

Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set

Elizabeth C. Kent,¹ Nick A. Rayner,² David I. Berry,¹ Michael Saunby,²
Bengamin I. Moat,¹ John J. Kennedy,² and David E. Parker²

Received 31 July 2012; revised 21 December 2012; accepted 27 December 2012; published 14 February 2013.

[1] An updated version of the Met Office Hadley Centre's monthly night marine air temperature data set is presented. It is available on a 5° latitude-longitude grid from 1880 as anomalies relative to 1961–1990 calendar-monthly climatological average night marine air temperature (NMAT). Adjustments are made for changes in observation height; these depend on estimates of the stability of the near surface atmospheric boundary layer. In previous versions of the data set, ad hoc adjustments were also made for three periods and regions where poor observational practice was prevalent. These adjustments are re-examined. Estimates of uncertainty are calculated for every grid box and result from measurement errors, uncertainty in adjustments applied to the observations, uncertainty in the measurement height, and under-sampling. The new data set is a clear improvement over previous versions in terms of coverage because of the recent digitization of historical observations from ships' logbooks. However, the periods prior to about 1890 and around World War II remain particularly uncertain, and sampling is still sparse in some regions in other periods. A further improvement is the availability of uncertainty estimates for every grid box and every month. Previous versions required adjustments that were dependent on contemporary measurements of sea surface temperature (SST); to avoid these, the new data set starts in 1880 rather than 1856. Overall agreement with variations of SST is better for the updated data set than for previous versions, supporting existing estimates of global warming and increasing confidence in the global record of temperature variability and change.

Citation: Kent, E. C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. J. Kennedy, and D. E. Parker (2013), Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set, *J. Geophys. Res. Atmos.*, 118, 1281–1298, doi:10.1002/jgrd.50152.

1. Introduction

[2] Air temperature observations have been made and recorded by ships' officers for centuries and are collated in data sets such as the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) [Woodruff *et al.*, 1998; 2011; Worley *et al.*, 2005] and in national archives including the Met Office Marine Data Bank [Parker *et al.*, 1995]. These observations have been used to generate long-term gridded air temperature data sets including the Global Ocean Surface Temperature Atlas (GOSTA) [Bottomley *et al.*, 1990], the Met Office Historical Marine Air Temperature data set version 4 (MOHMA4) [Parker *et al.*, 1997], the Hadley Centre Marine Air Temperature data set version 1 (HadMAT1) [Rayner *et al.*, 2005], and the Centennial in situ Observation-Based Estimates (COBE) [Ishii *et al.*, 2005]. The National

Oceanography Centre Surface Flux and Meteorological Data set v2.0 (NOCv2.0) [Berry and Kent, 2009] also contains a 40 year record of air temperature. Such air temperature data sets are an important part of the climate record, complementing long term data sets of sea surface temperature (SST) [e.g., Kaplan *et al.*, 1998; Rayner *et al.*, 2005, 2006; Smith *et al.*, 2008; Kennedy *et al.*, 2011a, 2011b], land air temperature [e.g., Menne *et al.*, 2009; Jones *et al.*, 2012], upper air temperature [Thorne *et al.*, 2003; Haimberger *et al.*, 2012], and global surface temperature [e.g., Hansen *et al.*, 2010; Brohan *et al.*, 2006; Smith *et al.*, 2008; Morice *et al.*, 2012]. Although night marine air temperature is less well sampled than SST the independence of the likely biases helps to build confidence in the global record of temperature variability and change. This paper describes an improved NMAT data set, HadNMAT2.

[3] In order to create a consistent record, the air temperature observations must be adjusted from the height of observation to a common reference height. Information on air temperature measurement height is therefore required to correct for possible spurious trends in the reported air temperature estimates due to the changing height of measurement. The height adjustment required depends directly on

¹National Oceanography Centre, Southampton, UK.

²Met Office Hadley Centre, Exeter, UK.

Corresponding author: E. C. Kent, National Oceanography Centre, Southampton, UK. (eck@noc.ac.uk)

©2013. American Geophysical Union. All Rights Reserved.
2169-897X/13/10.1002/jgrd.50152

the vertical temperature gradient in the atmosphere and indirectly on the vertical gradients of humidity and wind speed. These are estimated from boundary layer similarity theory, as represented, for example in *Smith* [1980, 1988]. The adjustment requires the specification of wind speed, air temperature (both at known heights), and SST. Atmospheric humidity has a smaller effect on the adjustment but should be used if available.

[4] Marine air temperature observations also contain biases due to heating of the ships' environment. In all but the best-exposed locations, daytime air temperatures are anomalously warm [Berry *et al.*, 2004; Berry and Kent, 2005]. Warm biases are caused either by the increased temperature of the surrounding ship environment warmed by the sun or by direct heating for those sensors not housed in a screen. These biases are accounted for either by adjustment [e.g., Berry *et al.*, 2004; Berry and Kent, 2009] or by excluding affected observations and constructing gridded products from nighttime data only [Bottomley *et al.*, 1990; Rayner *et al.*, 2005].

[5] The biases due to changing measurement height and solar heating are well understood. However, there are periods where the nighttime air temperature measurements exhibit other biases that are less easily explained [Bottomley *et al.*, 1990]. These include anomalously warm temperatures in the early part of the record in the Atlantic Ocean (up to 1885), in the period 1876 to 1893 in the Mediterranean Sea and North Indian Ocean, and globally during the Second World War (WW2). All are thought to result from non-standard, error-prone, observing practice.

[6] This paper will describe the data sources (section 2) and the adjustment methodologies for height (section 3) and for non-standard observing practice (section 4). The data set construction is described in section 5 along with the calculation of grid box uncertainty estimates. Section 6 presents comparisons of HadNMAT2 with HadMAT1 and with SST and land surface air temperature data sets. Section 7 is a summary and outlines outstanding issues requiring further research.

2. Data Sources

2.1. Air Temperature

[7] The observations of air temperature used to construct HadNMAT2 are from the ICOADS Release 2.5 [Woodruff *et al.*, 2011]. ICOADS contains observations of ocean near-surface conditions from a wide variety of ships, buoys, and other platforms. The earliest observations available are from 1662 and data from 2008 onward are presently available as real-time updates. Data are extremely sparse prior to the mid-1800s, and prior to 1855 reports do not routinely contain information on the time of observation.

[8] As its name implies, ICOADS contains a wide variety of in situ surface marine observations. Various source identifiers are available so that users can include or exclude different types of data to suit their particular application. HadNMAT2 is based on air temperature observations from all sources identified by ICOADS as being ship data based on their "platform type" code. Specifically this means all data from platform type codes 0–5, 10–12, and 17 (navy, merchant, ocean station vessels, light ship, unspecified type ship, and oceanographic station data). Data from the U.S.

Shipboard Environmental data Acquisition System (SEAS) have been excluded due to an error with data format conversion but are expected to be usable in future releases of ICOADS (S. Woodruff, personal communication). Data sources excluded (using the platform type code) were moored buoys, the small number of air temperature observations from surface drifters, coastal and island data, ice station data, and rig and platform data. Future investigation may allow the inclusion of some of these observation types: moored buoy and island station observations may be particularly valuable.

[9] To avoid biases due to spurious daytime heating, only air temperature observations taken between an hour after sunset and an hour after sunrise are used, as in HadMAT1. This minimizes the effects of solar heating more effectively than using actual hours of darkness [Parker *et al.*, 1995].

2.2. Atmospheric Stability

[10] Estimates of the atmospheric stability are required to compensate the air temperature observations for observing height. For the modern period, it is possible to make estimates of the local climatological stability of the near-surface atmosphere using observations from ICOADS. These estimates of atmospheric stability are generated using data from the NOC v2.0 flux data set [Berry and Kent, 2009, 2011] for the period 1970–2006. The individual ICOADS Release 2.4 observations [Worley *et al.*, 2005] used to construct the NOC v2.0 fields have been adjusted for known biases and, with the exception of pressure and SST, to the standard marine reference height of 10 m [Berry and Kent, 2009, 2011]. These observations are gridded using optimal interpolation (OI) to produce estimates of air temperature, SST, wind speed, humidity, and pressure on a 1° latitude-longitude daily grid. Daily 1° estimates are averaged over 37 years and to a 5° grid to produce climatological fields for each month. These fields are used in the height adjustment of the air temperature data described in section 3.1.

[11] To enable the calculation of uncertainty in the stability estimates, joint anomaly distributions of wind speed and air-sea temperature difference were formed. These two variables are the main determinants of stability, so humidity and pressure were held at their climatological monthly estimates. In each 5° area grid box, anomalies of 1° area daily wind speed and air-sea temperature difference distributions were constructed relative to the 5° 37 year monthly mean climatology. Wind speed anomalies were stored in 0.5 m s^{-1} bins ranging from -10 m s^{-1} to $+10 \text{ m s}^{-1}$, air-sea temperature difference anomalies in 0.2°C bins ranging from -10°C to $+10^\circ\text{C}$. Figure 1 shows an example of a joint distribution of wind speed and air-sea temperature differences used to estimate the uncertainty in the adjustment of air temperature to the 10 m reference height arising from the use of climatological monthly mean estimates of the atmospheric stability.

