slightly or significantly different practice. If data from different sources are mixed together, then these relative biases may partly cancel out. However, if one data source dominates, perhaps in a particularly data sparse time, or a new source floods into the record, the change in relative bias can be systematic.

From the 1950s onwards, ICOADS contains data from the World Ocean Data Base (Levitus et al. 1994; 2000); specifically from sub-surface ocean profilers and ocean stations. From the late 1970s onwards moored and drifting buoys are also included. Latterly these have made up a very large proportion of the total number of observations in the data base due, in part, to their much greater frequency of reporting relative to ships and also to the often delayed reporting of ships' data and the general decline in the numbers of reporting ships.

In addition, ships' routes have changed over time for socio-economic reasons, e.g. the opening of the Suez (1869) and Panama (1914) Canals. Also, despite recent efforts at digitisation of previously unavailable historical data, there remain large data gaps at times of large-scale conflict, e.g. 1914-18 and 1939-45.

This analysis attempts to understand and, where possible, correct for these factors in order to provide historical marine time series which are as homogeneous as possible and to which meaningful estimates of uncertainty can be attached. Our new dataset, HadSST2, uses similar analysis techniques to a previous analysis (MOHSST, Parker et al. 1995a), but it includes all of the extra data available in the new ICOADS data base. In addition, this analysis is available on a grid of variable spatial resolution

and comes with gridded fields of numbers of observations, standard deviation and uncertainty estimates.

Having homogeneous time series is essential in order to be able to quantify changes in climate on large and small spatial scales over several decades (see, for example, Hurrell et al. 2000; Stendel et al. 2000; Hurrell and Trenberth, 1998). Here, biases have been quantified and removed from the data, and uncertainties associated with those bias corrections have been calculated. These uncertainties, because they tend to be more spatially coherent than uncertainties due to random measurement errors and inadequate sampling within each grid box, are particularly important to applications such as climate change detection and attribution (Thorne et al. 2003; Allen and Stott, 2003; Hegerl et al. 2001). Other applications which use the data either on very small (e.g. Barton and Casey, 2005) or very large (e.g. Gregory et al. 2002) spatial scales, need to know how far to trust the mean temperature estimates at these scales.

Previous work has attempted to quantify all of these uncertainties in global or hemispheric averages (Folland et al. 2001a), on the regional scale (Folland et al. 2003) or in modern data only (Trenberth et al. 1992; Kent and Taylor, 2005; Kent and Challenor, 2005; Kent and Kaplan, 2005; Kent and Berry, 2005). Others have estimated partial (Jones et al. 1997; Ishii et al. 2003; 2005) or total (Smith and Reynolds, 2003; 2004) uncertainties for each location in their analysis, sometimes by comparison with similar analyses. Here a fresh attempt is made to address the total uncertainty by calculating sampling and measurement errors from the gridded data and by deconstructing the methods and testing the assumptions used to devise bias corrections.

There are five further sections to this paper: Section 2 describes how the gridded fields of SST were constructed; Section 3 corrects biases and assesses the uncertainties associated with the gridded fields; Section 4 presents the major results obtained from the new data and compares our analyses with those on the global and hemispheric scales in Folland et al. (2001b); Section 5 draws together the main lessons learnt and Section 6 gives a forward look and introduces other aspects of the data not addressed here.

## 2. Producing gridded fields

This section describes the data used and the methods employed to create quality-controlled, gridded fields of SST on various spatial resolutions. Bias corrections are discussed in Section 3.

### *a. Source of observations*

The analyses presented here are based upon the collection of *in situ* measured SST contained within the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Worley et al. 2005). This data base is the most comprehensive collection yet because it has blended archives held in many countries observation by observation. Consequently, greatly improved data coverage is seen in many periods relative to that available in MOHSST (Parker et al. 1995a; see Figure 1 for illustrative examples of periods with the largest improvement in SST data coverage).

The dataset has been extended beyond 1997 (the end of the period of ICOADS data available at the time of analysis) through to the most recent month using data gathered from the Global Telecommunication System (GTS) by the National Centers for Environmental Prediction (NCEP). Examination of fields for 1997, which were

generated from both data sources, has demonstrated no heterogeneity between the two.

All SST observations which passed the basic ICOADS quality check (i.e. checking for erroneous land locations, duplicates, etc) are included in our analysis, subject to further quality control (for a brief description see Section 2c).

SST data in ICOADS and the NCEP GTS collection were collected from various different platforms: Voluntary Observing Ships; naval vessels; near-surface observations from oceanographic profiles and moored and drifting buoys, along with other less extensive collections of data. ICOADS comprises data provided by various different nations: some of the data can be identified as having been provided by a particular nation but some are part of large international collections. The data base is dominated by different data sources at different times, often with abrupt changes between them (see Figure 2, which uses the "deck id" metadata to split the data into sources). As is shown later, this can have consequences for the changing biases in the analysis. Full information regarding the makeup of the ICOADS data base can be found in the online documentation at http://www.cdc.noaa.gov/coads/.

*b. Background climatology*

New background climatological fields, used as a reference in the quality control of the data (see Section 2c), have been created from the new data base. The new climatology was initially constructed as 1° latitude by 1° longitude (hereafter 1° area) pseudo-monthly average fields for the period 1961-90; pseudo-months are multiples of pentads (5-day periods), which closely approximate to calendar months (Bottomley et al. 1990). The monthly climatology was then interpolated to pentad resolution using a cubic spline fit to adjusted mid-month values. This produced pentad values which averaged to the monthly means (see below for a discussion and Taylor

et al. 2000). Linear interpolation to daily resolution provided background fields to which the individual observations were referenced during quality control (Section 2c).

The new climatology was created by iterative refinement of the GISST2.0 1961-90 climatology (Parker et al. 1995b), used for MOHSST (Parker et al. 1995a) and HadSST1 (Jones et al. 2001):

1. 1° area average quality-controlled SST anomalies for each pseudo-month between 1961 and 1990 were added to the initial calendar monthly background field (from GISST2.0). These were then augmented at polar latitudes with monthly-varying SST from grid boxes partially covered by sea ice in HadISST1 (Rayner et al. 2003)
2. These fields were then interpolated using the Laplacian (Reynolds 1988) of the background field to create globally complete fields.
3. The resultant fields were averaged separately for each calendar month

The new 1961-90 averages were then used as background fields in an iteration of steps 2 and 3. In all, six iterations were needed to ensure convergence to a stable result.

Figure 3 illustrates the improvements made to the monthly climatology in some regions. The sea ice fields used (taken from HadISST1) are more homogeneous in time than those used in the analysis of the GISST2.0 climatology and the large differences seen in the Southern Ocean in Figure 3(b) are a consequence of this. In the open oceans, the differences between the new and old climatologies are small but, on the grid box scale, Figure 3(e) and (f) demonstrate that the new climatology fits the new mix of observations better.

Adjusting the monthly climatology (see Taylor et al. 2000) prior to interpolation to pentad resolution has removed a spurious annual cycle arising in

previous anomaly datasets (e.g. MOHSST, HadSST1). Previously, pentad (and hence daily) climatology fields did not exactly average to the monthly fields and the interpolated annual cycle was too small. As observations were expressed as anomalies from the daily climatology prior to quality control, then aggregated to pentad and monthly averages, this led to larger monthly anomalies compared to those which would have been calculated from the true monthly climatology.

*c. Quality control*

Each individual observation is verified by a series of steps and quality flags assigned. Only observations passing all tests are used in the gridded dataset. These tests are those used successfully in previous versions of our datasets (e.g. Parker et al. 1995a) with one exception (see below).

Initially, the observation is checked for a meaningful location, date and time and that it is not surrounded on all sides by land.

Each ship or drifting buoy with an individual call-sign is tracked to verify its reported position, speed and direction; those observations for which these are suspicious are flagged. Reported positions, speeds and directions are checked for consistency along the voyage. Observations can only be checked in this way if they have a valid call-sign that is unique to one ship or buoy, so those without a call-sign or with the generic call-signs 'SHIP' or 'PLAT' are passed unchecked.

Each SST observation is checked that it is above the freezing point of sea-water (here taken to be -1.8°C), and that the observation is within ±8°C of the 1961-90 background climatology (Section 2b) interpolated to that day; this more than allows for extreme values associated with, for example, large ENSO events.

The final quality control (QC) is a “buddy check”. This compares the value of each individual anomaly, formed by subtracting the climatology for the relevant day

and 1° area grid box from the observation, to the mean anomaly from neighbouring observations. All neighbouring observations passed by the previous QC tests are gridded on to a 1° area pentad resolution by taking winsorised (see below) means of anomalies of all the observations in each grid-box. The search radius for neighbours starts at ±2 pentads and ±110km then opens out, if none are found, first to ±2 pentads and ±220km, then to ±4 pentads and ±110km and finally ±4 pentads and ±220km, as necessary. Individual observations differing too much from their gridded neighbours are flagged as bad. For each grid-box an acceptable range is calculated using a reference field of standard deviation for that box. If no neighbours are found within ±4 pentads and ±220km, the observation passes the check. As detailed above, the buddy check now utilises neighbouring observations from both forwards and backwards in time, rather than simply backwards in time as documented in Parker et al. 1995a.

The influence of any outliers remaining unflagged by the QC is limited by the use of winsorisation (Afifi and Azen, 1979; Bottomley et al. 1990) to create the gridded mean fields. The distribution of all the observations passing all QC tests in each 1° area pentad (or other, see Section 2d) resolution grid box is divided into quartiles. Any observations falling within the first (fourth) quartile are set to the value of the boundary between the first and second (third and fourth) quartiles. The mean is then taken of these values together with those falling within the central two quartiles. If there are fewer than four observations in the grid box, a simple average is taken. The climatology value for the appropriate pentad and 1° area grid box is then removed to give an anomaly. The technique is the same as used previously to produce MOHSST (Parker et al. 1995a) and minimises the effect of outliers whilst retaining

real information during climatologically unusual periods, so it is particularly well-adapted to climate change.

Exploration of the data rejected by the QC process has verified that our QC procedures are not introducing biases or other errors into the analyses.

The QC process does not remove unflagged duplicates, of which there are many in certain periods of ICOADS (see Ansell et al. 2005, manuscript submitted to *J. Climate*), so these are explicitly removed prior to QC. There are also a few groups of obviously misplaced Russian MARMET observations (from deck 732, source ID 57, see Table 1 for details) which report systematically lower temperature than observations from other sources in the same regions. The QC tests did not flag these data, probably because of the number of similarly erroneous data, so they were explicitly excluded from the regions and periods indicated.

Data quality varies in time, as measurement and transmission methods change. Kent and Challenor (2005) report particular problems with the quality of some data in the early period of the GTS, when transmission and archival methods were still developing. The fraction of observations rejected by the QC varies through time with these changes in quality, e.g. 10-15% of SST observations were rejected in the 1970s, compared to 5-10% in the 1950s and 60s. The QC system has proved very effective at removing these erroneous data (see an examination of the case of April 1973 in Figure 4).

*d. Flexible gridding*

The gridding system follows the process used for MOHSST (Parker et al. 1995a); it grids observations into anomalies on a 1° area pentad resolution, and then combines these "super-obs" into a winsorised mean. However, it generalises this by allowing fractions of a 1° area pentad to be included in a grid box. In such cases, each

super-ob contributing to the grid box average comprises only those observations falling within the grid box; i.e. the super-ob can be a fraction of a 1° pentad. These super-obs are combined in a weighted winsorised mean. Each weight is the fractional contribution each super-ob makes to the new grid box. Gridded fields can be produced on any spatial and temporal resolution.

This added flexibility makes it much easier to produce datasets for specific user requirements (see Figure 5 for an example for SST in the North Atlantic), although the basic product remains a 5° area monthly dataset. Users of the dataset should be aware that increased resolution (in the absence of interpolation) often comes at the cost of sparser data coverage, as illustrated in Figure 5.

**3. Correcting biases and quantifying uncertainties**

Folland and Parker (1995) developed a set of corrections to be applied to SST to account for the effect on the recorded temperatures before 1942 of the use and changing construction of buckets used to collect sea water samples and the changes in ship speed through time. These corrections have been previously verified: by Smith and Reynolds (2002), who arrived at a similar set of corrections using largely independent methods; by Folland et al. (2001a) and Folland (2005) who forced AGCM simulations using corrected and uncorrected SST to illustrate that the use of corrected SST yielded simulated land surface air temperature anomalies consistent with the observed record; by Folland et al. (2003) who compared corrected SST to island air temperatures; and by Hanawa et al. (2000) who found corrected SST agreed with independent coastal station SST data around Japan. Here they have been adapted for the new data base.

As a finite number of observations contribute to each grid box mean temperature, and as estimated corrections have been applied to those means to remove

relative biases through time, grid-box or regional mean temperatures are not known exactly. Therefore, in order to provide the user with a range of possible values for each gridded average, the uncertainties in the average due to under-sampling and the applied bias corrections have been calculated. All sources of uncertainty arising from decisions made within the methods used for gridding and bias correction are included, but the higher level uncertainties arising from the decision to use one technique rather than another (see Thorne et al. 2005, for a discussion of these "structural uncertainties" in the context of upper air data) are neglected. Unless many different techniques are applied to the data, producing multiple datasets, the contribution of this type of uncertainty cannot be fully evaluated, so we have chosen to neglect it here. However, an indication of the sizes of these can be obtained by comparing our global and hemispheric averages to those of an equivalent data set, both presented in Section 4.

The decisions taken in the construction of the bias corrections are well documented (Folland and Parker, 1995). We repeat their method here, varying each input parameter within its own likely uncertainty to generate multiple possible realisations of the corrections, from which their possible spread can be calculated (see Section 3a). Thus we quantify their uncertainties and adapt them for HadSST2.

As the method used for calculating the sampling error uses the data already gridded (see Section 3b), there is bound to be some small additional contribution from random measurement errors, so sampling and measurement uncertainties are treated together.

Before fields and time series of combined uncertainties can be assembled, their relationship to one another must be determined, i.e. whether they are

independent or correlated, both within and between grid boxes. This is discussed in Section 3c.

*a. Bias correction and its uncertainties*

Sea water has been sampled for temperature measurement on board ship by various different means at different times. This change from using insulated (wooden) to uninsulated (canvas) to partly insulated (rubber) buckets, engine room intakes and hull sensors, along with changes in ships' speeds, has introduced changing relative biases into the data base. Folland and Parker (1995) developed corrections to be applied to SST data between 1856 and 1941 to ameliorate the effect of these changes and to bring the older data into line with data from the modern mix of measurement methods. For details of the development of these corrections, the reader is referred to Folland and Parker (1995). Here we include a summary of the method, describe how it has been adapted for HadSST2, assess uncertainties in the corrections and then compare to other work.

1) FOLLAND AND PARKER (1995) BIAS CORRECTION METHODOLOGY

The Folland and Parker (1995, hereafter FP95) bucket corrections are based on a combination of a number of models for bucket thermodynamics and estimates of changes in ship speed and bucket type with time. Specifically, the following information is used:

(i) the proportions of fast (7m $s^{-1}$) and slow (4m $s^{-1}$) ships, which affects the rate of heat loss from the buckets before the measurement is made

(ii) modelled quantities of heat lost from wooden and canvas buckets between collecting the water sample and taking the measurement on fast and slow ships in climatological ambient conditions and

(iii) the proportions of wooden and canvas buckets used.