2.3. Air Temperature Observation Heights

2.3.1. Height Estimates Used in Previous Air Temperature Data Sets

[12] Bottomley *et al.* [1990] adjusted air temperatures to a reference height of 15 m, the assumed average over the period 1951–1980. Observations made before and during 1890 were assigned an observation height of 6 m. Later observations were assumed to have been made at heights

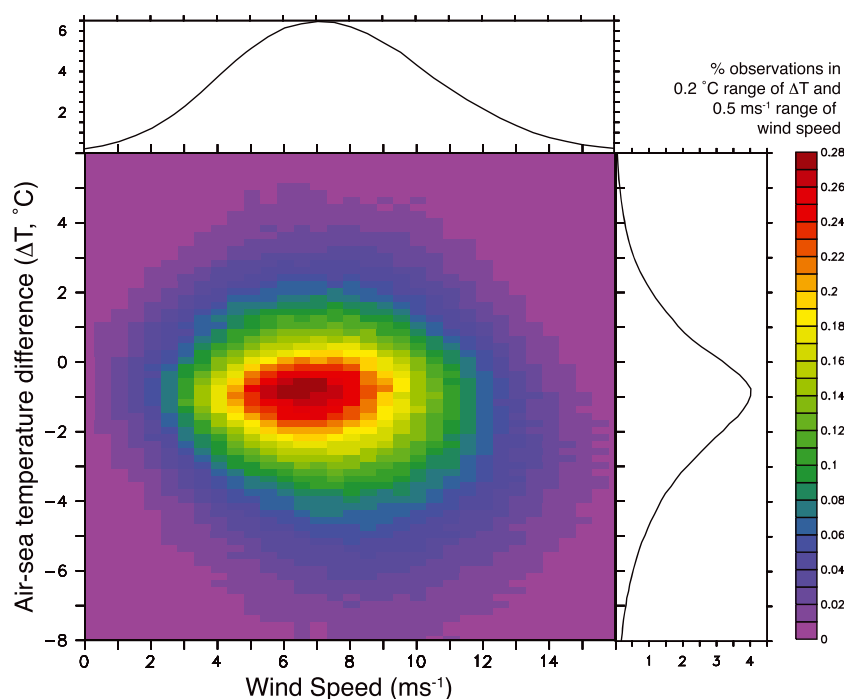


Figure 1. Joint distribution of air-sea temperature difference and wind speed for January over a 30° region centered on 45°N , 45°W . Shading represents the percentage of data falling within each 0.2°C and 0.5 m s^{-1} joint range of air-sea temperature difference and wind speed respectively. Line graphs represent the percentage of data falling within in each 0.5 m s^{-1} range of wind speed (top), and within each 0.2°C range of air-sea temperature difference (right). The anomaly joint distribution has been combined with the grid box mean wind speed and air-sea temperature difference to produce this joint distribution.

linearly increasing up to 15 m by 1931. Prior to 1891, this led to a cooling of the air temperature estimates by 0.15°C . Between 1891 and 1930, the adjustment linearly decreased to an adjustment of 0°C by 1930 after which no further adjustments were made. The heights used by *Bottomley et al.* [1990] were deduced using the history of deck elevation (and hence measurement height) from records of barometer-cistern elevations recorded in the logbooks of UK ships. The increase in measurement height was attributed to the transition from sail to steam and the accompanying increase in the size of ships. They concluded that the details of the transition would not seriously affect the relative changes in air temperature. Table 1 summarizes the information on observation heights used and Table 2 the approach taken to height adjustment.

[13] *Rayner et al.* [2005] used the same height estimates as *Bottomley et al.* [1990] for the period 1850–1939 but adjusted to a different reference height. The air temperatures were adjusted to be consistent with observations made at the average observing height for each grid box over the period 1961–1990 (cf. the period 1951–1980 used by *Bottomley et al.* [1990]).

2.3.2. Review of Information on Observation Height, 1850 to End of WW2

[14] There is evidence from the literature that the measurement height may have begun to increase prior to 1890, the date used by *Bottomley et al.* [1990] and *Rayner et al.* [2005]. The first iron merchant steam ships resembled wooden sailing traders [Marshall, 1989] and were likely to have had similar measurement heights. Images in *Lindsay*

[1876] suggest that the open deck of ships was typically two decks above the waterline, consistent with a measurement height of about 6 m. From around 1870, raised structures on the open deck were built to protect funnels and the engine room in bad weather. By 1890, steam ships were very different from the first iron steamers with “three island” cargo ships with raised forecastle, bridge deck amidships, and a raised structure aft on the poop deck [Marshall, 1989]. Estimation of likely measurement heights from ship images is consistent with those used by *Rayner et al.* [2005] in this period and we therefore use a similar linear increase although starting in 1870 rather than 1891 (Table 1).

[15] WW2 had a dramatic effect on shipping. The ships that were built to replace those lost were of standard types, the Liberty and Victory classes, and were significantly smaller than ships in general use before the war. Hog Island ships built during the First World War (WW1) were used in WW2. For this period the measurement height has been estimated as 11 m. Some naval ships were likely to be significantly higher, but the average height of naval ships is not clear so no attempt has been made to differentiate naval and non-military ships within ICOADS. Neither *Bottomley et al.* [1990] nor *Rayner et al.* [2005] allow for any discontinuity in observation height due to shipping loss during the WW2.

2.3.3. Review of Information on Observation Height, End of WW2 to Present

[16] After WW2, ships became more specialized and tended to increase in size. The height of the bridge above the deck is determined by the need to see over the bow, so the longer the ship the higher the bridge needed to be.

Table 1. Summary of Observation Heights (OH) and Uncertainties Used in Different Studies

Period	Observation Height [Bottomley <i>et al.</i> , 1990]	Observation Height [Rayner <i>et al.</i> , 2005]	Observation Height (This Study)	Uncorrelated Uncertainty in Observation Height (OH, $2\text{-}\sigma$)	Correlated Uncertainty in Observation Height ($2\text{-}\sigma$)
Before 1870	6 m	6 m	6 m	$0.6 \times \text{OH}$	2 m
1870 to 1890	6 m	6 m	linear increase	$0.6 \times \text{OH}$	2 m
1891 to 1909	linear increase	linear increase	linear increase	$0.6 \times \text{OH}$	2 m
1910 to 1930	linear increase	linear increase	linear increase	$0.6 \times \text{OH}$, linear decrease to $0.4 \times \text{OH}$ in 1930	2 m
1931 to 1937	15 m	15 m	linear increase	$0.4 \times \text{OH}$	2 m
1938	15 m	15 m	15 m	$0.4 \times \text{OH}$	2 m
1939 to 1945	15 m	15 m	11 m	$0.4 \times \text{OH}$	2 m
1946 to 1969	15 m	15 m	linear increase from 11 m	$0.4 \times \text{OH}$	2 m
1970 to 1994	15 m	Pub. 47 gridded	Pub. 47 individual plus gridded default	zero for known heights histograms or standard deviations for unknown heights	2 m
1995 to 2007	15 m (data set ends in 2003)	Pub. 47 gridded + 0.14 m yr^{-1}	as 1970–1994	as 1970–1994	2 m
Nov 2007	n/a	Pub. 47 gridded + 0.14 m yr^{-1}	as 1970–1994	as 1970–1994	2 m
Dec 2007 onwards	n/a	Pub. 47 gridded + 0.14 m yr^{-1}	2007 Pub. 47 gridded defaults	using 2007 histograms and gridded standard deviations	4 m

Container ships and gas tankers have the highest bridges and hence the highest observation heights [Kent *et al.*, 2007]. Modern container ships have reached a design limit in terms of the amount of cargo that can be carried in front of the bridge and now also carry cargo aft of the bridge.

[17] The World Meteorological Organization (WMO) has collected and published observational metadata on Voluntary Observing Ships (VOS) in the “International List of Selected, Supplementary and Auxiliary Ships,” WMO Publication No. 47 (Pub. 47) since 1955. Kent *et al.* [2007] summarized the information available in Pub. 47. Platform height was introduced as a field in 1968. In 1995, this field was dropped and the most relevant field that was introduced was the height of the barometer that in many cases would be similar to the height of temperature measurement. In 2002, the height of the temperature sensor was explicitly included as a field along with a field for “Visual wind/wave observation height.” Information on ship type was introduced as a field in 1995, although some countries listed their ships by type prior to this.

[18] Because Pub. 47 is an operational resource it contains some typographical errors and multiple entries; we used an amended version [Berry and Kent, 2006]. In this version, duplicate entries, where two different countries supplied information on the same ship, were reconciled or removed and instances of inconsistent entry of the same information were identified and rectified. The most significant example of this is where instrument heights were entered using the

wrong units, or relative to the wrong reference level. Where consistent entries for a particular ship could be identified across different editions, Berry and Kent [2006] interpolated metadata to fill gaps in information, or extrapolated to provide prior information on a newly introduced field. This enabled a small number of observation heights to be identified prior to the introduction of this information in 1968. This processing was performed for the period up to the first quarter of 2006; further details can be found in Berry and Kent [2006]. These metadata have been merged with ICOADS following Kent *et al.* [2007]. From the second quarter of 2006 onward, the metadata as provided by the WMO have been used and merged with ICOADS. From December 2007, it was not possible to associate measurement heights with ICOADS observations due to the masking of ship call signs after this date.

[19] Measurement heights from the merged ICOADS/Pub. 47 data set were averaged in 5° monthly grid boxes and the mean height, number of heights, and their standard deviation calculated. These data were then smoothed using triangular weighting with a width of 61 months (i.e., a weight of 0 for data 31 months before and after the observation time rising linearly to a weight of 1 at the observation time); missing data were masked in the filtering. From December 2007 onward, the heights used were the average of the filtered values for 2007. Values were not increased in line with previous trends as the unmatched metadata in Pub. 47 indicated that no noticeable increase in measurement height

Table 2. Approach Taken to Air Temperature Height Adjustment in Different Studies

	Bottomley <i>et al.</i> [1990]	Rayner <i>et al.</i> [2005]	This Study
Reference height	15 m	average height over 1961 to 1990 $\sim 16 \text{ m}$, spatially varying	10 m
Climatological reference period	1951 to 1980	1961 to 1990	1961 to 1990
Adjustment method	Boundary layer similarity (after Large and Pond [1982]) based on one representative atmospheric profile	Boundary layer similarity (after Fairall <i>et al.</i> [1996]) based on 30 representative atmospheric profiles	Boundary layer similarity [Smith, 1980, 1988] using average monthly atmospheric conditions for each 5° grid-box [Berry and Kent, 2009]
Uncertainty in height adjustment	None calculated	None calculated	Including uncertainty in measurements heights and atmospheric stability

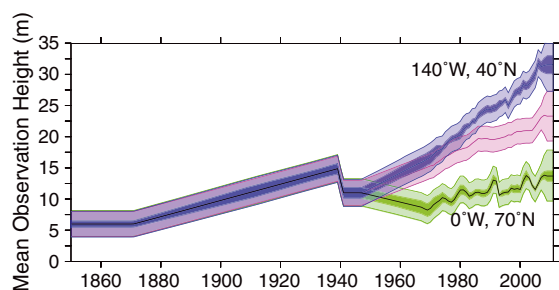


Figure 2. Annual mean observation heights (m, solid lines), and 2- σ uncertainties, 1850–2010. The correlated uncertainty (darker shading) and uncorrelated uncertainty are combined in quadrature to give the total uncertainty (lighter shading). Uncorrelated and correlated uncertainties are assumed to be uncorrelated with each other. Presented are the global mean (pink: uncorrelated component of the uncertainty is small and not visible) and uncertainties for two 5° boxes, one in the Pacific [blue: 140°W, 40°N] where observation heights are higher than the global mean and one in the Norwegian Sea [green: 0°W, 70°N] where observation heights are low. In the earlier period when heights are globally invariant the information visible is for the Pacific grid box. The global uncertainty in observation height will be close to the estimated correlated uncertainty of 2 m for this period.

had occurred since the loss of call signs in ICOADS (December 2007) and the latest available metadata (second quarter of 2008). This should be revisited when call signs and metadata become available. Figure 2 shows annual mean measurement height averaged over the globe and for two example 5° grid boxes, all with uncertainty estimates. Figure 3 shows the mean observation height and its uncertainty for four sample years.

2.3.4. Uncertainty in Individual Observation Heights

[20] In addition to assigning a measurement height to each observation, an estimate of the uncertainty in that measurement height is also required. The uncertainty in the observation height in the early period is difficult to estimate but we have done so assuming two different components. The first accounts for the variation in measurement height between different ships and hence around a mean value. We have assumed that this is uncorrelated between observations from different ships. The second represents the uncertainty in our knowledge of the mean value itself and this has been assumed to be correlated across all observations.

[21] For data prior to 1910, the standard deviation of the heights of barometer cisterns used by *Bottomley et al.* [1990] suggested that an uncorrelated contribution to the 2- σ overall uncertainty of approximately $0.6 \times$ observation height (OH) was appropriate. The 2- σ standard deviation of height values in Pub. 47 in 1970 was 6 m and the mean observation height 15 m, giving an uncorrelated uncertainty estimate of $0.4 \times$ OH. The standard deviation of the barometer cistern heights in 1930 also indicates an uncorrelated uncertainty of $0.4 \times$ OH. We therefore linearly reduce the fractional uncorrelated uncertainty from 0.6 to 0.4 over the period 1910–1930 and use an uncorrelated uncertainty estimate of $0.4 \times$ OH between 1930 and 1970. These uncertainties are summarized in Table 1.

[22] After 1970, the uncorrelated uncertainty is estimated from observation heights in Pub. 47. If an observation can be matched to a height in Pub. 47 the uncorrelated uncertainty in the observation height is set to zero. If there is no observation height information available and the observation is in a 5° grid box where there are more than 100 identified heights over a centered 5 year period, then the standard deviation of 5000 values randomly selected (with replacement) from the distribution of known heights in the target grid box and centered 5 year period is calculated. This standard deviation is smoothed in time as described in section 2.3.3 and used as the uncorrelated uncertainty in the observation height. If there are fewer identified heights then the uncorrelated uncertainty is set to the 5 year smoothed standard deviation of the available heights. In data-sparse regions the uncorrelated uncertainty is the 5° latitude zonal standard deviation, or if there are no heights in the zonal band, the global value.

[23] In addition to the estimates of uncertainty due to the variability of measurement height amongst the different ships we also need to allow for the overall uncertainty in the mean measurement height. This cannot be done objectively and we therefore estimate this correlated contribution to the total uncertainty. Correlated 2- σ uncertainty in the measurement height has been estimated to be 2 m, except for the period after 2007 when no call sign information exists and the estimate of correlated uncertainty in the measurement height is increased to 4 m (see Figure 2).

3. Air Temperature Height Adjustment and Its Uncertainty

3.1. Mean Air Temperature Height Adjustment

[24] Individual air temperature observations from ICOADS have been adjusted to a standard reference height of 10 m using boundary layer similarity theory [e.g., *Dyer*, 1974]. The parameterization of *Smith* [1980] has been used for the underlying drag coefficient and *Smith* [1988] for the heat and moisture transfer coefficients (Table 2). The height adjustment requires knowledge of observation height, wind speed, the air and sea temperatures, and atmospheric humidity. Not all of this information is available for all ICOADS air temperature observations. The availability of and uncertainty in the required information varies over time, and also regionally. We here attempt to take a consistent approach over the entire period and use climatological monthly and 5° area average estimates of atmospheric variables to make the height adjustment (section 2.2). Where air temperature observation heights can be identified from Pub. 47 (almost exclusively after 1970), these are used, otherwise default values are assumed as described in sections 2.3.2 and 2.3.3 and Table 1. Prior to 1946, globally invariant estimates of air temperature observation height are used as described in section 2. After 1970, 5° gridded observation heights have been used based on the ICOADS/Pub 47 merge and smoothed in time as described in section 2.3.3. Between 1946 and 1970, the measurement height has been assumed to linearly increase in each 5° grid box from 11 m in 1946 to the 1970 5° value. The adjustment has been calculated following *Smith* [1980, 1988] using climatological mean values of air-sea temperature difference, humidity and wind speed (all at a reference height of 10 m; section 2.2) and