The first two of these are independent pieces of information and the third is derived partly using the other two (see below).

The proportions of fast and slow ships was estimated in FP95 from the literature. The modelled heat losses were calculated using combinations of models of different sized buckets in different conditions. Specifically, correction fields for wooden buckets use an average of models for thick and thin buckets and those for canvas buckets combine models for large and small buckets, integrated over different times to mimic the effect of time delays between hauling the bucket and taking the sample, whilst sitting either fully or partially exposed to the sun on the deck.

Proportions of wooden and canvas buckets were obtained in FP95 by maximising the fit between nighttime marine air temperature (NMAT) and SST anomalies averaged over two regions of the tropics between 1856 and 1920. These regions were carefully chosen to avoid places where NMAT corrections (see Parker et al. 1995a for details) were dependent on SST. The method used by FP95 to fit the relationship was simply to vary the proportion of wooden buckets in 1856 from 0 to 100% in steps of 20%, assuming a linear increase from that value in 1856 to 100% in 1920. Out of these options, the linear trend that provided best agreement between corrected SST and NMAT over this period was chosen. FP95 cite evidence in the literature for fixing the proportion of canvas buckets at 100% in 1920 and thereafter.

Fields of corrections are obtained by combining the four basic sets of correction fields, i.e. those for: wooden buckets on slow ships; wooden buckets on fast ships; canvas buckets on slow ships and canvas buckets on fast ships, according to the changing fraction of fast and slow ships, wooden and canvas buckets for each year:

$$SST_{corr} = SST_{uncorr} + p_f p_c C_{fc} + p_f p_w C_{fw} + p_s p_c C_{sc} + p_s p_w C_{sw} \qquad (1)$$

where $p_c$ = proportion canvas buckets, $p_w$ = proportion wooden buckets ($=1-p_c$), $p_s$ = proportion slow ships, $p_f$ = proportion fast ships ($=1-p_s$), $C_{fc}$ = correction fields (a different one for each calendar month) for fast ships with canvas buckets, $C_{fw}$ = correction fields for fast ships with wooden buckets, $C_{sc}$ = correction fields for slow ships with canvas buckets and $C_{sw}$ = correction fields for slow ships with wooden buckets.

2) FITTING FOLLAND AND PARKER (1995) BIAS CORRECTIONS TO HADSST2

The same four basic correction fields and time series of proportions of fast and slow ships were used here as in FP95. However, the fitting of the proportions of wooden and canvas buckets was tailored to the new SST data base.

The NMAT dataset used as the reference in this fit, HadNAT2, was created from the ICOADS data base in a similar manner to HadSST2 (see Section 2). Corrections were applied to HadNAT2 to correct for the effect on the data of changing observation height as ships have generally become taller (see Rayner et al. 2003 for details of the corrections applied). Here, the same fitting technique has been used as in FP95 (i.e. still assuming a linear trend), but the proportions of wooden and canvas buckets in 1856 were not required to be a multiple of 20%.

In MOHSST, the SST anomaly (relative to 1961-90) before bias correction jumps suddenly from negative to positive values after December 1941 (see Figure 7(a)). This was previously attributed by FP95 to a sudden switch in the U.S. Navy from using buckets to sample SST to measuring the temperature of engine room intake (ERI) water during this part of World War II. In HadSST2, this change is more

of a gradual rise between the highly negative SST anomalies recorded in the late 1930s and values similar to those found in MOHSST in the early 1940s. This difference in behaviour was identified as being due to the substantially different mixture of multi-national data in ICOADS from that included in MOHSST at that time, most particularly in the addition of the newly digitised U.S. Merchant Marine data collection. The U.S. Merchant Marine SST now contributes between 60 and 80% of all SST data in ICOADS between 1938 and 1942, so is a significant alteration to the data base. Where measurement method is indicated in the metadata for this collection, it indicates that these data were largely measured from ERI water. The sharp discontinuity in MOHSST was therefore probably a result of the absence of U.S. ERI data in the data base before December 1941 and their presence thereafter, rather than a change in actual measurement practice.

It was clear that the FP95 corrections needed modification in this period for the new analysis, so they were altered to gradually decrease to zero between January 1939 and January 1942, rather than the previous continued increase to December 1941 and then sudden end at December 1941 (see Figure 7(b)). This was achieved by decreasing the annual proportion of canvas buckets from 100% in 1939 to 50% in 1940 to 25% in 1941. Smith and Reynolds (2004) linearly decreased their bucket corrections to zero over this period for the same reason. Plotting the difference in global mean SST in versions of HadSST2 which include and exclude these new U.S. Merchant Marine data and overlaying the difference in the new corrections relative to those of FP95 (see Figure 7(c)) confirms that this bias-correction change accounts for the differences caused by the introduction of the new U.S. Merchant Marine data.

When the new U.S. Merchant Marine data are further split into their constituent decks, it is found that, relative to corrected MOHSST data, one deck (705)

contains data which are generally unbiased (but tending to become warmer in the Atlantic), another (706) is completely unbiased in the Atlantic, but negatively biased in the Pacific and the third (707) displays a gradually reducing negative bias in both oceans (not shown). This indicates a different method of SST data collection in the different U.S. Merchant Marine decks and is an illustration of the complicated nature of the biases in historical SST data.

The newly bias-corrected HadSST2 time series can be seen for the globe in Figure 7(d) compared against the FP95-corrected MOHSST series; they are now very similar during 1938-41. Clearly, the FP95 correction procedure is robust and the corrections require only these small adaptations to be applicable to HadSST2.

Further work (see Section 6), has shown that this is not the only effect on overall biases of the inclusion of new sources of data. However, it is a major effect (albeit over a short period) and was relatively easy to correct for.

## 3) UNCERTAINTY IN BIAS CORRECTIONS

The evidence and assumptions used in the construction of the FP95 bias corrections were studied to assess likely uncertainties in each of the input parameters. Assuming that the input parameters can all be drawn from normal distributions (with variance equal to the square of one standard error in the input), the total uncertainty in the FP95 corrections was calculated by generating multiple realisations of the corrections by Monte Carlo simulation, drawing each input parameter randomly (with replacement) from its distribution.

The fit of appropriate proportions of wooden and canvas buckets depends on the uncertainties in: the ships' speeds; the four correction fields; and the NMAT data. The reader is referred to the appendix for the details of these uncertainties.

The fit was performed 1000 times, randomly sampling with replacement the canvas and wooden correction fields, the proportion of fast ships and the corrected tropical NMAT anomaly time series from their likely distributions. Sampling and measurement uncertainties in the SST were neglected here, as the calculation of those uncertainties involves using bias-corrected SST, which would lead to an inter-dependence of the two sets of uncertainties. This interdependence is difficult to assess and its contribution is, nevertheless, likely to be small relative to that of other components.

The result of this process was a distribution of possible canvas/wooden bucket proportions for each year which, along with distributions of the other input parameters, was used to create 1000 realisations of possible bucket corrections. Although the input parameters were drawn from normal distributions, their combination results in a distribution which is not always normal, owing for example to the multiplications in Eq. (1). So we use the median from our 1000 realisations at each spatial and temporal location to produce a "best" correction. The 95% confidence interval gives the uncertainty, shown in Figure 6 for selected months in 1938. It is clear that the uncertainty varies both geographically and seasonally, as do the corrections themselves.

## 4) COMPARISON TO EARLIER WORK

Figure 8 compares the global area-weighted mean bias corrections of FP95 and this study. The median HadSST2 bias correction is slightly greater than the FP95 correction in 1856 (by about 0.02ºC). This could be due to a combination of the effects of changes in the data base of SST and a change in the deck height corrections applied to NMAT, since FP95, resulting in an increase in the inferred proportion of canvas buckets in 1856 from 20% in FP95 to 30% in this study, or simply an artefact

of the fact that FP95 tested 20% increments in canvas bucket proportion. In any case, this difference in canvas bucket proportion is well within the uncertainty estimated by Folland et al. (2001a) of ±20%. The resultant difference in the two sets of bias corrections is within our 95% confidence interval, given by the 2.5$^{th}$ percentile and 97.5$^{th}$ percentile HadSST2 corrections (Figure 8). This difference is smaller than the difference between the corrections of FP95 and those of Smith and Reynolds (2002).

Folland (2005) drew tentative conclusions about the mix of canvas and wooden buckets in the late nineteenth century which would be required to improve agreement between atmospheric-model-simulated and observed land surface air temperature. His conclusion was that the fraction of canvas buckets should be less than in FP95, rather than more, as deduced here. To test this, the experiments of Folland (2005) might be repeated using a forcing dataset based on the analysis presented here, comparing to the improved land surface air temperature data now available (Jones and Moberg, 2003).

Smith and Reynolds (2002) discussed uncertainties in SST bias corrections in their comparison of the FP95 corrections to their own assessment. They found that the correction uncertainty was largest in the nineteenth century and in the 1940s. Their calculated 60°N-60°S average correction is about 70% larger than that of FP95 in 1856, whereas our new global average correction is about 30% larger then. However, the Smith and Reynolds (2002) correction relied heavily on NMAT data adjusted using an earlier set of deck height corrections (Bottomley et al. 1990). Rayner et al. (2003) demonstrates that the NMAT corrections used here result in NMAT data which are 0.05°C cooler on the global average in 1856 than those used by Smith and Reynolds (2002), assuming the same data composition. So, if they were to recalculate their SST bias corrections using NMAT data with Rayner et al. (2003) corrections,

their SST bias corrections would be about 0.05°C smaller, which would bring them into line with ours.

Folland et al. (2001a) also estimated the global mean uncertainty in SST anomaly due to the FP95 bias corrections. They included consideration of differences due to size and exposure between bucket models of the same type (i.e. canvas or wood), and of the uncertainty owing to the lack of specific knowledge of the relative proportions of canvas and wooden buckets. They neglected the uncertainty in the length of time chosen for integration of the canvas bucket model, which we include (see the appendix). This is the major part of our calculated uncertainty in the correction fields for canvas buckets and leads to an increasing total bias correction uncertainty through 1939 (c.f. our 95% confidence interval of ±0.09°C in 1939 with their ±0.06°C). We include the uncertainty in the proportion of canvas or wooden buckets by explicitly calculating it from the fit to NMAT (see above). This results in a smaller uncertainty than that obtained from the Folland et al. (2001a) conservative standard error of ±20%, which was based on a conclusion from the literature that buckets were "in fairly general use" from 1870 onwards (Folland and Parker, 1995). This contribution to the Folland et al. (2001a) bias correction uncertainty decreased from 0.06°C to zero between 1890 and 1920, as the deviation from the "mostly canvas" assumption decreases. So, our confidence interval increases through time from 0.03°C in 1856 to 0.09°C in 1939, as the contribution of canvas buckets (which involve the largest uncertainties in our calculation) increases. The Folland et al. (2001a) uncertainties decrease from 0.13°C in 1856 to 0.06°C in 1939, as their conservative uncertainty in the proportion of canvas and wooden buckets (their largest uncertainty) decreases.

As we have improved the SST bias corrections, tailoring them to the new data base and employing NMAT data with improved bias corrections in our wooden/canvas bucket proportion fit, our uncertainty estimates are smaller than those of Smith and Reynolds (2002) and Folland et al. (2001a), as discussed above.

***b. Sampling and measurement error***

When looking at multi-decadal time series of *in situ* measured temperature, it is important to recognise that the data base of observations is inconsistent in number from month to month and from place to place. Figure 9 illustrates the temporal variability of data numbers contributing to our SST analysis through time. It is clear that the confidence which can be placed in (or, conversely, the numerical uncertainty which can be assigned to) the gridded temperature averages also varies in space and time. The figure also shows why 1850 is the starting point of our SST analysis, as there is a steep decline in the geographical coverage of available observations prior to this time.

The uncertainty in a grid box average temperature anomaly value due to measurement error and under-sampling will depend on the number of observations which contribute to that average, with more observations giving a lower uncertainty. The standard error of the average of $n$ independent well-spread observations is given by the standard deviation of the observations divided by root $n$. However, our gridded fields are not simple, equally-weighted averages of all available observations because of the quality control procedures (Section 2c). Therefore, to assess uncertainties, it is necessary to use an indirect procedure that is based on the properties of the gridded averages themselves.

A time series of SST anomaly from a given grid box (see an example in Figure 10(a)) will generally show several sorts of variability: a long-term trend, interdecadal

variability, and high frequency variability which is due in part to measurement and sampling uncertainty. To isolate the variability due to measurement and sampling uncertainty, it is necessary to first remove the large low frequency variability components. The best way to do this will depend on the data being analysed. Here a moving six-year average is subtracted from each point (Figure 10(b)). It is clear from the example time series that such detrended SST anomalies appear more variable when data are sparse than when data are plentiful.

Assume that each grid box consists of $n$ randomly distributed point measurements. Let the true climatological variance of any point SST anomaly (from a fixed climatology) in the grid box be assumed constant at $c^2$. Here "true" implies no measurement errors. The constancy of $c^2$ within a grid box should be a very good assumption, except perhaps in some coastal grid boxes or where two very strongly differentiated water masses interact on the grid box scale. In general, the true anomalies within a grid box will be quite strongly spatially correlated on monthly and longer time scales. Let the observations be randomly distributed in space and let the average correlation of the time series of every true point anomaly with every other true point time series be $\overline{r}$. Let the variance of the independent random errors be similarly constant across the grid box at $m^2$. Then the total point variances will all be:

$$s^2 = m^2 + c^2 \tag{2}$$

Kagan (1966; see an English translation in Yevjevich 1972) show that the variance of the grid box average of $n$ spatially correlated values with variance $c^2$ is:

$$S_{nt}{}^2 = \frac{c^2(1+(n-1)\overline{r})}{n} \tag{3}$$

As expressed here, this is the variance of the grid box average of the true point values. Assuming the $n$ measurement errors are spatially uncorrelated, the variance of the grid box average measurement error is

$$S_{ne}^{\ 2} = \frac{m^2}{n} \tag{4}$$

The total variance of the grid box average is then

$$S_n^{\ 2} = \frac{c^2 + m^2 + c^2(n-1)\overline{r}}{n} \tag{5}$$

As $n \to \infty$ then

$$S_\infty^{\ 2} = c^2\overline{r} \tag{6}$$

Thus $c^2\overline{r}$ is the true variance of the grid box average. Thus, for *n* observations the additional error variance due to measurement and under-sampling is

$$S_E^{\ 2} = S_n^{\ 2} - S_\infty^{\ 2} = \frac{c^2(1-\overline{r}) + m^2}{n} \tag{7}$$

Eq. (7) shows that the error variance of a grid point average falls as $n^{-1}$. Accordingly, the slope of the relationship between $S_E^{\ 2}$ and *n*, $dS_E^{\ 2}/dn$, should be proportional to $n^{-2}$. The true variance of the grid box average is the right hand asymptote ($n \to \infty$) of a plot of $S_n^{\ 2}$ (y axis) versus *n* (x axis), and the additional variance in the grid box average for any lesser value of *n* is the difference between the right hand asymptote and the value of $S_n^{\ 2}$ (Eq. (7)). For a single observation, the combined sampling and measurement error variance in the grid box average is the difference between the variance of one observation and the true variance. This is the difference on the y axis between the left and right hand extremes of a plot of $S_n^{\ 2}$ versus *n*. In this paper, the sampling error standard deviation and true standard deviation are estimated by plotting $S_n$ versus $n$ and fitting a curve (e.g. Fig 10(c)).