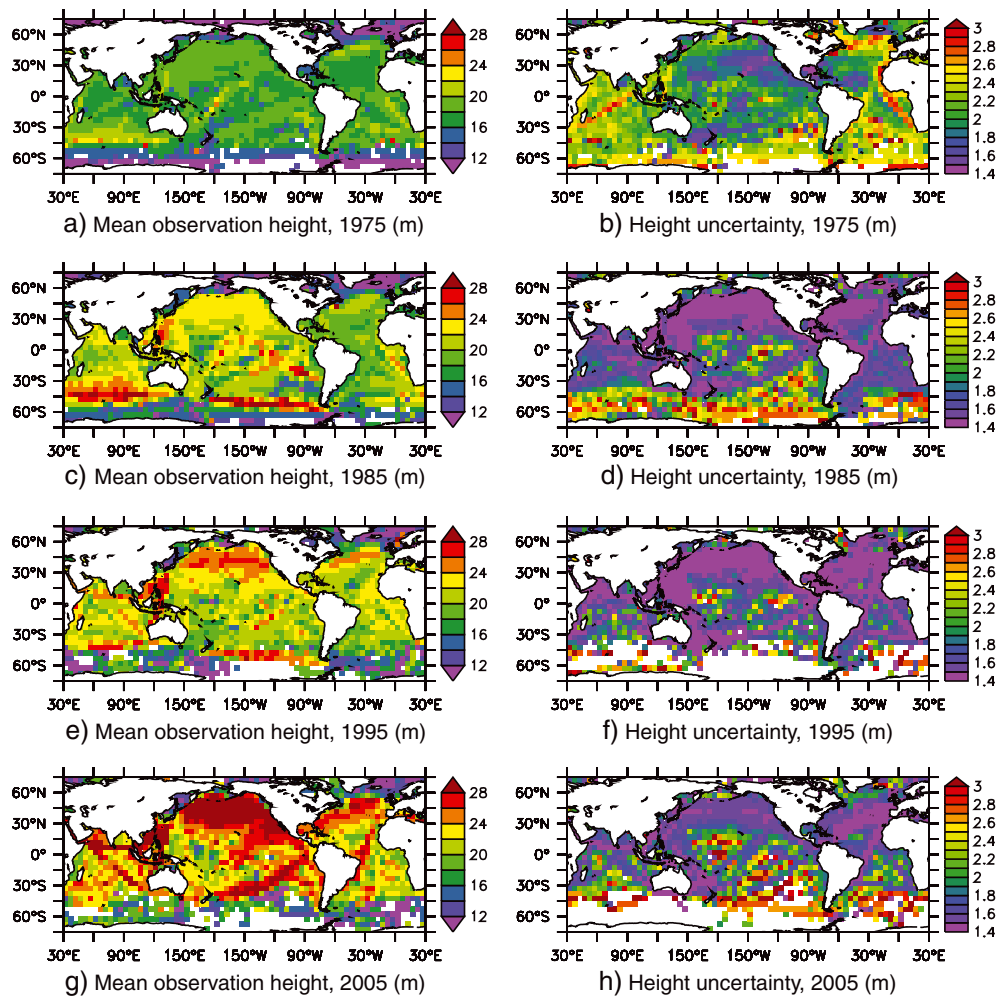


Figure 3. Air temperature observation heights and their estimated 2- σ uncertainty for four sample years (m); (a) mean height for 1975; (b) uncertainty in annual mean height for 1975; (c) mean height for 1985; (d) uncertainty in annual mean height for 1985; (e) mean height for 1995; (f) uncertainty in annual mean height for 1995; (g) mean height for 2005; and (h) uncertainty in annual mean height for 2005.

assuming that the humidity just above the ocean surface is 98% of the saturation specific humidity at the SST.

[25] Figure 4 compares the global and annual mean height adjustments from this study with those used by *Bottomley et al.* [1990] and *Rayner et al.* [2005]. Globally the new adjustment ranges from a reduction of slightly over 0.06°C in the 1850s to an average increase of nearly 0.17°C by 2009. This range of 0.23°C is larger than the 0.15°C assumed by *Bottomley et al.* [1990] because their adjustment methodology assumed no systematic changes in measurement height after 1930, which we know is not appropriate. The range of 0.23°C in the global adjustment from the present study is approximately 0.05°C smaller than that from *Rayner et al.* [2005]. Tests using stability estimates calculated using the same fixed profiles as in *Rayner et al.* [2005] accounted for most of this difference (0.036°C). The remaining unexplained difference is smaller than the estimate of correlated uncertainty that has been attributed to the height adjustment procedure. It should also be noted that the height adjustments of *Rayner et al.* [2005] will vary regionally only due to varying measurement heights and not due to regional variations in atmospheric stability.

[26] The adjustment is offset relative to that used by *Rayner et al.* [2005] as here the adjustment has been made to the more commonly used standard reference level of 10 m. *Rayner et al.* [2005] applied their adjustment to homogenize the record at each grid box to the average measurement height in that grid box in their climatological period of 1961 to 1990. For comparison, Figure 5a shows the equivalent spatially varying reference height calculated from the heights used in this study, which should be broadly similar to those used by *Rayner et al.* [2005]; the mean value over this period is 16.3 m.

[27] In order to retain consistency with HadMAT1, anomalies are presented relative to a 1961–1990 climatology based on ICOADSv2.0 data, adjusted as for HadMAT1 [*Rayner et al.*, 2005].

[28] Figure 6 shows the adjustment required to the reported air temperature observations averaged over four example decades. In the two earliest periods shown, the spatial distribution of the adjustment depends on stability only as the heights are globally invariant. In the first period (1880–1889), the adjustment acts to decrease the reported air temperature as heights are below the reference of 10 m and the sea surface

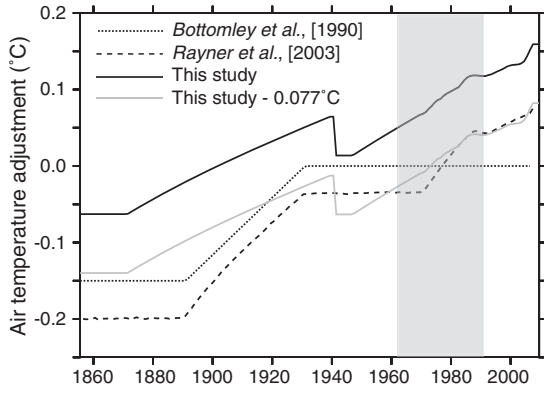


Figure 4. Global annual mean air temperature adjustments for differences from the standard reference height of 10 m, as used in this study, by *Bottomley et al.* [1990] and by *Rayner et al.* [2005]. The grey shading indicates the climatological period over which the adjustments of *Rayner et al.* [2005] average to zero. *Bottomley et al.* [1990] used a climatology period of 1951–1980. Note that the adjustments used in this study (black solid line) do not average to zero over the climatological period and the mean value of the height adjustment over the period 1961 to 1990 of 0.077°C has been subtracted to aid comparison with the other time series (grey solid line).

is typically warmer than the air above it. In the second period (1920–1929), the assumed measurement height is greater than 10 m and the adjustment acts to increase the reported air temperature. In the two later periods shown (1960–1969 and 2000–2009), the adjustments now vary with both stability and observation height giving a more complex regional picture.

3.2. Uncertainty in Air Temperature Height Adjustment

[29] The uncorrelated or random contribution to the grid box uncertainty in the air temperature height adjustment is assumed to come from two sources—that due to the uncertainty in the observation height and that due to the uncertainty in the atmospheric stability estimate.

[30] The contribution of random uncertainty in the observing height was estimated in the following way. Whenever the measurement height was known, which was sometimes the case from 1968 onward, the uncertainty was set to

zero. Otherwise, 5000 realizations of observing height were generated as described in section 2.3.4. These realizations of observing height were then used to calculate realizations of air temperature adjustments, given climatological stability as described in section 2.2, following *Smith* [1980, 1988]. These realizations defined the contribution of random uncertainty in observing height to the random uncertainty in air temperature height adjustment of a single observation.

[31] The contribution of uncertainty in stability was estimated by sampling 5000 times from the joint wind speed/air-sea temperature distributions described in section 2.2 given climatological mean humidity and using observed heights where available, otherwise the default heights described in sections 2.3.2 and 2.3.3. Each time, an air temperature height adjustment was calculated following *Smith* [1980, 1988]. These realizations defined the contribution of uncertainty in stability to the random uncertainty in air temperature height adjustment of a single observation.

[32] The overall grid box random uncertainty in air temperature height adjustment is based on these random uncertainties for single observations. The contribution of random uncertainty in observing height will reduce if measurement heights are known, or if multiple ships are known to provide measurements in the particular grid box and month; but not if observations are all from unidentified ships, because it is possible that there may be only one ship, preventing reduction of scatter by averaging within the grid box (equation (1)). The adjustment uncertainty due to uncertainty in the atmospheric stability will reduce if there are many observations within the grid box and month (equation (2)). These two contributions to the grid box uncertainty are assumed to be uncorrelated with each other (equation (3)).

$$\sigma_{g,h,rand} = \sqrt{\frac{\frac{n_{ht} \times 0}{n_{ship_{ht}}} + \frac{n_{id} \times \sigma_{h,rand}^2}{n_{ship_{id}}} + n_{unk} \times \sigma_{h,rand}^2}{n_{obs}}} \quad (1)$$

$$\sigma_{g,s,rand} = \sigma_{s,rand} / \sqrt{n_{obs}} \quad (2)$$

$$\sigma_{g,rand} = \sqrt{\sigma_{g,h,rand}^2 + \sigma_{g,s,rand}^2} \quad (3)$$

where $\sigma_{g,h,rand}$ is the grid box random uncertainty due to uncertainty in the observation height, $\sigma_{g,s,rand}$ is the grid box random uncertainty due to uncertainty in the atmospheric

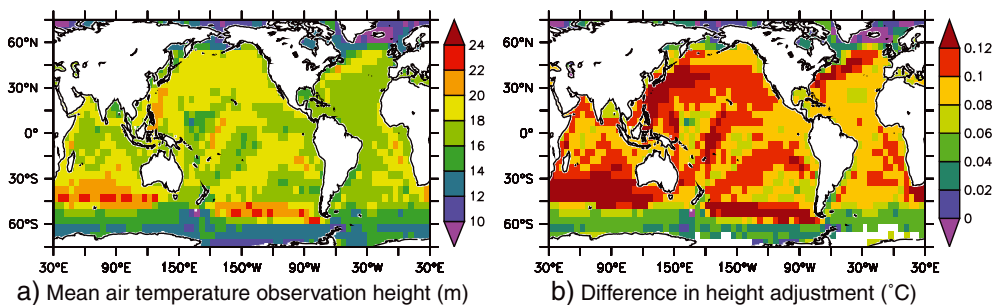


Figure 5. (a) Estimate of reference height used in *Rayner et al.* [2005] calculated as the average height in the period 1961 to 1990. (b) Increment to adjustments applied using reference height in Figure 5a when changing to a reference height of 10 m.

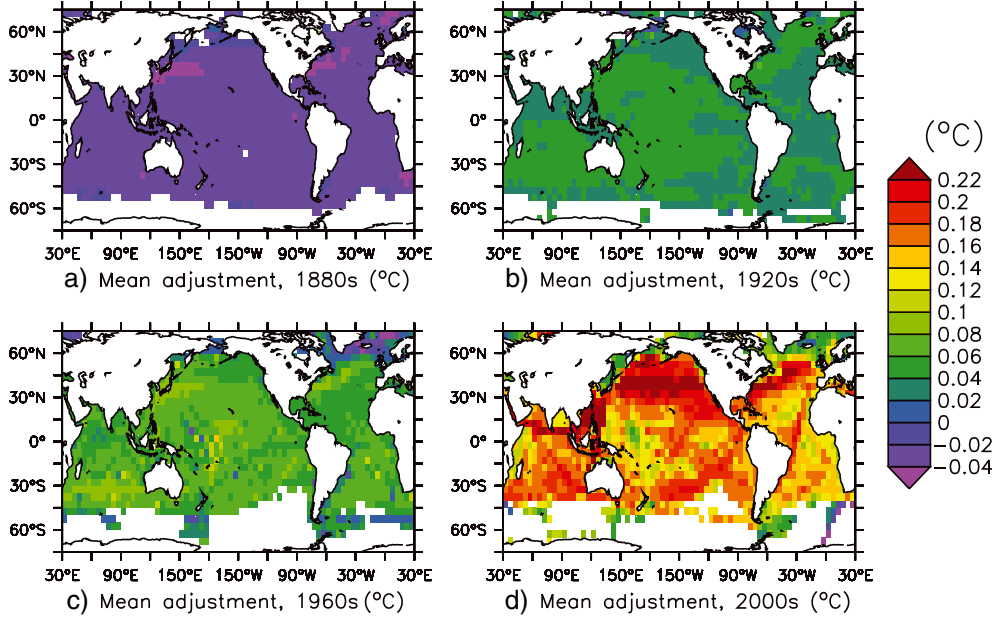


Figure 6. Average adjustment of air temperature (°C) when changing to a reference height of 10 m in four sample decades. (a) 1880–1890, plotted when there is at least 1 month present in grid box; (b) as Figure 6a but for 1920–1929; (c) as Figure 6a but for 1960–1969 and plotted when there are at least 24 months present; and (d) as Figure 6c but for 2000–2009.

stability, $\sigma_{g,rand}$ is the total random grid box uncertainty, $\sigma_{h,rand}$ is the random uncertainty in a single observation due to uncertainty in the measurement height and $\sigma_{s,rand}$ is the random uncertainty in a single observation due to uncertainty in the stability. n_{ht} is the number of observations in the grid box with known observation height, $n_{ship_{ht}}$ is the number of different ships contributing to n_{ht} , n_{id} is the number of observations in the grid box from identifiable ships without information on observation height, $n_{ship_{id}}$ is the number of different ships contributing to n_{id} , n_{obs} is the total number of observations, and n_{unk} is the number of observations that cannot be associated with an individual ship (e.g., reports with generic or missing identifiers).

[33] The uncorrelated uncertainties will reduce when spatial averages are made as $n_g^{-1/2}$, where n_g is the number of grid boxes. The uncorrelated component of the uncertainty is therefore very small in the global mean, but may be large for a small area average or individual grid box (e.g., Figure 2).

[34] Figure 7 shows maps of the annual average uncorrelated uncertainty in air temperature height adjustment due to the combined effects of stability and measurement height in four sample years. The uncertainty in the height adjustment due to uncertainty in the stability scales with the difference in height from the 10 m reference height and is relatively larger in regions where the variability in

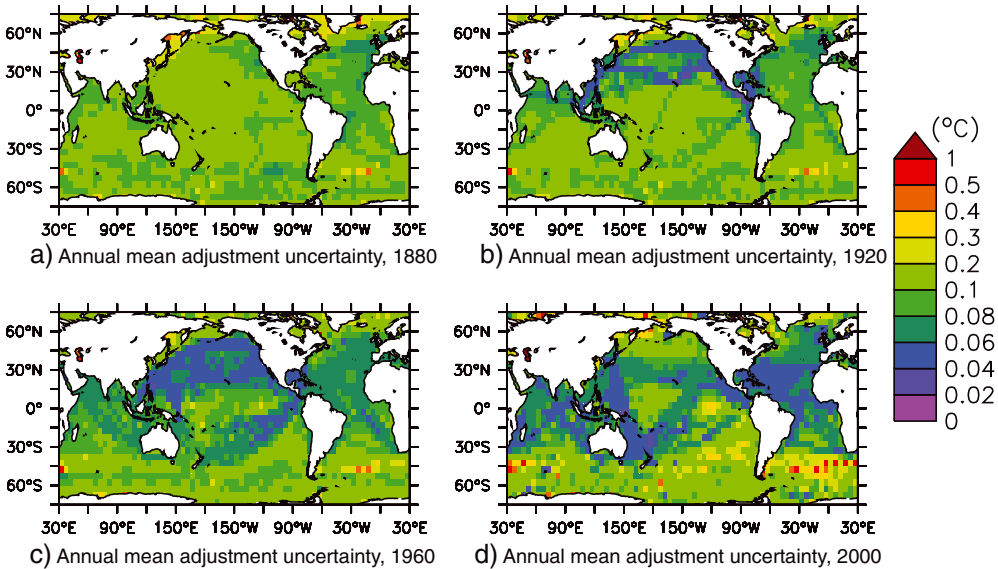


Figure 7. Annual mean 2- σ uncorrelated uncertainty in height adjustment (°C) in four sample years.

Table 3. Estimates of Correlated Contribution to Total 2- σ Uncertainty Used for Calculation of Correlated Uncertainty in Air Temperature Height Adjustment^a

Variable	Bias Uncertainty
Observation height	1855–2007: 2.0 m 2008–2011: 4.0 m
Air temperature	0.4 °C
SST	$\max(0.3, 0.2 * SST - Air\ Temp.)$ °C
Wind speed	0.4 ms ⁻¹

^aNote that 1- σ uncertainties are quoted in *Berry and Kent* [2009].

the stability is larger (e.g., regions with western boundary currents, not shown). The contribution of uncertainty in the measurement height scales with the uncertainty in the height, but is greater for measurement heights low in the boundary layer where vertical gradients of temperature are larger.

[35] Estimates of the correlated contribution to the total uncertainty in the air temperature adjustment are correlated among all grid boxes and may also be correlated over time. They are based on a correlated contribution to measurement height uncertainty of 2 m (2- σ) prior to 2008 and 4 m thereafter (see section 2.3.4 and Table 1). For the meteorological parameters used to calculate the atmospheric stability, the estimates of the correlated uncertainty of *Berry and Kent* [2009] are used (Table 3).

4. Adjustments for Non-standard Observing Practice

4.1. Overview

[36] The importance of using well-exposed instruments to measure air temperature is now well understood, albeit sometimes difficult to implement given the practical constraints of instrument siting. Early in the record, there are particular regions and periods where air temperature observations show characteristics that suggest poor instrument exposure and observing practice might have been prevalent. It is of course possible that some of the air temperature records that have been digitized were never intended to be measurements of the ambient atmosphere but were measured for information on the ship [*Chenoweth*, 2000]. Adjustments are therefore made in an attempt to recover the true values of air temperature [*Bottomley et al.*, 1990; *Rayner et al.*, 2005]. These adjustments typically rely on the expected relationships among SST and daytime and nighttime MAT. The rationale for these adjustments has been previously described [e.g., *Bottomley et al.*, 1990; *Rayner et al.*, 2005], here we assess whether these adjustments are still appropriate given the large amounts of newly available data in ICOADS Release 2.5.

4.2. Adjustments to Early Air Temperature Observations

[37] *Bottomley et al.* [1990] found high air temperature anomalies in the Atlantic Ocean in the period 1856 (the start of their data set) to about 1885. Unexpectedly they found anomalies were warmer in windy conditions and they suspected that observers were making measurements in sheltered positions when conditions were bad. NMAT anomaly in the region 40°N–50°N, 50°W–20°W was seen to exhibit

a different relationship with wind speed in the period 1856–1880, as compared to that seen in 1881–1900 and subsequent periods. *Bottomley et al.* [1990] and *Rayner et al.* [2005] therefore constrained NMAT anomalies in each grid box over much of the Atlantic to have the same 1856–1885 calendar-monthly averages as adjusted SST anomalies, while allowing interannual variations of air-sea temperature difference.

[38] *Chenoweth* [2000] describes conditions of thermometer exposure in the period 1795–1843 which may have persisted into the start of the period of interest here. *Chenoweth* [2000] found examples of temperatures being measured in cabins and concluded that cabin temperatures would be anomalously warm in higher latitudes if cabins were heated, and show lower wind speed dependence than deck temperatures (as any windows were likely to be shut in bad weather). The Brussels conference in 1853 [*Mauray*, 1854] specified the need for air temperature observations to be taken in the open air and a report of the meeting [*Board of Trade*, 1857] cites that “Great care should be taken in making observations of the temperature that the situation is in an exposed but shaded place, where the heated air of the decks of the ship or a cabin cannot influence the result, and where the sun has no effect upon the instrument.” Observing instructions from the UK and US in the period 1860 onward mention good exposure and the use of screens.