The fit is performed separately for each grid box, using data for all months in the time series, yielding two fields: one of true standard deviation (see Figure 11(a)) and one of error standard deviation due to under-sampling and measurement

uncertainties for the special case when the grid box average is based on only one observation (Figure 11(b)). Combined sampling and measurement uncertainty standard deviation fields for each month are then calculated by dividing the values in the latter field by the square root of the number of observations in each grid box in that month. Figures 11(c) and (d) show examples of such fields for SST anomaly in September 1853 and September 2003.

This method does not address the issue of systematic changes in the spatial distribution of observations within each grid box which can affect $\overline{r}$ for finite $n$. However, over the small grid boxes considered here, these effects are likely to be small. Indeed, given that ships' routes are generally found in the same place year after year (excepting the large changes brought about by e.g. the opening of the Panama and Suez Canals), the error estimates will include the average effects of any non-random distribution.

Figure 3 of Kent and Berry (2005) compares combined sampling and measurement uncertainties calculated using our method with measurement errors calculated using the method of Kent and Challenor (2005). Their zonal average analysis shows that measurement error forms a large part of the combined uncertainty in tropical regions, but is relatively less important at higher latitudes.

*c. Combining uncertainties*

Examination of the correlation structure of individual observations has shown that sampling and measurement uncertainties for monthly 5° area grid boxes are independent between grid boxes. It is assumed that bias correction-related uncertainties are perfectly correlated within a month from grid box to grid box, as the assumptions made in the construction of the bias corrections (see Section 3a) apply to

all grid boxes equally, but that they are uncorrelated with the measurement and sampling uncertainties.

Therefore, measurement/sampling uncertainties are added to bias correction errors in quadrature when calculating the total uncertainty for an individual grid box (see examples in Figure 12). However, for uncertainties in larger regional averages the measurement/sampling uncertainties are input to an optimum averaging procedure and we use the many realisations of the bias corrections and hence of HadSST2 to calculate a distribution of possible regional averages from which we can explicitly calculate their uncertainty (see Section 4b).

Figure 12 also shows the relative contribution to the total uncertainty in gridded SST anomalies of sampling and measurement versus bias correction uncertainties. In many 5° areas, the bias correction uncertainty is relatively unimportant, as the sampling and measurement uncertainty is substantially larger. However, when the bias corrections and their uncertainties are largest and the quantity of SST data has increased, so decreasing the sampling and measurement uncertainty, bias correction uncertainties can be comparable to sampling and measurement error.

## 4. Key results

This Section brings together the main aspects of this study, assessing the effect of including new data sources on estimates of uncertainty and calculating overall uncertainties on global, hemispheric and regional averages.

### *a. Benefits of data digitisation*

Much time and effort went into creating the new ICOADS data base, digitising historical data sources and amalgamating these with existing collections, whilst ensuring that duplicate observations were removed, where possible. But, what effect

do all these new data have on our confidence in our understanding of marine climate variability and change?

A version of HadSST2 was created which excluded all data sources which were easily identified as being new to ICOADS. Observations with Source IDs 22 and 24 and above were excluded; the reader is referred to http://www.cdc.noaa.gov/coads/ for the details). Figure 13(a) illustrates the decrease in data numbers this entailed. This sub-sampled dataset was then used to verify that our model for sampling and measurement uncertainty (Section 3b) was robust, by re-running the model using the sub-sampled data base and finding that it made insignificant differences to the fields depicted in Figure 11 (a) and (b). However, as the numbers of observations contributing to each grid box average are different in the full and sub-sampled analysis, there were different actual sampling/measurement uncertainty estimates for each grid box in each month and on the global mean (see Figure 13(b)). It is clear that the largest uncertainties have been reduced significantly by the large increases in data numbers at some of the previously most poorly represented times and, therefore, that the newly incorporated data have made an important contribution to our knowledge of climate variability and change. Section 4b shows significant changes to large-scale averages in these periods. On the regional scale, decreases in uncertainty associated with these increases in data availability are much larger than in global averages (see the examples in Figure 13(c) and (d) for June 1940).

*b. Global and hemispheric averages*

Global and hemispheric SST averages are calculated here using both a simple area-weighting method and optimum averaging (OA, see Folland et al. (2001a) for the version used here). The OA method utilises the sampling and measurement uncertainties to weight the data according to their reliability, and the information

contained in EOFs of the SST anomalies to weight them according to their contribution to the mean, in order to produce the "optimum" average. It also produces a sampling error estimate.

The OA procedure was followed as in Folland et al. (2001a), except:

1. only marine temperature data were included and
2. EOFs calculated from data for 1870-2004 were used to define the covariance structure of the gridded SST data, rather than the shorter period of 1948-99 used by Folland et al. (2001a). Our EOF period encompasses the full range of secular variability over the last 150 years, leading to a better representation of this in our averages than in averages calculated by Folland et al. (2001a). Complete fields of SST anomaly needed for calculation of EOFs were created by filling data voids in HadSST2 using the Laplacian of HadISST1 (Rayner et al. 2003) SST anomalies. HadISST1 is globally complete from 1870 onwards, so this is achievable for SST, but no equivalent dataset currently exists for combined land and marine data.

Figure 14 illustrates the results and compares the HadSST2 time series with the simple, area-weighted SST anomaly averages used by Folland et al. (2001b, hereafter IPCC TAR) and those from the ICOADS gridded enhanced summaries (Worley et al., 2005), bias corrected using Smith and Reynolds (2002) corrections and interpolated to our 5° latitude by 5° longitude grid. The global averages (Figure 14(a)) are very similar in HadSST2 and IPCC TAR, except for the periods 1945-60 and 1865-80, when differences are as large as 0.1ºC (with HadSST2 cooler (warmer) than IPCC TAR in the 1945-60 (1865-80) period). In the Northern Hemisphere (Figure 14(b)), there are also differences between these two data sets of a similar magnitude

in the first two decades of the twentieth century (HadSST2 cooler), which is a period of particularly improved data coverage in the new dataset. In the Southern Hemisphere (Figure 14(c)), differences are about 0.2ºC (here with HadSST2 warmer) between 1870 and 1890. This is the main large-scale change relative to IPCC TAR. The OA gives slightly warmer conditions than a simple average at this time, but this may be due to the weighting of the data to take account of both the large data gaps and the uncertainty in the data. Gridded values with large uncertainties were common in this hemisphere at this time. Because the EOFs used in the OA extend back into the nineteenth century and not just back to 1948, as before (Folland et al. 2001a), we are less likely than Folland et al. (2001a) to be underestimating relatively cool temperatures at this time. The OA and simple averages of HadSST2 are almost always closer together than to the IPCC TAR, so we assess that it is very likely that the new data from ICOADS are correctly assessing warmer late nineteenth century SSTs than shown in Folland et al. (2001b).

Differences are as expected between the HadSST2 and ICOADS gridded summaries through 1941, because of the different bias corrections applied. However, there are also interesting differences emerging between these two data sets in the last few years of the record (with HadSST2 warmer). This may be due to subtle differences in data sources and should be explored further. Generally, the differences seen in these large-scale averages arising from the QC and gridding methods are almost always less than 0.1°C.

The tropical Pacific region shown here is [10°N-10°S, 180-120°W], because the EOFs used in our OA are on a 10° latitude by 20° longitude spatial resolution. The time series for this region is shown for 1875 onwards, because there are very few data in this region prior to this. The new data confirm the relatively large ENSO

fluctuations before 1920, followed by the reduction in variance between 1920 and about 1980 and subsequent increase during the last thirty years seen, for example, in Kestin et al. (1998). In this particular area of the central and east tropical Pacific, the pre-1920 fluctuations are as large as those of recent decades, allowing for the modest warming tendency, so do not provide evidence of a long-term increase in ENSO variance. Further analyses of the enhanced data base in other regions representative of ENSO are needed to confirm this, exploiting the new tropical Pacific data for this period.

The 95% confidence intervals shown in Figure 14 are a combination of uncertainty from sampling of the region (as calculated by the OA procedure) and from bias corrections. 1000 optimally averaged time series with uncertainties were produced for each region from 1000 realisations of HadSST2 by applying each of the 1000 possible sets of bias corrections (see Section 3a) in turn to the uncorrected SST. These 1000 averages and uncertainties were combined into a single distribution at each time point. The median is plotted as the solid black curve and the $2.5^{th}$ and $97.5^{th}$ percentile values give the confidence intervals. The resultant spread in the large-scale averages is smaller than seen in the bias corrections themselves (Figure 8) because the OA uses the same set of EOFs each time, based on the version using the best estimate corrections, which may tend to bias the result towards this. In any case, the OA always acts to minimise the error in the average, so will weight the data accordingly. The uncertainty estimate produced by the OA relates only to the accuracy of that average and its veracity is determined by the assumptions made, e.g. the appropriateness of the EOFs. So, these uncertainties are not expected to encompass the differences between all three time series because they do not take into account differences in data sources and in averaging method.

Linear trends in the OA time series were calculated by Cochrane-Orcutt estimation (Wei 1990), modelling the residuals about the trend as either a first-order autoregressive (AR(1)) process (for the Globe, Northern Hemisphere, North Atlantic and Nino region) or a third-order autoregressive (AR(3)) process (for the Southern Hemisphere and Indian Ocean). The trends in all realisations were then aggregated to produce a single trend estimate with an overall uncertainty. Calculated linear changes in the $50^{th}$ percentile average for each region in Figure 14 for 1850-2004 and 1901-2004 and their 95% confidence intervals are given in Table 1. Owing to the large interannual and interdecadal variability, a linear trend is not a good approximation of the behaviour of these time series and can give a false impression of the overall change. Thus, the uncertainty in the calculations arises mostly from the inter-decadal time-scale residuals about the trends and not from the uncertainties in the bias corrections or from under-sampling.

The poor fit of the linear trends can be seen most particularly in Figure 14(c)-(e). The consequent underestimation of the change in these non-linear time series can be largely overcome by instead fitting a smoothed curve to the annual data (see Figure 14) and calculating the difference between the start and end of the period. As is the case with the linear trend, these differences are highly dependent on the start and end of the period chosen. The result will also depend on the smoothing. Here we use the method employed by Folland et al. (2001b), which eliminates fluctuations with period less than a decade. Over the particularly non-linear period of 1850-2004 in the Southern Hemisphere, we calculate a change in SST from the filtered curve of 0.64±0.07°C, rather than 0.46±0.29°C from fitting the linear trend (see Table 1). The effect is similar in the Indian Ocean over this period. However, when the time series are more linear, e.g. between 1901 and 2004, the results converge. The exception in

this period is the North Atlantic average (Figure 14(d)), where the changes are mostly non-linear. The quasi-periodic fluctuations about the linear trend in this average have been termed the Atlantic Multidecadal Oscillation (Enfield et al. 2001) and have been shown to be related to climatic fluctuations over the USA, variations in North Atlantic hurricane activity (Goldenberg et al. 2001), decadal variations of Sahel rainfall (e.g. Folland et al. 1986) and may be a key component of the natural fluctuations of the thermohaline circulation (Knight et al. 2005), so are not variations that it would be appropriate to dismiss as uncertainty in a trend.

The goodness of fit of the linear changes is reflected in the size of the uncertainties (see Table 1), which are an order of magnitude larger than those of the filtered differences. The uncertainties in the filtered differences are a simple combination of the uncertainties in the annual values for the start and end of the period. Hence, we can now present a more accurate assessment of how SST has changed over any period, assuming that our annual uncertainties are an accurate reflection of the uncertainties in the data.

**5. Conclusions**

This paper presents a new flexible analysis of SST based on an improved data base (ICOADS) and assesses the validity of previous assumptions about biases. It presents comprehensive estimates of uncertainty arising from under-sampling of variability and the uncertainty in the estimated bias corrections. The new data base has refined our knowledge of changes in global and hemispheric averages and extended that analysis back to 1850.

The previously used bias corrections to SST are still largely valid, but some modification was required to those corrections between the late 1930s and early 1940s

and in the late nineteenth century. The former change was due to the incorporation of a large new data source (U.S. Merchant Marine).

Uncertainties in gridded SST anomalies due to under-sampling and measurement errors have been quantified for each grid box in each month. Locally, they dominate the bias-correction uncertainties in most regions at most times. Bias correction uncertainties have been calculated by exploring the assumptions made in their derivation and can be up to 30% of the size of the correction locally, so are not negligible, but, despite their spatial coherence, have only a modest effect on the uncertainties of calculated global and hemispheric trends.

The amalgamation of existing data bases and the addition of newly digitised data sources which formed the ICOADS data base, has been shown to be very beneficial to our knowledge of marine surface climate variability and change, both on local and global scales. However, there is still scope for improvement and many more data remain undigitised in archives around the world, not least in the U.K. at the National Archive, which hopefully will be digitised in the near future.

We propose a method for summarising changes in temperature over any particular period, which takes account of non-linear fluctuations in the time series and allows us to estimate those changes with greater accuracy.

The new analysis presented in this paper is updated every month and can be obtained from http://www.hadobs.org

**6. Remaining issues**

Using metadata in the ICOADS it is possible to compare the contributions made by different countries to the marine component of the global temperature curve. Different countries give different advice to their observing fleets concerning how best to measure SST. Breaking the data up into separate countries' contributions shows

that the assumption made in deriving the original bucket corrections, i.e. that the use of uninsulated buckets ended in January 1942, is incorrect. In particular, data gathered by ships recruited by Japan and the Netherlands (not shown) are biased in a way that suggests that these nations were still using uninsulated buckets to obtain SST measurements as late as the 1960s. By contrast, it appears that the U.S. started the switch to using engine room intake measurements as early as 1920 (Section 3a).

So, the next step will be to revisit the SST bias corrections and refine them, making use of the new information uncovered concerning national measurement practices and new analysis techniques that allow for more accurate corrections in areas and at times where there are few data. In particular, small post-1941 corrections will be developed to take account of the Japanese and Dutch practices. Because the Dutch data make a relatively small contribution to the total number of observations in the 1950s and 1960s and because the area of major influence of the Japanese data at the same time is limited to the North Pacific, these relatively small biases have not been corrected for here.

From the late 1970s there has been a steadily increasing number of SST observations made by buoys. By 1997, buoy observations made up around 65% of all the observations in the ICOADS (see purple shaded area in Figure 2). Measurements made by buoys are generally biased cold by around 0.1 to 0.2˚C relative to measurements made by ships (not shown). The size of this bias varies regionally. There are also systematic differences between reports from moored and drifting buoys. Moored buoys off the east coast of the U.S. are biased cold relative to SST from ships by 0.4ºC and relative to nighttime marine air temperature (HadNAT2, see Section 3a) by 0.5-0.6ºC since the late 1980s (see Figure 15(a)). Moored buoys off the west coast of the U.S. are biased cold relative to ship SST and NMAT by 0.2-0.5ºC

(Figure 15(b)). Separating these U.S. moored buoys into their different types (3m DISCUS, 6m NOMAD, 10m DISCUS and 12m DISCUS), shows that none is relatively unbiased (not shown). By contrast, moored buoys in the equatorial Pacific (Figure 15(c)) and around the U.K. (Figure 15(d)) are within 0.1ºC and 0.2°C of ship SST after 1990, respectively. SST from some buoys is clearly more biased than from others relative to ship SST and so might reasonably be excluded from the analysis but, then good data could be excluded along with bad. Consequently, our analysis includes all buoy SST in ICOADS passing our QC tests.