[39] Air temperature observations from ICOADS Release 2.5 show the same behavior prior to 1880 noted by *Bottomley et al.* [1990] suggesting that the data in the Atlantic would need adjustment if they were to be used in HadNMAT2, as in previous versions. Because observations in the Pacific are sparse, this implies that most observations would require adjusting in this period. We therefore choose to start HadNMAT2 in 1880 rather than 1856. Further analysis of the data prior to 1880 to try to identify whether any of the data can be made usable would be highly desirable, however this is made more difficult by missing ship or logbook identifiers in much of the data.

4.3. “Suez” Adjustment

[40] *Bottomley et al.* [1990] identified a warm bias in air temperature in the period 1876–1893 in the Mediterranean Sea and North Indian Ocean. They attributed this warm bias to the practice of piling cargo on deck rather than in the hold to avoid taxes at the Suez Canal, thus restricting the airflow around the air temperature thermometer. Tariffs were reduced in 1893. *Bottomley et al.* [1990] replaced the biased air temperature anomalies with collocated SST anomalies which appeared to be unbiased.

[41] The ICOADS Release 2.5 also shows the air temperature bias in the region around the Suez Canal, but the bias was only present in one of the ICOADS data sources, known as “Deck 193.” Deck 193 (Netherlands Marine) contains the bulk of air temperature observations prior to about 1878 and about half the observations between 1879 and 1882. Figure 8a shows the difference averaged over 1880–1892 between all air temperature observations (selected as described in section 2.1) and all observations with Deck 193 excluded. Warm biases are seen in Deck 193 in the Mediterranean and Red Sea regions and extend into the Indian Ocean and North Atlantic. *Bottomley et al.* [1990] did not detect the extension into the North Atlantic.

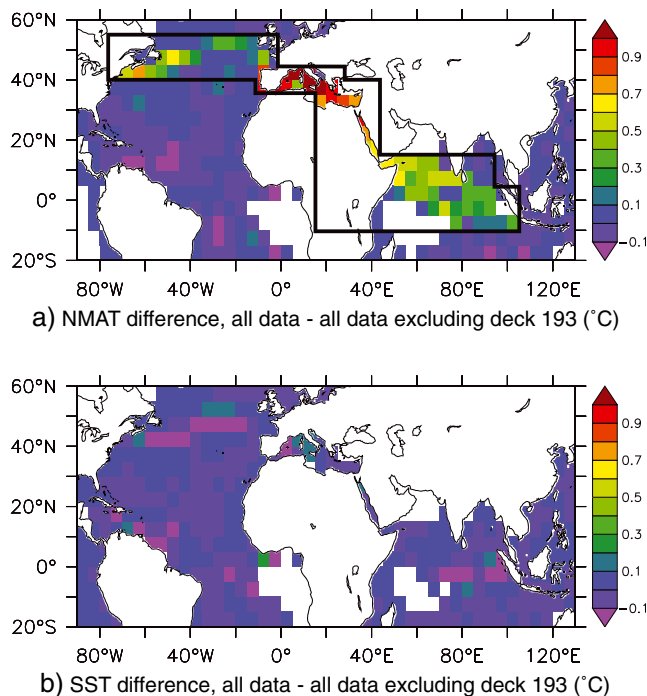


Figure 8. Mean differences resulting from the exclusion of ICOADS Deck 193 (Netherlands Marine) in (a) night marine air temperature and (b) SST over the period 1880–1892. Figure 8a indicates with a heavy line the region for which air temperature observations from this Deck are excluded from HadNMAT2 for the period 1880–1892.

It would be interesting to examine anomalies separately for ships in Deck 193 that passed through the Suez Canal and those which did not, but Deck 193 does not contain ship identifiers so tracking individual ships would require further research. Figure 8b shows that SSTs were not affected by the bias.

[42] The HadNMAT2 differs from HadMAT1 in that no adjustments are applied to account for warm biases in the early period of operation of the Suez Canal. Instead, data from Deck 193 are excluded in the region defined by Figure 8 from the start of HadNMAT2 in 1880 until 1893. This approach, combined with the later start of HadNMAT2 compared with HadMAT1, means that HadNMAT2 is not dependent on contemporary SST observations for any adjustments over its full period of record. The adjustment of air temperature observations to the standard 10 m reference height uses climatological monthly estimates of the air-sea temperature difference (described in sections 2.2 and 3). This dependence of the adjustment to 10 m reference height on climatological SST rather than time-varying SST means that the adjustment for height should not introduce time-varying biases into HadNMAT2. Any real change in the air-sea temperature difference will not be accounted for in the adjustment, but any effect should be small.

4.4. WW2 Adjustment

[43] There are warm biases in both air temperature and SST during 1939–1945 due to non-standard observing practices during wartime [e.g., Bottomley *et al.*, 1990; Thompson

et al., 2008]. Rayner *et al.* [2005] replaced nighttime (defined as hours of darkness) air temperature anomalies with daytime air temperature anomalies adjusted so that from 1939 to 1941 (1942–1945) the day-night difference equaled its local average for 1929–1938 (1946–1955).

[44] Data from ICOADS Release 2.5 still show additional warmth in their nighttime air temperature anomalies during WW2, from 1942, and the latter part of the adjustment applied by Rayner *et al.* [2005] is therefore still required and appropriate. Here, we amend it slightly and replace NMAT anomalies between 1942 and February 1946 with DMAT anomalies, adjusted according to the difference between DMAT and NMAT anomalies over the period 1947–1956. Additionally, daytime air temperature anomalies for Deck 195 (U.S. Navy Ships Logs) were anomalously warm compared with data from other Decks and are excluded. An adjustment prior to 1942 appears not to be required due to the addition of many recently digitized measurements for this period. Figure 9 shows time series of monthly unadjusted and adjusted NMAT anomalies for the period 1929–1955, along with the daytime air temperature data used in the adjustment process.

[45] Uncertainties for the WW2 adjustment can be estimated from the standard deviations of the day-night differences in the period used to calculate the differences (1947–1956).

5. The Construction of HadNMAT2 and Its Uncertainty

5.1. Quality Control and Gridding

[46] The adjustments for the effect of changing deck height were applied to each individual NMAT measurement used. Only unadjusted measurements passing quality control (QC) checks (summarized below) were included. A Winsorized mean [Dixon, 1960] of all measurements passing QC in each monthly 5° latitude by 5° longitude grid box was used to create gridded fields of NMAT anomaly (as described for SST in Rayner *et al.* [2006]). The WW2 adjustment was applied to those gridded fields between 1942 and February 1946.

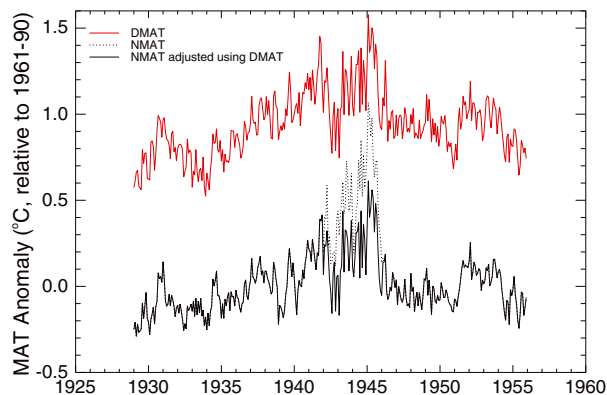


Figure 9. Monthly average marine air temperature anomaly (relative to the same 1961–1990 climatology) for (red) daytime, (black dotted) nighttime, and (black solid) nighttime, with NMAT anomalies for 1942–February 1946 replaced by adjusted DMAT anomalies.

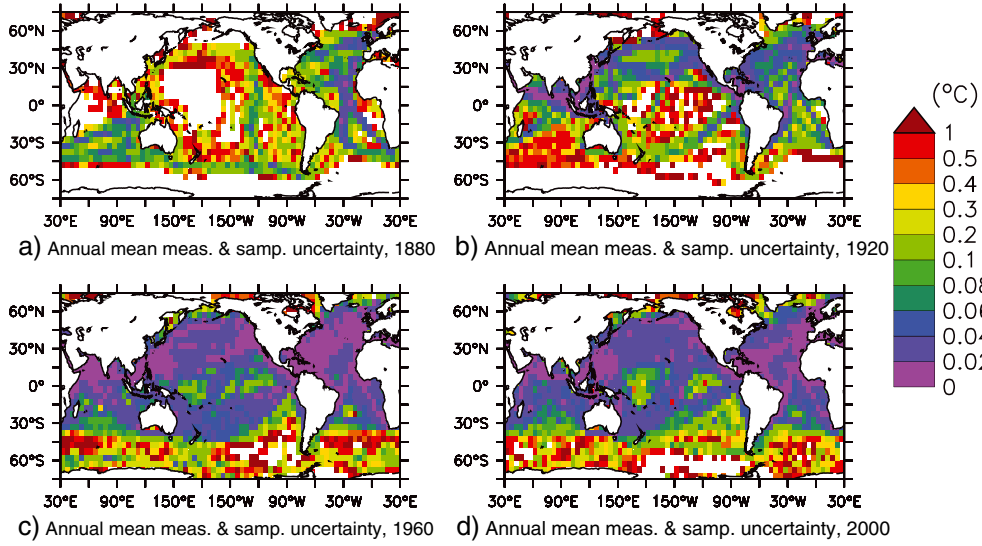


Figure 10. As Figure 7 but for sampling and random measurement uncertainty ($^{\circ}\text{C}$).