Recent work (Kent and Taylor, 2005; Kent and Kaplan, 2005) has shown that modern SST buckets also lose heat, especially when air-sea temperature differences are large. They also show that engine intake SST has been biased warm in the past, but more recently (in the 1990s) shows a slightly cold bias. A further step will thus be to apply the requisite corrections.

These extensions of the corrections will be aided by the extra wealth of metadata concerning measurement methods available in the ICOADS and in the WMO Publication No. 47 (WMO, 1955-2004) for data from the second half of the twentieth century. Early issues have recently been digitized and their content assessed in Kent et al. (2005, manuscript submitted to *J. Atmos. Oceanic Technol.)*. These new concerns are not currently reflected in our uncertainty estimates, but are likely to make only a small contribution to their overall size. The improved modern data will allow the computation of improved climatologies and, in turn, improved historical anomalies.

For now, our combined sampling and measurement error estimates (Section 3b) are likely to include contributions from these remaining biases.

## APPENDIX

### Uncertainty in inputs to bias corrections

Uncertainty in the speed of ships was obtained from literature cited in FP95. For each period cited, the standard error in average ships' speed is ~0.2 m $s^{-1}$. This translates to an uncertainty in proportion of fast (7m $s^{-1}$), or slow (4m $s^{-1}$), ships of 0.07 (i.e. 7%), because 0.2m $s^{-1}$ is about 7% of the difference between 7m $s^{-1}$ and 4m $s^{-1}$.

It is assumed that the only uncertainty in the modelled corrections for the wooden buckets stems from the averaging of fields of corrections pertaining to thick and thin buckets. The standard deviation (and hence the standard error) of the annual mean corrections given in Table 1(b) of FP95 is ~30% for both the fast and slow ship cases. So, an uncertainty of 30% of the value of the wooden bucket correction field in any grid box is assigned to that grid box.

The uncertainty in the average canvas bucket correction fields has two components. Firstly, as for wooden buckets, there is a contribution from the averaging of corrections for different bucket sizes: the standard deviation of the idealised annual mean corrections listed in Table 1(a) of FP95 is ~4% for both the fast and slow ship cases, confirmed to be globally applicable by inspection of their Figure 17. However, the main contribution stems from the choice of integration time of the models, obtained by minimising the ratio of annual cycle variance to total variance of monthly SST anomalies in certain areas and periods. Figure 16 of FP95 plots the spread of possible integration times for one model. From this, and an assumption that all these possible times are independently arrived at, an uncertainty of ~13% in integration time is inferred. Assuming a linear relationship between integration time and resultant correction field leads to an uncertainty in the canvas correction fields of 13% of the

correction. Now, each of the four correction fields (for models using different sized buckets and different exposure to the sun) which contribute to the average correction field for canvas buckets on (for example) fast ships has an integration-time-related uncertainty of 13% of its value at each grid point. Combining this in quadrature with the spread between the four models (our previously calculated 4%), a combined uncertainty of 13.6% of the size of the canvas corrections at each grid box is obtained from these two factors. Thus total uncertainty in the canvas correction fields is dominated by uncertainty in the length of time elapsed between obtaining the water sample and taking the measurement.

Sampling and measurement uncertainties have been estimated for the NMAT data using the same method as for SST (see Section 3b).

**Acknowledgements**

The ICOADS and NCEP GTS databases and ICOADS gridded summaries were provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their website at http://www.cdc.noaa.gov/. We wish to thank Gil Compo, Briony Horton, Peter Stott, Scott Woodruff, Steve Worley and David Sexton for useful discussions and suggestions, and Jim Arnott for preliminary work on the QC and gridding system. The suggestions of two anonymous reviewers helped to clarify the text. This work was funded by the U.K. Government Meteorological Research contract and by the U.K. Department for Environment, Food and Rural Affairs, under Contract PECD/7/12/37. This work is British Crown Copyright.

**References**

Afifi, A.A. and S.P. Azen, 1979: *Statistical analysis. A computer oriented approach*. Academic Press, New York, pp xx+442.

Allen, M.R. and P.A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part 1: theory. *Climate Dyn.*, **21**, 493-500.

Barton, A.D. and K.S. Casey, 2005: Climatological context for large-scale coral bleaching. Coral Reefs, *in press.*

Bottomley, M., C.K. Folland, J. Hsiung, R.E. Newell, and D.E. Parker, 1990: *Global ocean surface temperature atlas*. Her Majesty's Stn. Off., Norwich, UK. 20 + iv pp. and 313 color plates,

Enfield, D.B., A.M. Mestas-Nuñez and P.J Trimble, 2001: The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental US. *Geophys. Res. Lett.*, **28**, 2077-2080.

Fiorino, M., 2004: A multi-decadal daily sea surface temperature and sea ice concentration data set for the ERA-40 Reanalysis. ECMWF ERA-40 Project Report Series No. 12, European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, RG2 3AX, U.K.

Folland, C.K., 2005: Tests of bias corrections to sea surface temperature using a climate model. *Int. J. Clim.*, **25**, 895-911.

Folland, C.K., D.E. Parker and T.N. Palmer, 1986: Sahel rainfall and worldwide sea temperatures 1901-85. *Nature*, **320**, 602-607.

Folland, C.K. and D.E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319-367.

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. Nicholls, and D.M.H. Sexton, 2001a: Global temperature change and its uncertainties since 1861. *Geophys. Res. Lett.*, **28**, 2621-2624.

Folland, C.K. and Coauthors, 2001b: Observed climate variability and change in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. I. Johnson, Eds., Cambridge University Press, chapter 2, 99-181.

Folland, C.K., M.J. Salinger, N. Jiang and N.A. Rayner, 2003: Trends and variations in South Pacific island and ocean surface temperatures. *J. Climate*, **16**, 2859-2874.

Goldenberg, S.B, C.W. Landsea, A.M. Mestas-Nuñez and W.M. Gray, 2001: The recent increase in Atlantic Hurricane activity: causes and implications. *Science*, **293**, 474-479.

Gregory, J.M., R.J. Stouffer, S.C.B. Raper, P.A. Stott and N.A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Climate*, **15**, 3117-3121.

Hanawa, K., S. Yasunaka, T. Manabe, and N. Iwasaka, 2000: Examination of correction to historical SST data using long-term coastal SST data taken around Japan. *J. Meteor. Soc. Jpn.*, **78**, 187-195.

Hegerl, G.C., P.D. Jones and T.P. Barnett, 2001: Effect of observational sampling uncertainty on the detection of anthropogenic climate change. *J. Climate*, **14**, 198-207.

Hurrell, J.W., S.J. Brown, K.E. Trenberth and J.R. Christy, 2000: Comparison of tropospheric temperatures from radiosondes and satellites: 1979-98. *Bull. Am. Meteor. Soc.*, **81**, 2165-2177.

Hurrell, J.W. and K.E. Trenberth, 1998: Difficulties in obtaining reliable temperature trends: reconciling the surface and satellite microwave sounding unit records. *J. Climate*, **11**, 945-967.

Ishii, M., M. Kimoto and M. Kachi, 2003: Historical ocean subsurface temperature analysis with error estimates. *Mon. Wea. Rev.*, **131**, 51-73.

Ishii, M., A. Shouji, S. Sugimoto and T. Matsumoto, 2005: Objective analyses of SST and marine meteorological variables for the 20$^{th}$ century using ICOADS and the Kobe Collection. *Int. J. Clim.*, **25**, 865-879.

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

Jones, P.D. and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206-223.

Kagan, R.L., 1966: *An Evaluation of the Representativeness of Precipitation Data* (in Russian). Gidrometeoizdat, 191 pp.

Kent, E.C. and D.I. Berry, 2005: Quantifying random measurement errors in voluntary observing ships meteorological observations. *Int. J. Clim.*, **25**, 843-856.

Kent, E.C. and P.G. Challenor, 2005: Towards estimating climatic trends in SST, part 2: random errors. *J. Atmos. Oceanic. Technol.*, in press

Kent, E.C. and A. Kaplan, 2005: Towards estimating climatic trends in SST, part 3: systematic biases. *J. Atmos. Oceanic. Technol.*, in press

Kent, E.C. and P.K. Taylor, 2005: Towards estimating climatic trends in SST, part 1: methods of measurement. *J. Atmos. Oceanic. Technol.*, in press

Kestin, T.S., D.J. Karoly, J.I. Jano and N.A. Rayner, 1998: Time-frequency variability of ENSO and stochastic simulations. *J. Climate*, **11**, 2258-2272.

Knight, J.R., R.J. Allan, C.K. Folland, M. Vellinga and M.E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate, *accepted for publication in Geophys. Res. Lett.*

Levitus, S., J.I. Antonov and T.P. Boyer, 1994: Interannual variability of temperature at a depth of 125 meters in the North Atlantic Ocean. *Science*, **266**, 96-99.

Levitus, S., J.I. Antonov, T.P. Boyer and C. Stephens, 2000: Warming of the world ocean. *Science*, **287**, 2225-2229.

Maury, M.F., 1858: *Explanations and sailing directions to accompany the wind and current charts*. Vol.I, Cornelius Wendell, Washington, pp.383+51pls.

Maury, M.F., 1859: *Explanations and sailing directions to accompany the wind and current charts*. Vol.II, Cornelius Wendell, Washington, pp.874+7pls.

McAvaney, B.J. and Coauthors, 2001: Model evaluation in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. I. Johnson, eds., Cambridge University Press, chapter 8, 471-523.

Nicholls, N., G.V. Gruza, J. Jouzel, T.R. Karl, L.A. Ogallo and D.E. Parker, 1996: Observed climate variability and change in *Climate change 1995: The Science of Climate Change. Contribution of Working Group I to the second assessment report*

*of the Intergovernmental Panel on Climate Change*, edited by J.T. Houghton et al, pp133-192.

Parker, D.E., C.K. Folland and M. Jackson, 1995a: Marine surface temperature: observed variations and data requirements. *Climatic Change*, **31**, 559-600.

Parker, D.E., M. Jackson, and E.B. Horton, 1995b: The GISST2.2 sea surface temperature and sea ice climatology. *Clim. Res. Tech. Note CRTN 63*, Hadley Centre, Met Office, Exeter, EX1 3PB, UK.

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 SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, 10.1029/2002JD002670

Reynolds, R.W., 1988: A real-time global sea surface temperature analysis. *J. Climate*, **1**, 75-86.

Sexton, D.M.H., H. Grubb, K.P. Shine and C.K. Folland, 2003: Design and analysis of climate model experiments for the efficient estimation of anthropogenic signals. *J. Climate*, **16**, 1320-1336.

Smith, T.M. and R.W. Reynolds, 2002: Bias corrections for historical sea surface temperatures based on marine air temperatures. *J. Climate*, **15**, 73-87.

Smith, T.M. and R.W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854-1997). *J. Climate*, **16**, 1495-1510.

Smith, T.M. and R.W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466-2477.

Stendel, M., J.R. Christy and L. Bengtsson, 2000: Assessing levels of uncertainty in recent temperature time series. *Climate Dyn.*, **16**, 587-601.

Taylor, K.E., D. Williamson, and F. Zwiers, 2000: The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations. *PCMDI Report No. 60*, Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, 25 pp

Thorne, P.W., P.D. Jones, S.F.B. Tett, M.R. Allen, D.E. Parker, P.A. Stott, G.S. Jones, T.J. Osborn, and T.D. Davies, 2003: Probable causes of late twentieth century tropospheric temperature trends. *Climate Dyn.*, **21**, 573-591.

Thorne, P.W., D.E. Parker, J.R. Christy, and C.A. Mears, 2005: Uncertainties in climate trends: lessons from upper-air temperature records. *In press in Bull. Am. Met. Soc.*

Trenberth, K.E., J.R. Christy and J.W. Hurrell, 1992: Monitoring global monthly mean surface temperatures. *J. Climate*, **5**, 1405-1423.

Wei, W.S., 1990: *Time Series Analysis: Univariate And Multivariate Methods*, Addison Wesley, 624pp.

World Meteorological Organization (WMO), *International list of selected, supplementary and auxiliary ships*, WMO Publ. 47, Geneva, 1955-2004.

Worley, S.J, S.D. Woodruff, R.W. Reynolds, S.J. Lubker and N. Lott, 2005: ICOADS release 2.1 data and products, *Int. J. Clim.*, **25**, 823-842

Yevjevich, V., 1972: *Probability and Statistics in Hydrology*. Water Resources Publications, 302 pp.

**Figure Captions**

**Figure 1.** Example improvement in data coverage for decades 1850-59 (top) and 1910-19 (bottom): (a) and (c) MOHSST6D (b) and (d) HadSST2. Contours are of the proportion of months containing data in each decade at each grid box.

**Figure 2.** (a) Number and (b) fraction of SST observations in the ICOADS, 1826-1997, split between different sources. Sources of ship data are identified via the "deck id" metadata in ICOADS and grouped according to country. The "miscellaneous international" collection is a grouping of decks which are large international collections. "Other" refers to non-ship data. Colours refer to the same sources in both panels.

**Figure 3.** Comparison of HadSST2 and GISST2.0 1961-90 SST climatologies in January (top) and July (middle): (a) and (c) HadSST2; (b) and (d) HadSST2 – GISST2.0. Also shown are annual cycles in the climatologies and the 1961-90 average of *in situ* data input to HadSST2 at: (e) [30ºS, 90ºW] and (f) [0,0].

**Figure 4.** The effect of QC on SST observations for April 1973, a month characterised by particularly poor quality data: (a) all observations; (b) observations passing QC. 19% of the observations were flagged as bad; a much larger fraction than is usual (see text).

**Figure 5.** HadSST2 North Atlantic SST fields (ºC), December 1997, at various spatial resolutions: (a) 5º latitude by 5º longitude; (b) 2.5º latitude by 3.75º longitude; (c) 1º latitude by 1º longitude and (d) 0.5º latitude by 0.5º longitude.

**Figure 6.** Seasonal cycle of SST bias correction uncertainty, as given by the difference between the 97.5$^{th}$ percentile and median bucket correction (ºC), taken from the 1000 realisations for 1938: (a) January; (b) April; (c) July and (d) October.

**Figure 7.** Adaptation of bias corrections for SST, 1938-41: (a) uncorrected global average SST in HadSST2 and MOHSST6, 1930-46; (b) FP95 versus adapted bias corrections, 1856-1941; (c) difference in global mean SST in HadSST2 with and without the recently digitised U.S. Merchant Marine SST versus global mean alterations to bucket corrections (multiplied by -1), 1910-45 and (d) comparison of corrected global average SST in HadSST2 and MOHSST6, 1850-2004.

**Figure 8.** Comparison of global area-weighted mean bucket corrections (ºC), 1850-1941: FP95 bucket corrections (from 1856, thin black line) and HadSST2 (thick black line). The grey shading is the 95% confidence interval of HadSST2 corrections.

**Figure 9.** Time series of global SST data availability for each month 1826-2004: (a) number of 5º area grid boxes containing data; (b) number of observations.

**Figure 10.** SST anomaly time series at 0-5°S, 30-35°W: (a) monthly bias-corrected SST anomaly; (b) as (a), but detrended (black solid line) along with number of observations (grey dots) used each month; (c) standard deviation of detrended SST anomalies (°C) binned by number of observations.

**Figure 11.** Results of fit of standard deviation versus number of observations relationship to HadSST2: (a) estimated true standard deviation (°C); (b) sampling and measurement uncertainty assuming one observation used in each monthly 5° area grid box; (c) actual sampling and measurement uncertainty (1 s.e., °C) for September 1853 and (d) as (c), but for September 2003.

**Figure 12.** HadSST2 SST uncertainties (2 s.e., ºC) due to under-sampling and measurement errors ((a) and (e)), bias corrections ((b) and (f)) and a combination of the two ((c) and (g)). (d) and (h) depict the ratio of bias correction uncertainty to combined sampling and measurement uncertainty. (a)-(d) September 1863 and (e)-(h) September 1938.

**Figure 13.** Effect on uncertainty in SST of inclusion of recently digitised data in ICOADS: (a) number of 5° area grid boxes containing data, 1850-2003; (b) global area-weighted rms sampling and measurement uncertainty (1 s.e., °C) with and without new data sources; (c) grid box measurement and sampling uncertainties (1 s.e., °C), June 1940, without new data sources and (d) as (c) but with new data sources.

**Figure 14**. Large-scale annual SST anomalies (ºC, relative to 1961-90), 1850-2004: (a) Globe; (b) Northern Hemisphere; (c) Southern Hemisphere; (d) North Atlantic; (e) Indian Ocean and (f) [10°N-10°S; 180-120°W]. Black: HadSST2(OA). Green: simple HadSST2 average ((a)-(c) only). Red: Folland et al. (2001b), denoted IPCC TAR ((a)-(c) only). Cyan: ICOADS enhanced summaries with Smith and Reynolds (2002) bias corrections ((a)-(c) only). Blue shaded areas are 95% confidence intervals of HadSST2(OA). Smooth black line is HadSST2(OA) smoothed with a 21-point binomial filter (see text). Linear trends are fit to the HadSST2(OA) time series (see text).

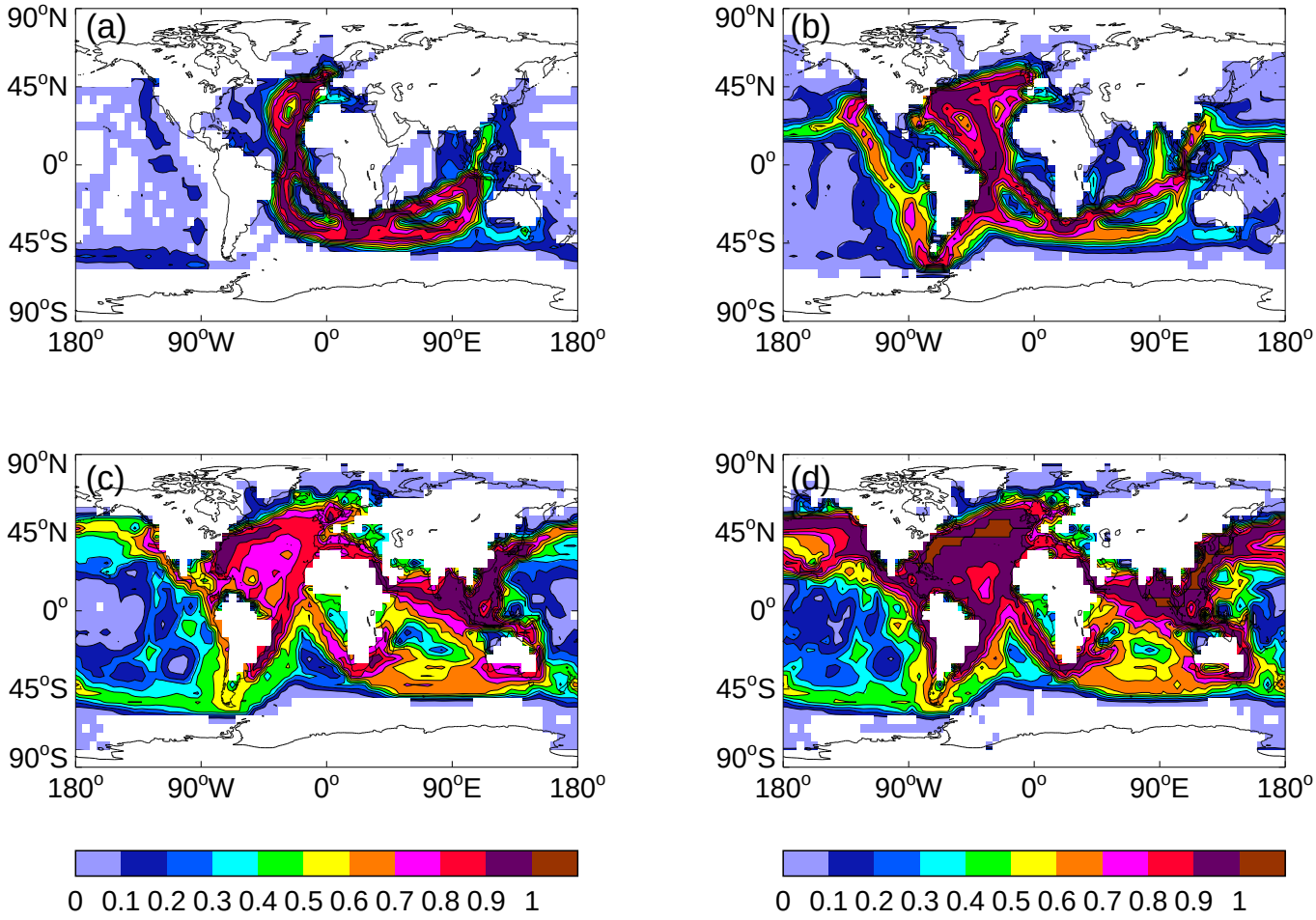
**Figure 15.** Comparison of SST anomalies (ºC, relative to 1961-90) from moored buoys with SST from ships and NMAT (HadNAT2), 1970-2004: (a) east coast of U.S.; (b) west coast of U.S.; (c) equatorial Pacific and (d) around the U.K. Smoothed with low-pass filter to isolate variability on timescales of at least 5 years, monthly values also shown for the moored buoys. Anomalies are relative to climatologies for 1961-90: HadNAT2 relative to HadNAT2 and moored buoys and ships relative to HadSST2.

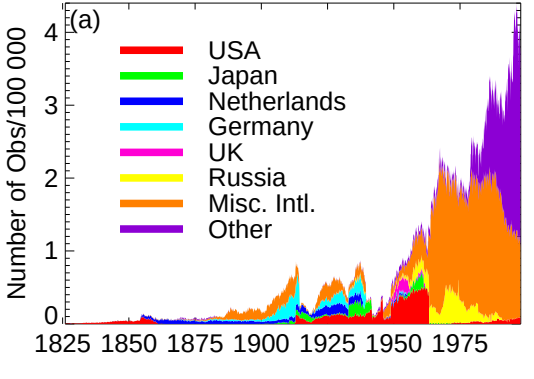
**Table 1.** Details of erroneous MARMET data removed explicitly from HadSST2.

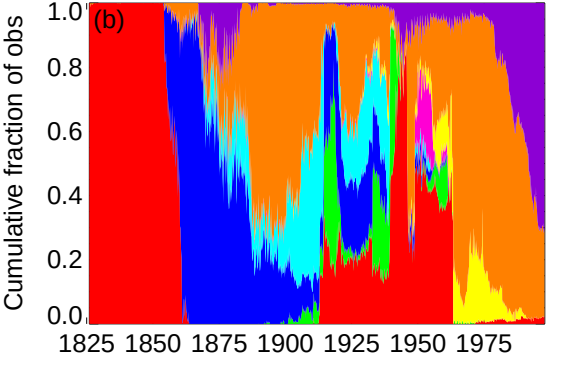
| Region | Period |
|---|---|
| North Pacific [40-45°N, 170-175°W] | 1960s |
| [40-45°N, 160-165°W] | 1960s |
| Cape of Good Hope [37-40°S, 30-34°E] | April-August 1971 |
| [37-40°S, 6-9°E] | March 1961-66, 1969, 1971-74 |

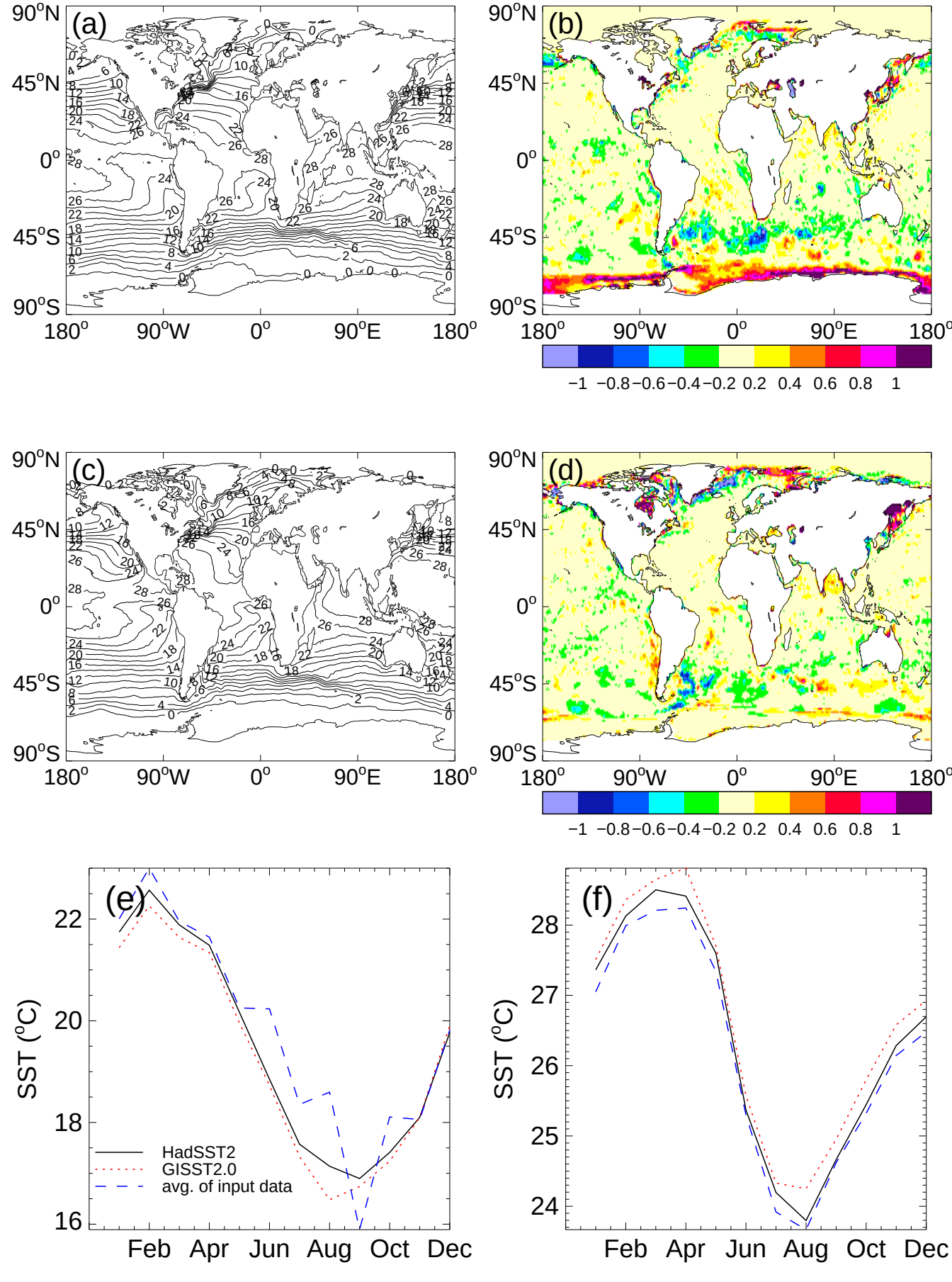
**Table 2.** SST anomaly changes (°C, relative to 1961-90) and 95% confidence intervals of large-scale and regional averages over the periods indicated. The 1850-2004 change for the Nino region is not calculated because of data voids in the late nineteenth century.

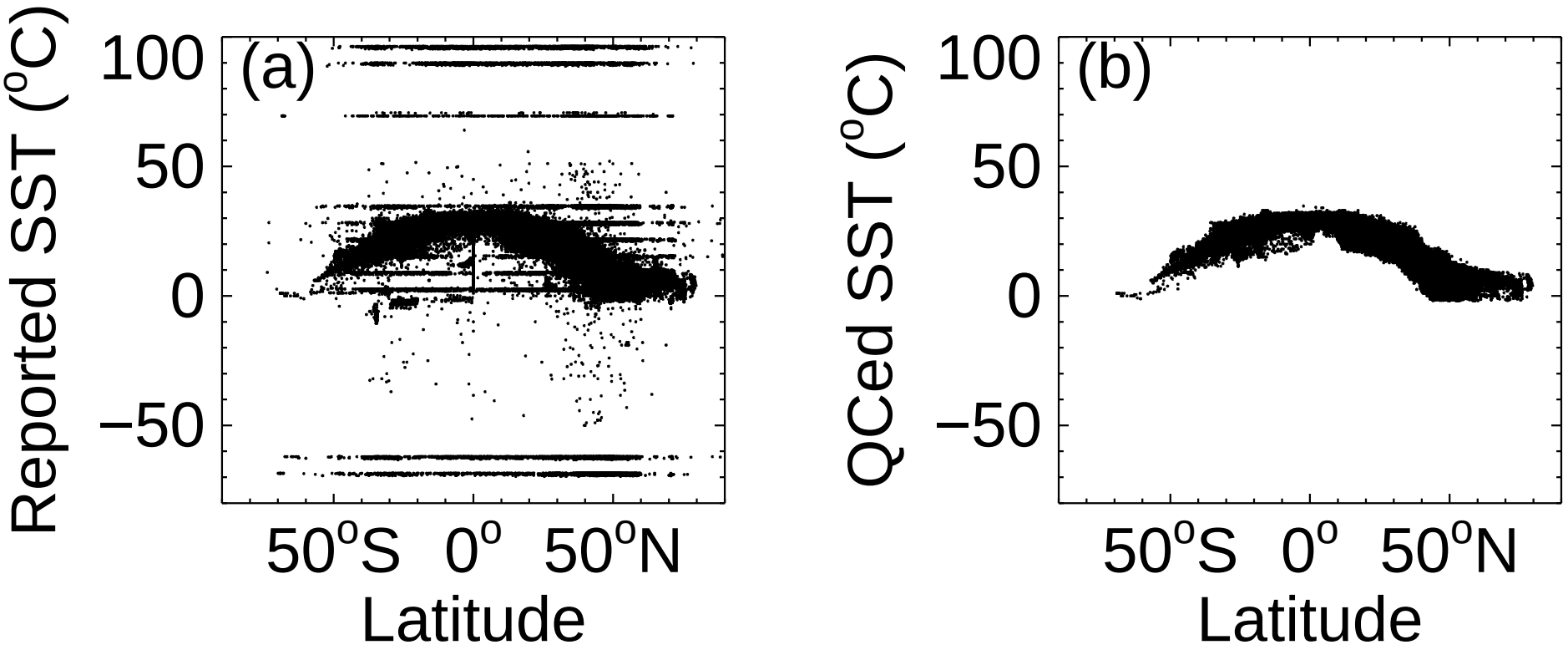
| Region | Temperature change in linear trend | | Temperature change in filtered curve | |
|---|---|---|---|---|
| | 1850-2004 | 1901-2004 | 1850-2004 | 1901-2004 |
| Globe | 0.52±0.19 | 0.68±0.13 | 0.67±0.04 | 0.67±0.02 |
| Northern Hemisphere | 0.59±0.20 | 0.66±0.19 | 0.71±0.06 | 0.74±0.03 |
| Southern Hemisphere | 0.46±0.29 | 0.68±0.18 | 0.64±0.07 | 0.63±0.03 |
| North Atlantic | 0.48±0.23 | 0.58±0.27 | 0.59±0.06 | 0.76±0.04 |
| Indian Ocean | 0.35±0.35 | 0.72±0.22 | 0.56±0.08 | 0.75±0.04 |
| Nino region (10°N-10°S, 180-120°W) | Not calculated | 0.49±0.42 | Not calculated | 0.24±0.10 |