[47] The quality control procedures employed were the same as those used by *Rayner et al.* [2006], but with different acceptable limits to those used for SST. Each observation was checked to ensure that it has a meaningful location, date, and time and that it is not land locked, that its position is consistent with the previous track of the ship, and that the measurement is within $\pm 10^{\circ}\text{C}$ of the climatological value for the relevant 5 day period, interpolated to the day of observation. A buddy check is also performed, comparing each observation to the mean of its neighboring observations (see *Rayner et al.* [2006] for details). Observations from ICOADS Deck 780 with platform type 5 (originating from the World Ocean Database) [*Boyer et al.*, 2009] were found to be erroneous (Z. Ji and S. Worley, personal communication, 2011) and were excluded from HadNMAT2.

5.2. Estimation of Uncertainties

[48] There are three components to the uncertainty in each gridded NMAT anomaly value. These components are combined to give an estimate of the total uncertainty in each gridded value, assuming they are uncorrelated with each other.

[49] The first uncertainty component is the sampling and random measurement uncertainty; sampling uncertainty accounts for the possible error in each gridded value arising from under-sampling of the possible variability in the grid box. We use the same method as in *Rayner et al.* [2006] and assume the sampling and random measurement uncertainty to be uncorrelated between grid boxes. First, we estimate the uncertainty in each grid box average which would arise should that average have been calculated from only one observation. This value is then divided by the square root of the number of measurements comprising each monthly grid box average to give the sampling and uncorrelated measurement uncertainty in that average. Figure 10 shows annual average sampling and random measurement uncertainty for four example years.

[50] As discussed in section 3.2, the adjustment made to each NMAT measurement to account for the effect of

changing measurement height also has an uncertainty, calculated via the generation of an ensemble of possible such adjustments for each measurement (for the uncorrelated component) and from estimation (for the correlated component). These estimates of the second component of uncertainty are aggregated separately to produce an estimate of the correlated and uncorrelated components of the uncertainty in the grid box average height adjustment. The uncorrelated component of the height adjustment uncertainty is also uncorrelated *between* grid boxes and the correlated component is completely correlated.

[51] Uncertainties in the WW2 adjustment are calculated as described in section 4.4 and are assumed to be perfectly correlated between grid boxes. This third component is added to the other two in quadrature to produce fields of total NMAT uncertainty in each monthly 5° area.

[52] Our method of sampling and random measurement uncertainty calculation follows *Rayner et al.* [2006] and not the more advanced error model presented by *Kennedy et al.* [2011a]. Figure 11 shows the time-varying ratios of measurement and sampling uncertainties calculated following *Kennedy et al.* [2011a] (HadSST3) to those calculated following *Rayner et al.* [2006] (HadSST2). The more advanced error model results in increased uncertainty over the entire period, and especially toward the start and end of the record. The uncertainties calculated for HadNMAT2 are therefore expected to be underestimates, and the ratios

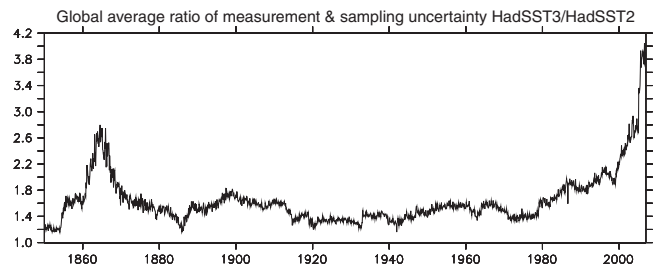


Figure 11. Ratio of mean measurement and sampling uncertainty HadSST3/HadSST2.

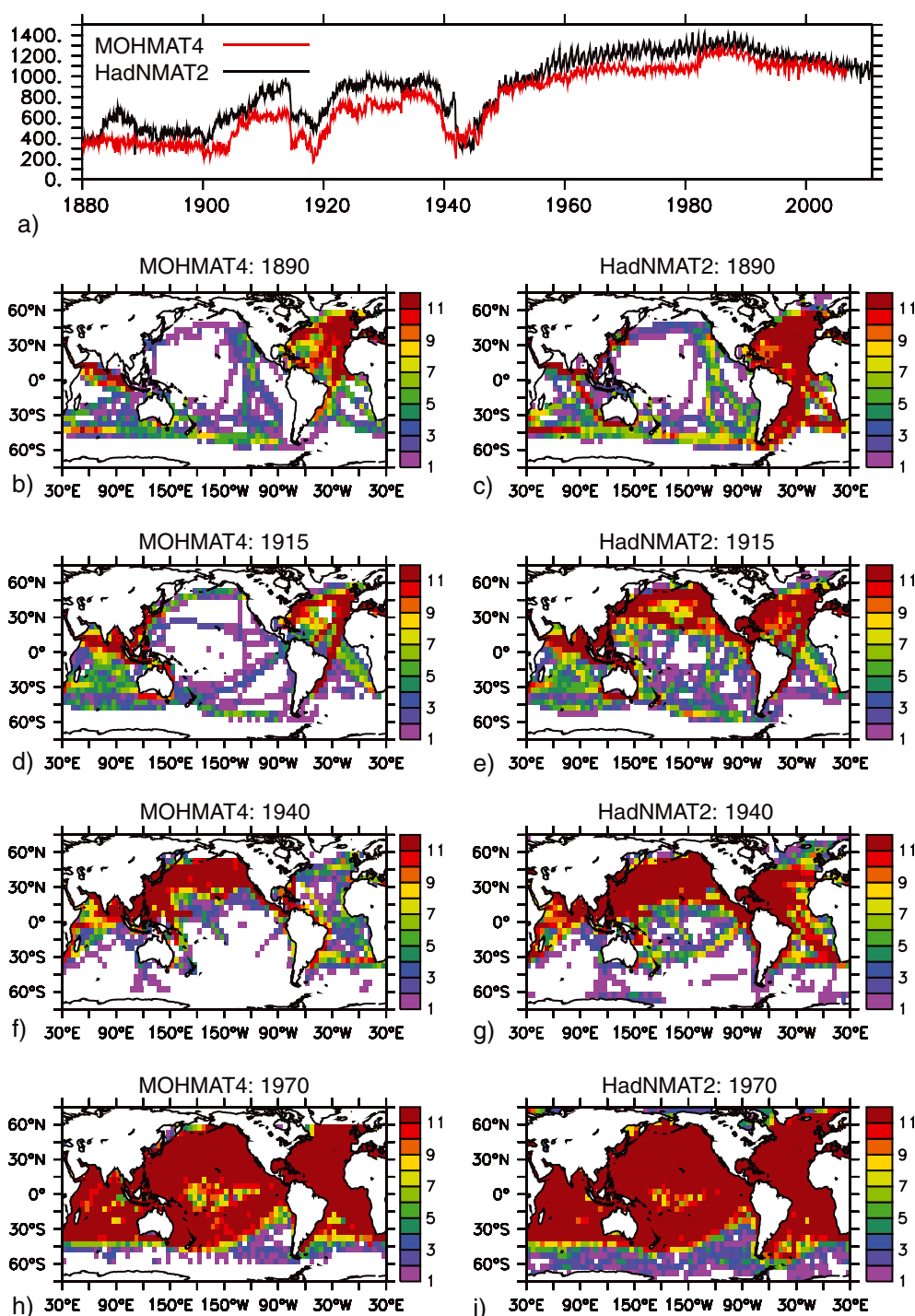


Figure 12. Comparison of data coverage in MOHMAT4 and HadNMAT2 data sets. (a) Number of 5° grid boxes sampled each month in MOHMAT4 (red) and HadNMAT2 (black); (b) number of months sampled in 1890 in MOHMAT; (c) as Figure 12b but for HadNMAT2; (d) MOHMAT4 1915; (e) HadNMAT2 1915; (f) MOHMAT 1940; (g) HadNMAT2 1940; (h) MOHMAT4 1970; and (i) HadNMAT2 1970.

in Figure 11 are likely to be indicative of the underestimation of measurement and sampling uncertainty in HadNMAT2.