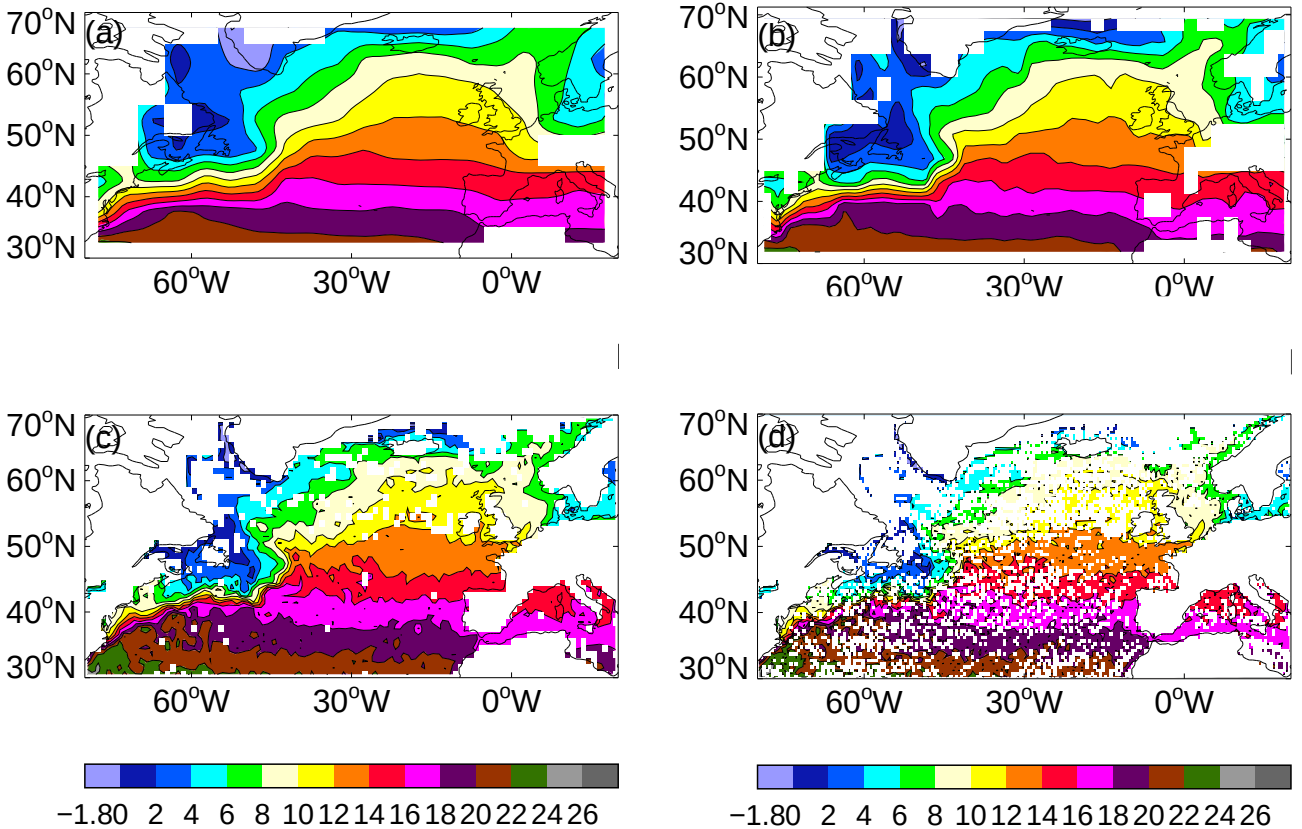

(a)
(b)
(c)
(d)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

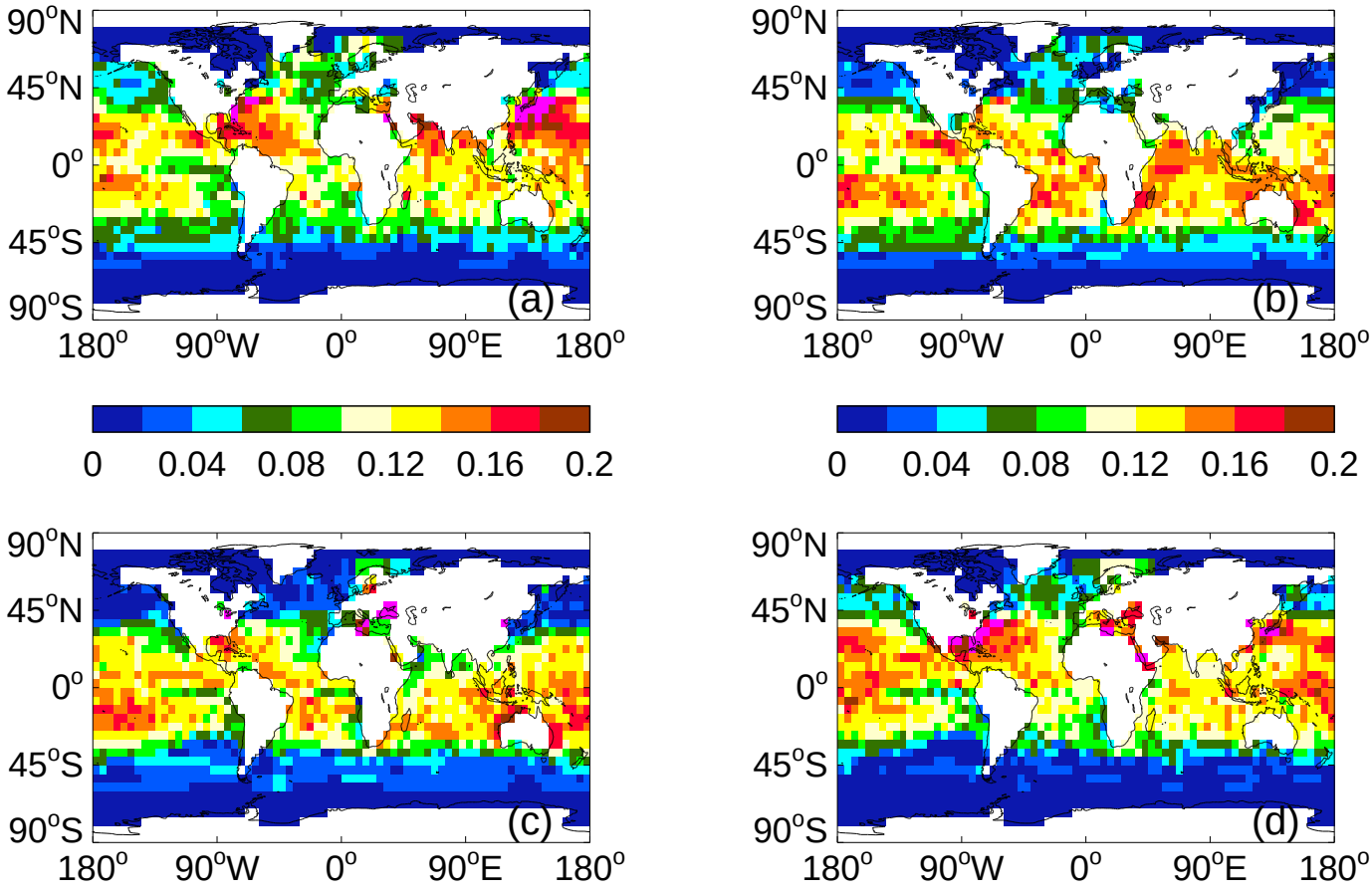

(a)
Number of Obs/100 000
0
1
2
3
4
USA
Japan
Netherlands
Germany
UK
Russia
Misc. Intl.
Other
1825 1850 1875 1900 1925 1950 1975

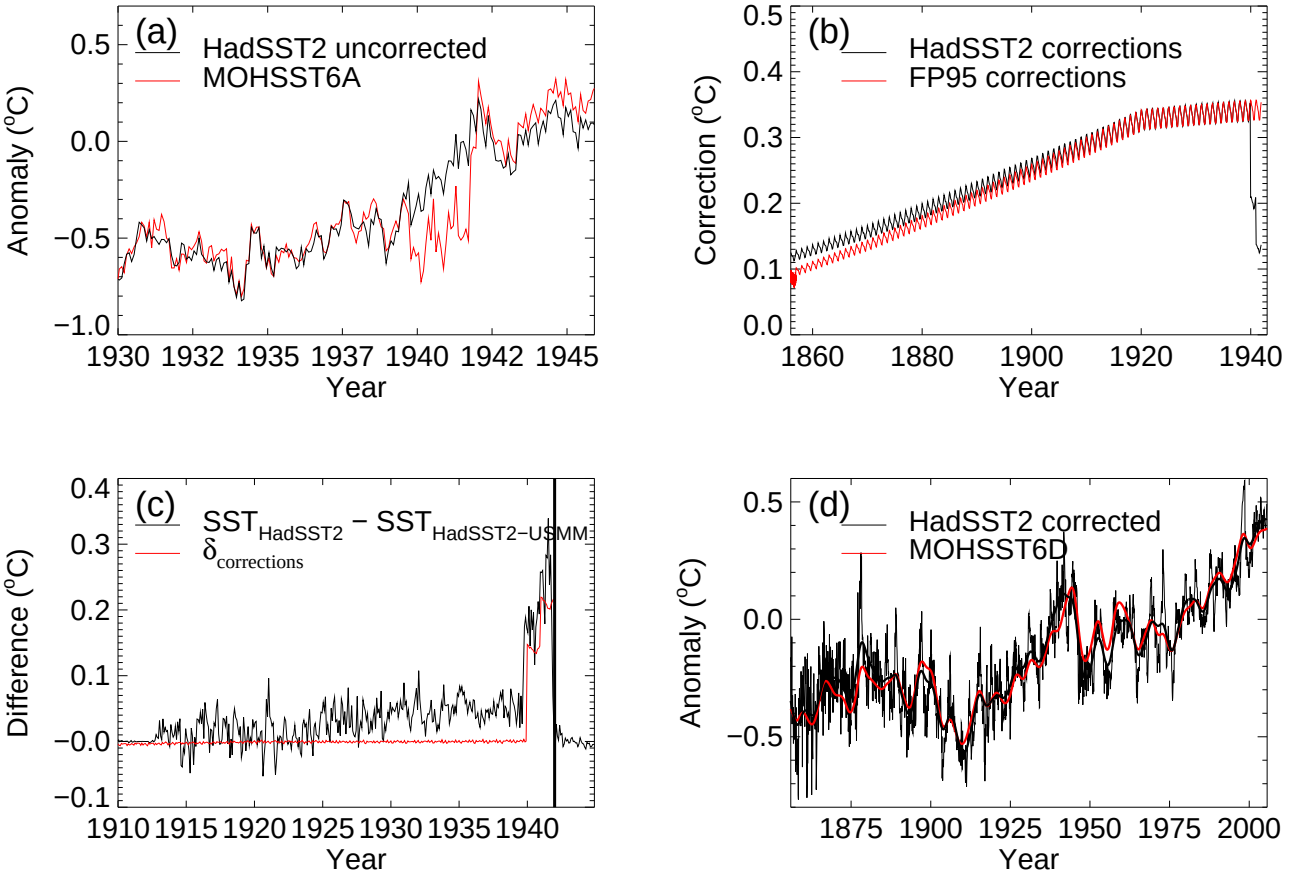

(b)
Cumulative fraction of obs
0.0
0.2
0.4
0.6
0.8
1.0
1825 1850 1875 1900 1925 1950 1975

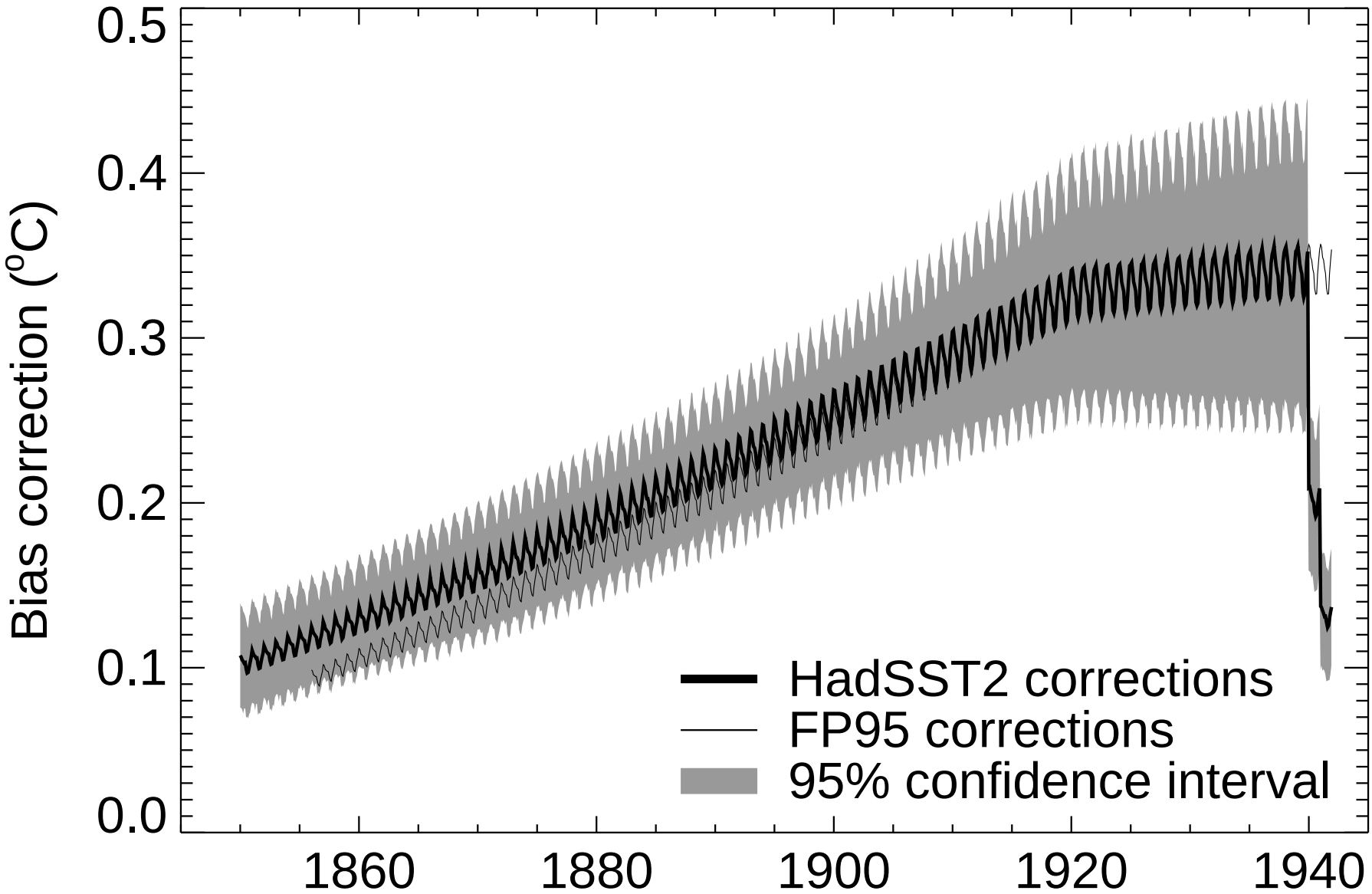

(a)
(b)
(c)
(d)
(e)
(f)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
−1 −0.8 −0.6 −0.4 −0.2 0.2 0.4 0.6 0.8 1
SST (°C)
HadSST2
GISST2.0
avg. of input data
Feb Apr Jun Aug Oct Dec

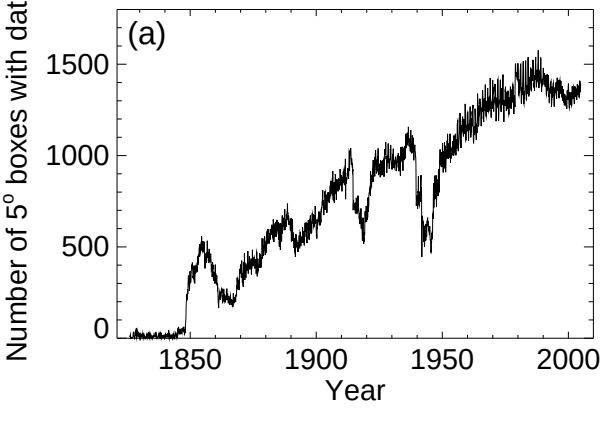

(a)
(b)
Reported SST (°C)
QCed SST (°C)
100
50
0
−50
50°S
0°
50°N
Latitude

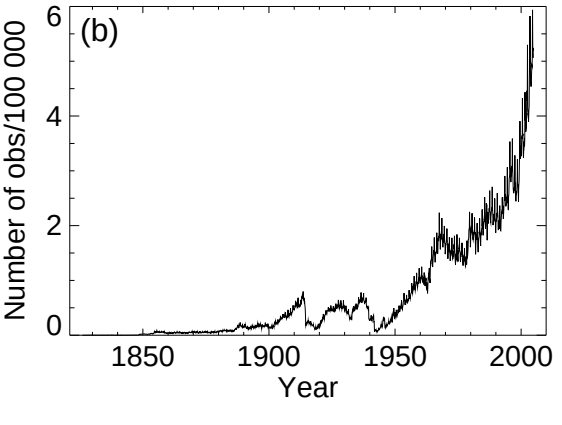

(a)
(b)
(c)
(d)
70°N
60°N
50°N
40°N
30°N
60°W
30°W
0°W
−1.8 0 2 4 6 8 10 12 14 16 18 20 22 24 26

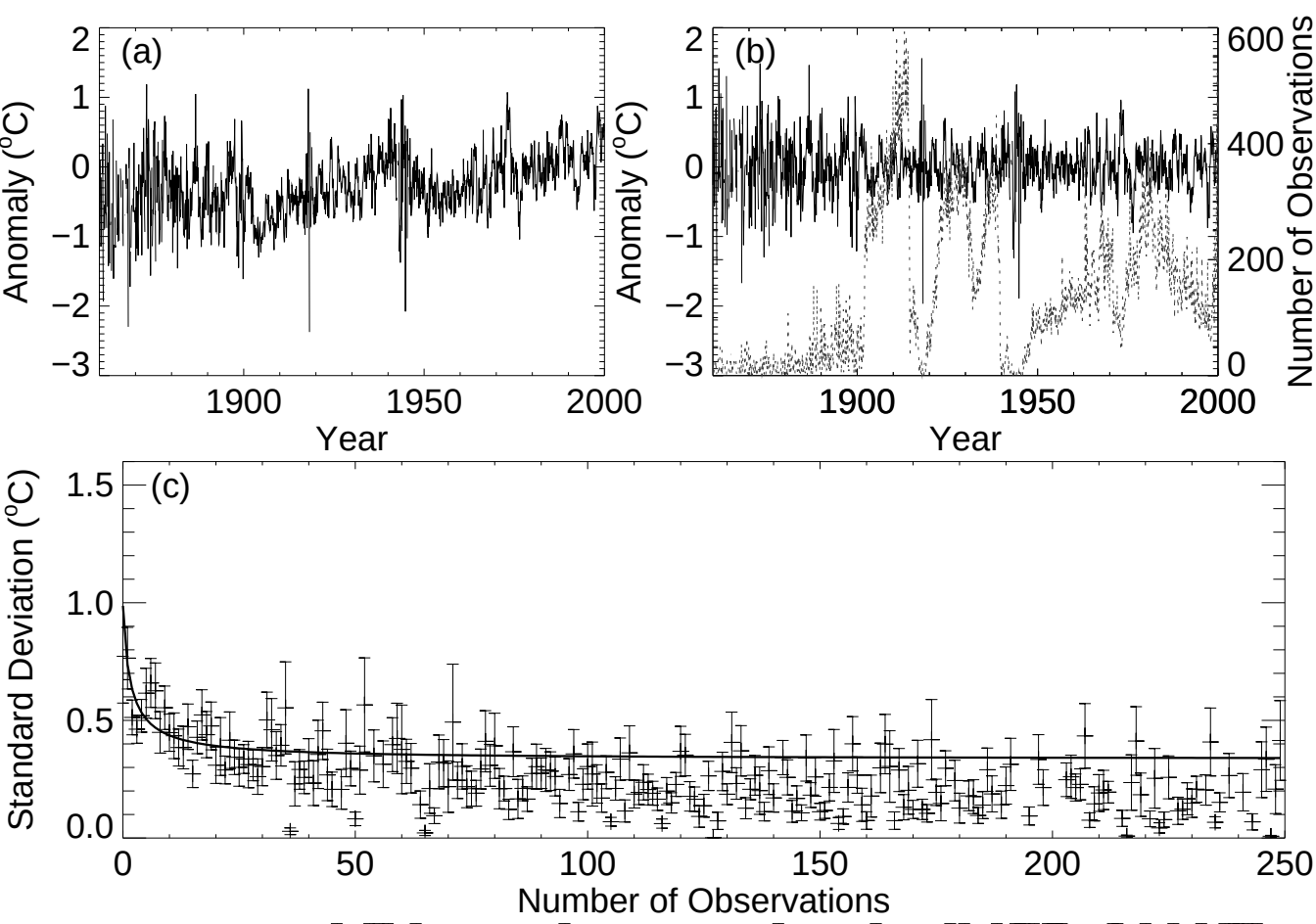

90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
(a)
(b)
(c)
(d)
0
0.04
0.08
0.12
0.16
0.2

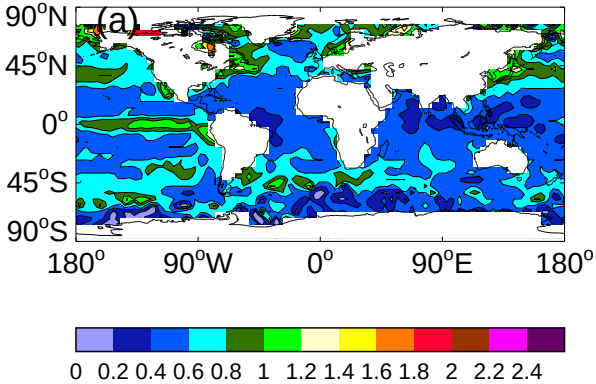

(a)
HadSST2 uncorrected
MOHSST6A
Anomaly (°C)
0.5
0.0
−0.5
−1.0
1930 1932 1935 1937 1940 1942 1945
Year
(b)
HadSST2 corrections
FP95 corrections
Correction (°C)
0.5
0.4
0.3
0.2
0.1
0.0
1860 1880 1900 1920 1940
Year
(c)
SST HadSST2 − SST HadSST2−USMM
δ corrections
Difference (°C)
0.4
0.3
0.2
0.1
−0.0
−0.1
1910 1915 1920 1925 1930 1935 1940
Year
(d)
HadSST2 corrected
MOHSST6D
Anomaly (°C)
0.5
0.0
−0.5
1875 1900 1925 1950 1975 2000
Year

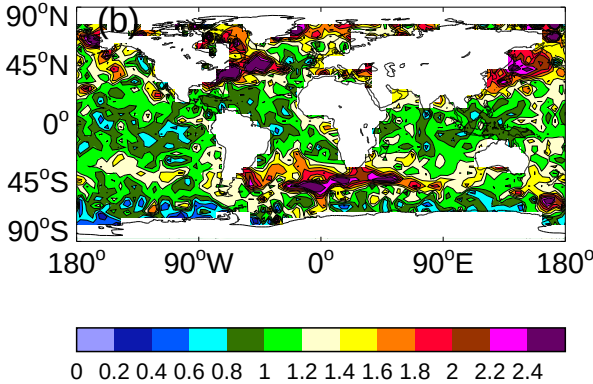

Bias correction (°C)
0.5
0.4
0.3
0.2
0.1
0.0
1860
1880
1900
1920
1940
HadSST2 corrections
FP95 corrections
95% confidence interval

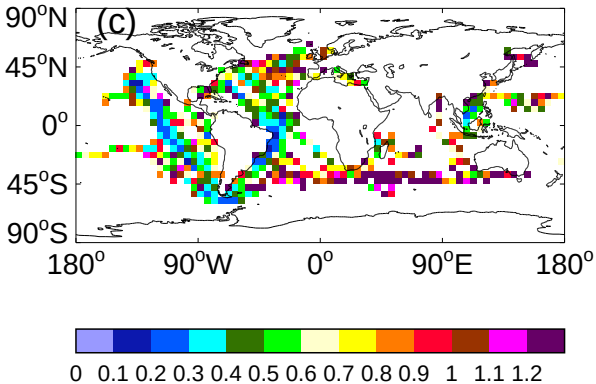

(a)
Number of 5° boxes with data
0
500
1000
1500
1850
1900
1950
2000
Year

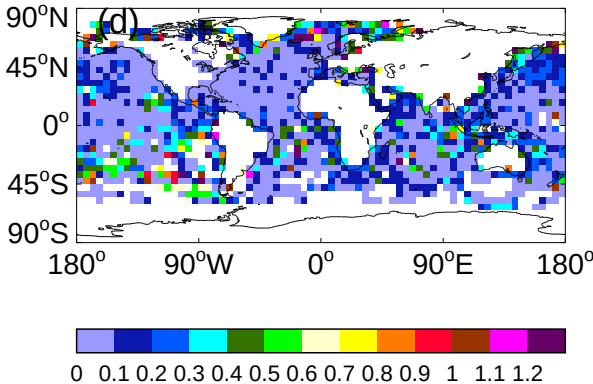

(b)
Number of obs/100 000
0
2
4
6
1850
1900
1950
2000
Year

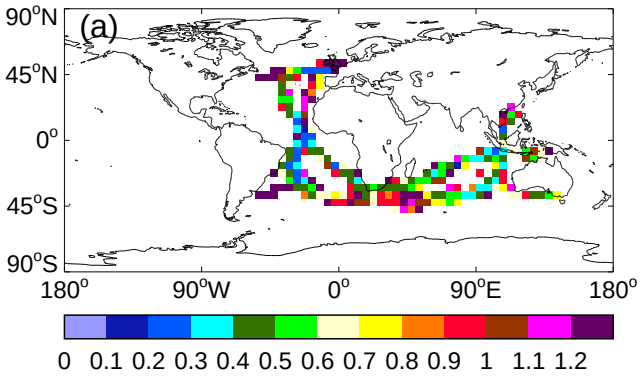

(a)
Anomaly (°C)
Year
(b)
Anomaly (°C)
Number of Observations
Year
(c)
Standard Deviation (°C)
Number of Observations

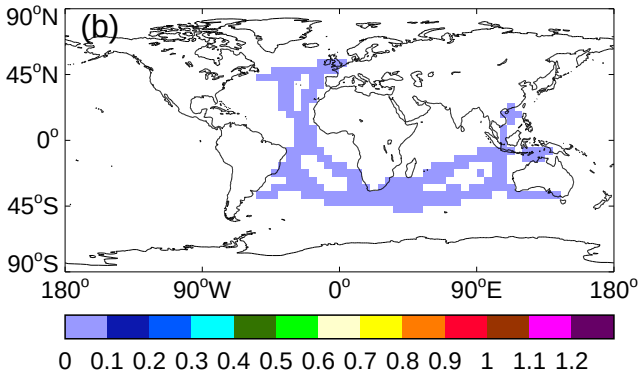

(a)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4

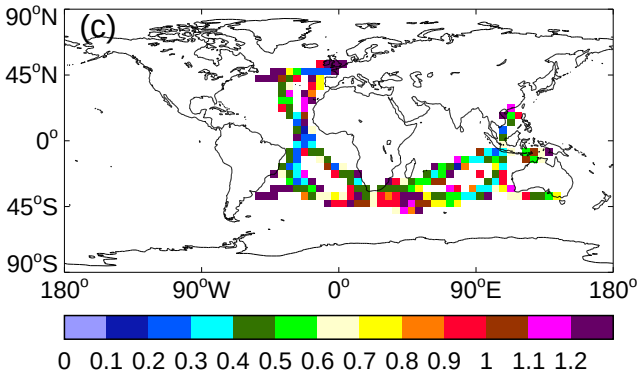

(b)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4

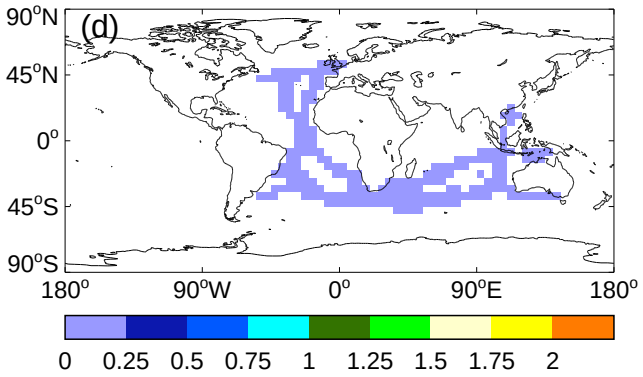

(c)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

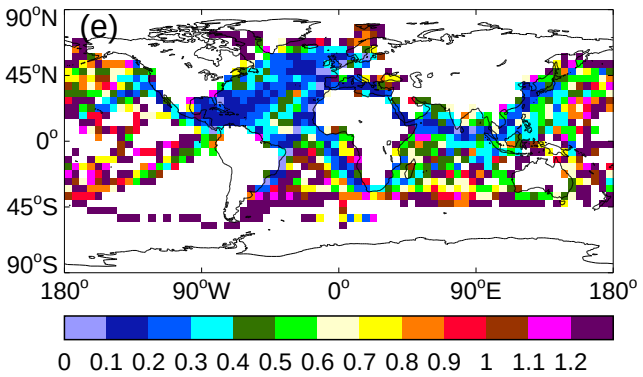

(d)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

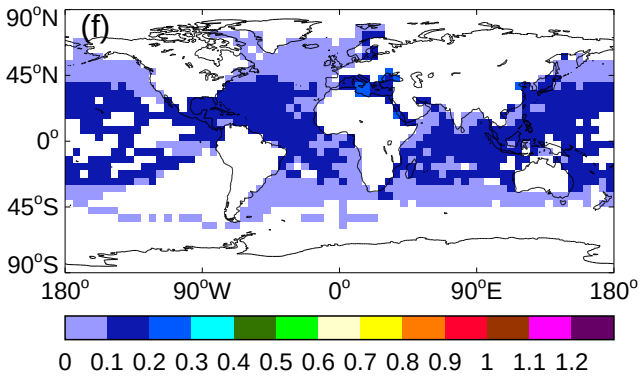

(a)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

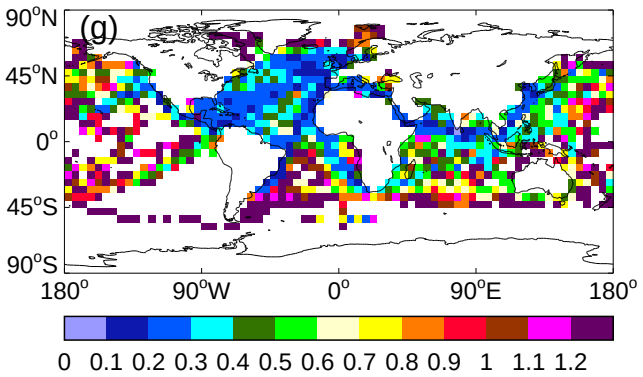

(b)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

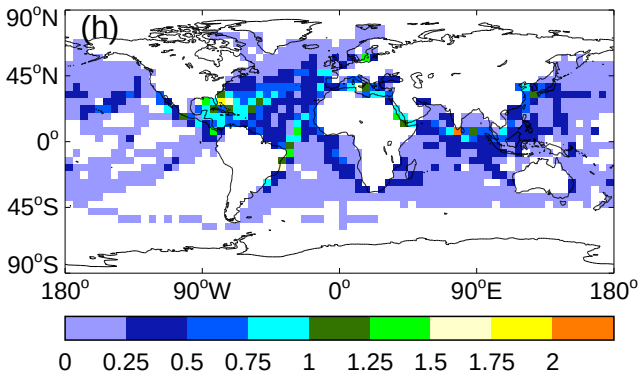

(c)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

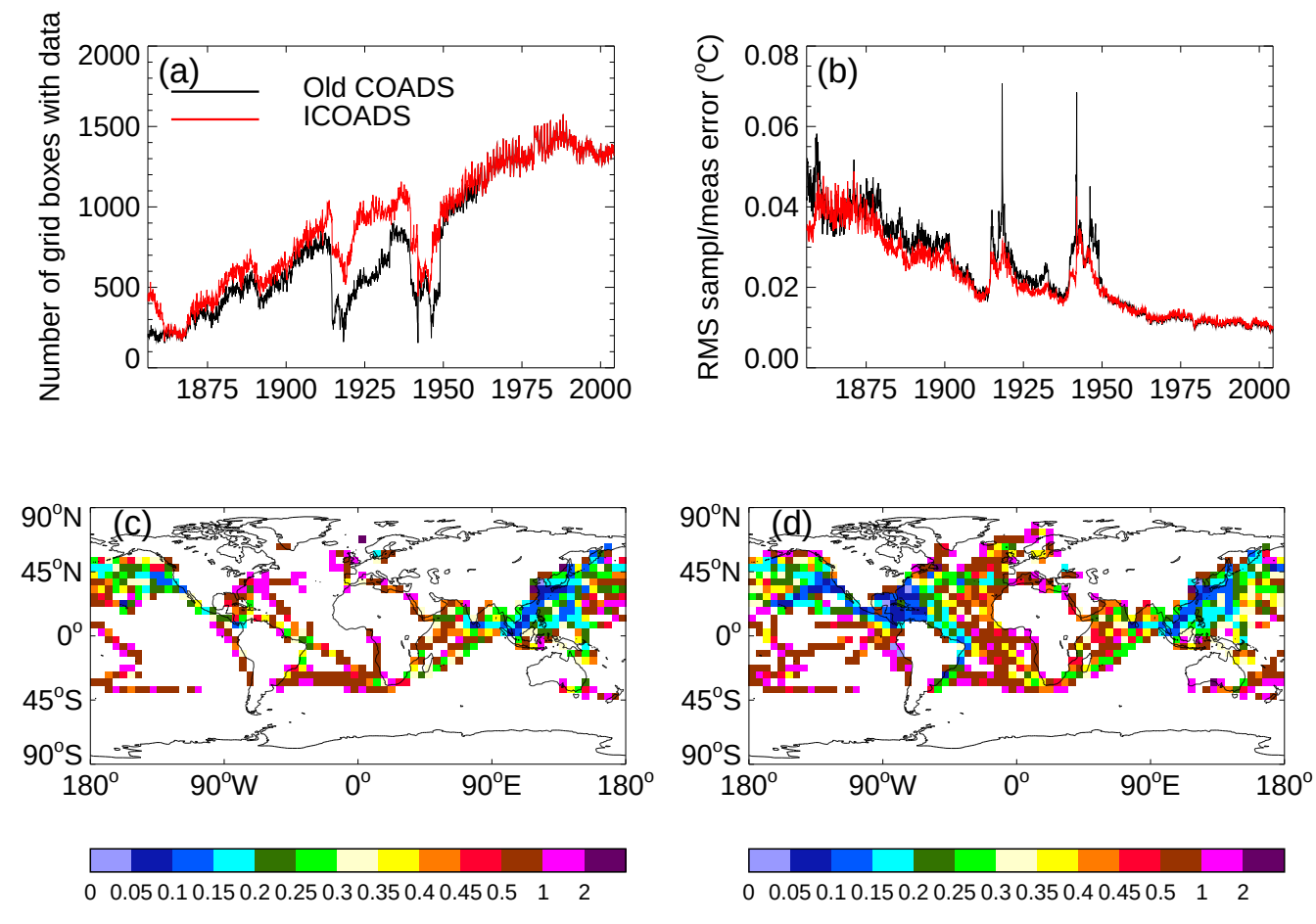

(d)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.25 0.5 0.75 1 1.25 1.5 1.75 2

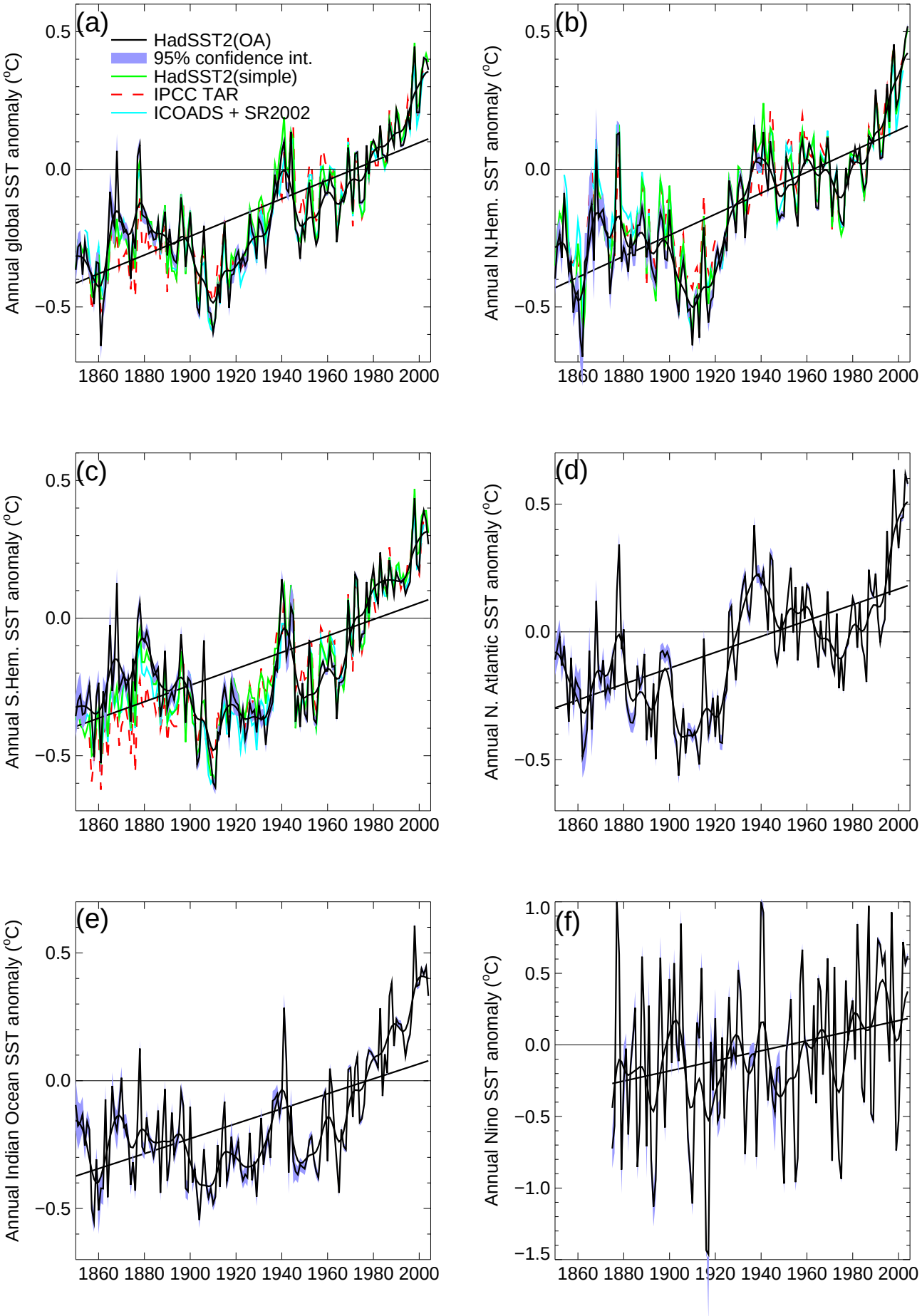

(e)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2

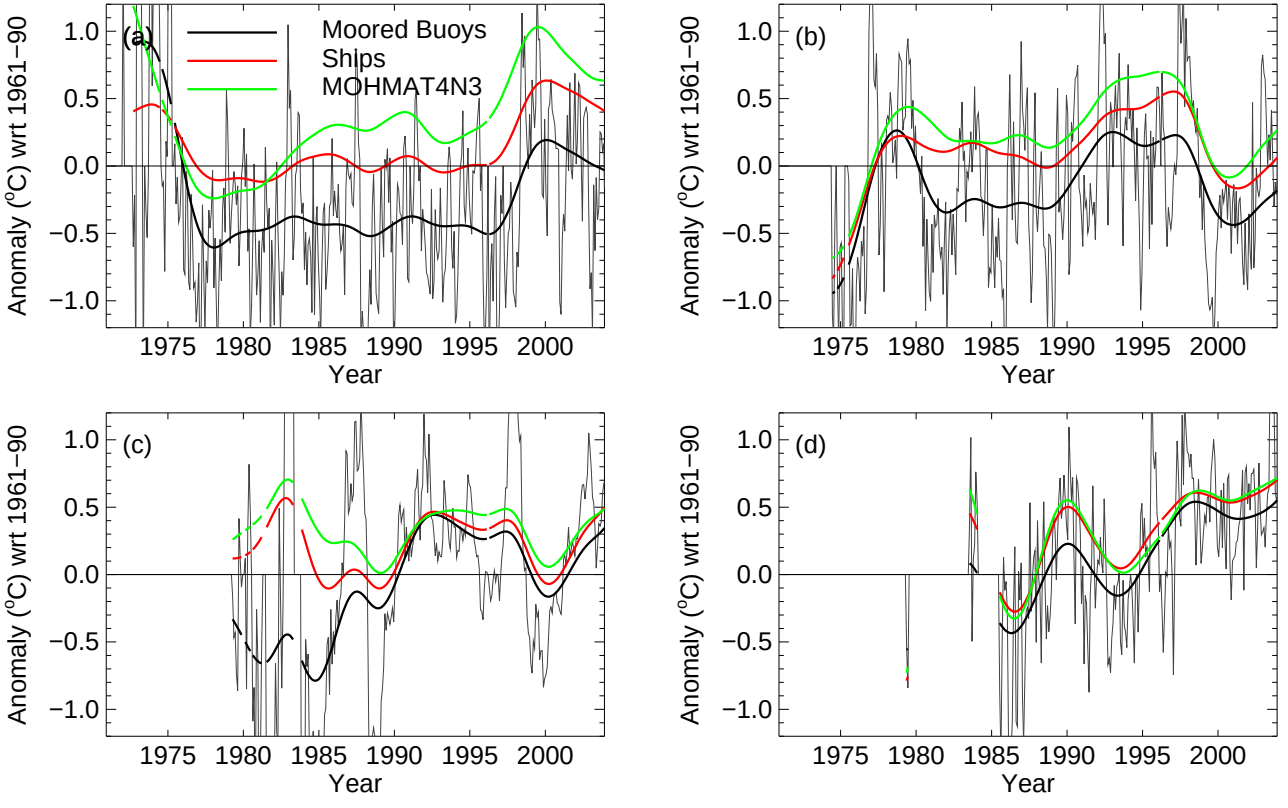

(f)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2


(g)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2


(h)
90°N
45°N
0°
45°S
90°S
180°
90°W
0°
90°E
180°
0 0.25 0.5 0.75 1 1.25 1.5 1.75 2


(a)
Old COADS
ICOADS
Number of grid boxes with data
2000
1500
1000
500
0
1875 1900 1925 1950 1975 2000
(b)
RMS sampl/meas error (°C)
0.08
0.06
0.04
0.02
0.00
1875 1900 1925 1950 1975 2000
(c)
(d)
90°N
45°N
0°
45°S
90°S
180° 90°W 0° 90°E 180°
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 1 2


(a)
HadSST2(OA)
95% confidence int.
HadSST2(simple)
IPCC TAR
ICOADS + SR2002
Annual global SST anomaly (°C)
(b)
Annual N.Hem. SST anomaly (°C)
(c)
Annual S.Hem. SST anomaly (°C)
(d)
Annual N. Atlantic SST anomaly (°C)
(e)
Annual Indian Ocean SST anomaly (°C)
(f)
Annual Nino SST anomaly (°C)
0.5
0.0
−0.5
1.0
−1.0
−1.5
1860 1880 1900 1920 1940 1960 1980 2000


(a)
Moored Buoys
Ships
MOHMAT4N3
(b)
(c)
(d)
Anomaly (°C) wrt 1961–90
Year
1.0
0.5
0.0
−0.5
−1.0
1975
1980
1985
1990
1995
2000