6. The Characteristics of HadNMAT2

6.1. Improvements in Coverage

[53] In recent years, many new marine observations have been digitized [Woodruff *et al.*, 2005; Brohan *et al.*,

2009; Wilkinson *et al.*, 2011; Allan *et al.*, 2011]. Figure 12 compares the number of grid boxes sampled in MOHMAT4 and HadNMAT2. The total number of sampled 5° grid boxes is on average about 20% greater in the new dataset than MOHMAT4, but the amount of extra data coverage varies over time (Figure 12a). The maps in Figures 12b–12i show the numbers of months sampled in four sample years in the two data sets. In 1890 coverage increases by ~20% but much

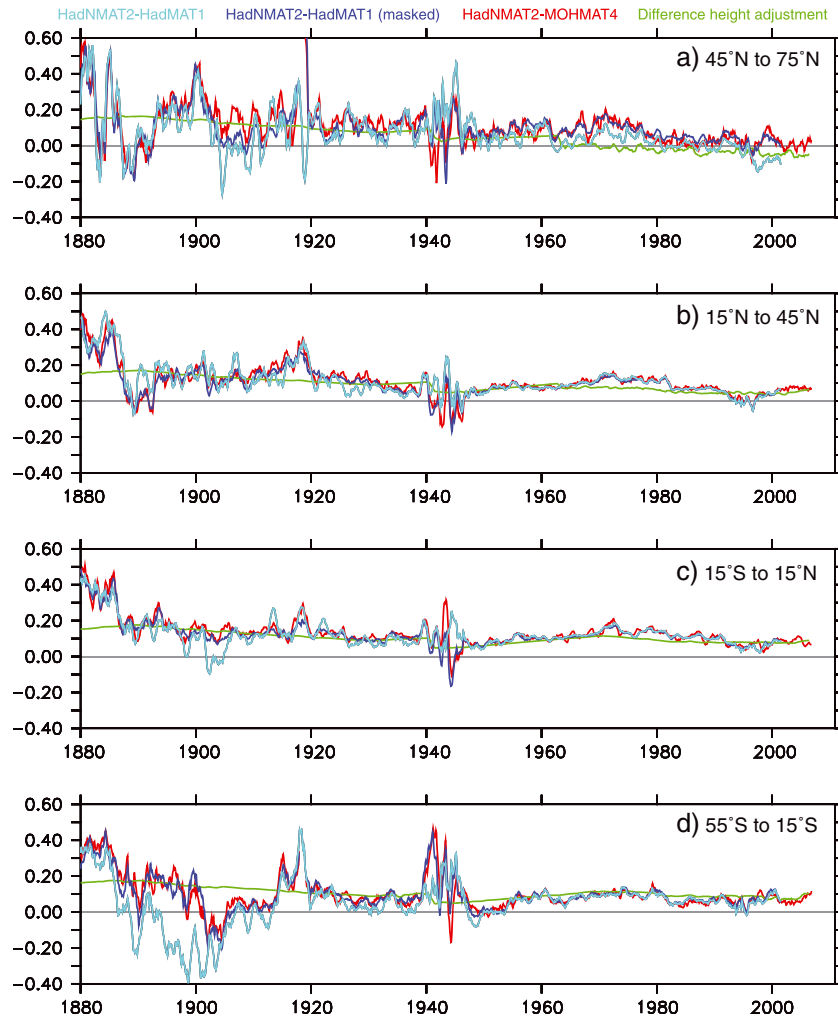


Figure 13. Difference between HadNMAT2, HadMAT1 and MOHMAT4 averaged in four latitude bands and smoothed with a 12-point running mean filter. HadNMAT2-HadMAT1 (light blue); HadNMAT2-HadMAT1 masked where MOHMAT4 contains no information (blue); HadNMAT2-MOHMAT4 (red); difference in height adjustment HadNMAT2-HadMAT1 (green, see also Figure 4). (a) 45°N to 75°N; (b) 15°N to 45°N; (c) 15°S to 15°N; and (d) 55°S to 15°S.

of the central Pacific remains unsampled. In 1915 coverage increases by ~30% and sampling in the North Pacific is substantially improved. There is a reduction in coverage in the western south Atlantic between 1890 and 1915 that is probably due to a change in shipping routes following the opening of the Panama Canal in 1914. In 1940 the increased coverage (27%) is most noticeable in the Atlantic. Figure 12a however shows that there are fewer observations in HadNMAT2 than MOHMAT between 1942 and the end of WW2, probably due to the exclusion of Deck 195 (section 4.4). In 1970, the increased coverage (~27%) is more uniform and is probably due to the ingestion of delayed mode observations by ICOADS.

6.2. Comparisons with Previous Versions: MOHMAT4 and HadMAT1

[54] Figure 13 compares the air temperatures from the new data set with those from the uninterpolated MOHMAT4 and the interpolated HadMAT1 data sets as averages in four latitudinal bands. Difference fields have been calculated and then the average of the differences in all grid boxes

where information from the relevant data set pairs exist is taken; i.e., these are collocated differences. Also plotted is the difference between the height adjustments applied in HadNMAT2 and HadMAT1.

[55] Differences are noisy in the most northerly and southerly latitude bands. Here, differences are expected to be large due to relatively poor data coverage; any new observations available in HadNMAT2 will have the greatest impact on area averages in the most data sparse regions. Differences here between HadMAT1 and MOHMAT4 are also greatest as a reconstruction method was used in HadMAT1 to infer information in unobserved grid boxes [Rayner *et al.*, 2005].

[56] The mean difference between the adjustments applied to account for differences in observation height indicates the expected value of the difference between HadNMAT2 and either HadMAT1 or MOHMAT4. Variation about this expected difference will be due to differences in the input observations, the different approaches taken to adjust for non-standard observing practice (prior to 1893 and during WW2) and to the effects of the HadMAT1 reconstruction approach. In well-sampled regions and periods, the data

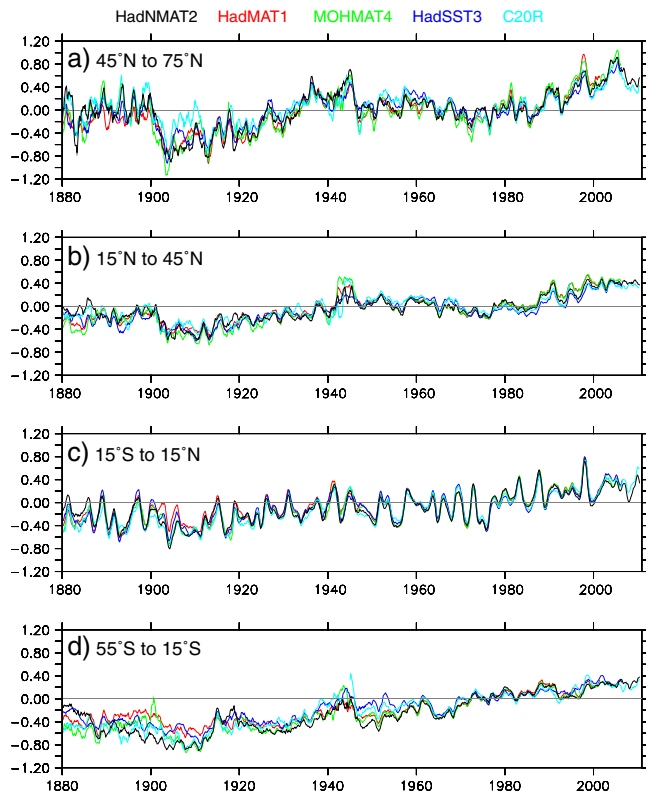


Figure 14. Near-surface temperature anomaly ($^{\circ}\text{C}$, relative to 1961–1990) 1880–2010 in four latitude bands: HadNMAT2 (black); HadMAT1 (red); MOHMAT4 (green); HadSST3 median (blue); and C20R (light blue). HadNMAT2 anomalies been adjusted to average to zero over the period 1961–1990. HadMAT1 and MOHMAT4 anomalies are as provided. C20R masked where HadNMAT2 contains no information, but otherwise the time series are not collocated. (a) 45°N to 75°N ; (b) 15°N to 45°N ; (c) 15°S to 15°N ; and (d) 55°S to 15°S .

set differences are typically close to the expected difference. Before about 1886, HadNMAT2 is typically warmer by about 0.4°C than the other data sets. This is probably related to the non-standard observing practice described in section 4.2; HadNMAT2 is likely too warm in this period. Indeed, HadMAT1 was adjusted through 1885 using SST anomalies for this reason. However, we note that no adjustments have been applied to HadNMAT2 to account for these effects.

6.3. Comparisons with Other Data Sets: HadSST3, C20R, CRUTEM4

[57] In addition to showing HadNMAT2 average anomalies for our four latitude bands with MOHMAT4 and HadMAT1, Figure 14 compares to air temperature anomalies from the Twentieth Century Reanalysis Project (C20R, v2) [Compo *et al.*, 2011] and SST anomalies from the HadSST3 median [Kennedy *et al.*, 2011a, 2011b]. Anomalies are calculated for C20R using the long term monthly mean data provided, then these anomalies have been adjusted to average to zero over the period 1961–1990 by subtracting their respective grid box monthly means over this period. Anomalies for MOHMAT4 were adjusted in the same way. There is broad agreement among the data sets in these large

regional averages. The higher temperatures prior to 1886 of HadNMAT2 relative to MOHMAT4 and HadMAT1 are less obvious when viewed against the range of mean anomalies among the different data sets. The warm biases in WW2 are most obvious in MOHMAT4 in the midlatitudes (Figures 14b and 14d).

[58] Large differences are seen between HadNMAT2 and both the HadSST3 median and MOHMAT4 in the mid-1940s to 1950s, in the region 55°S to 15°S with HadNMAT2 relatively cooler. Maps of 5 year mean temperature anomaly (not shown) indicate that a warm anomaly present in the South Pacific in the HadSST3 median is missing from HadNMAT2, but weakly present in MOHMAT4.

[59] The largest divergence between the mean temperature anomalies is in the region 55°S to 15°S prior to about 1915 (Figure 14d). Here HadNMAT2 mean anomalies are warmer than -0.2°C in 1880 and cooler than -0.8°C in 1900; none of the other data sets show such a strong trend over this short period. HadNMAT2 agrees well with the HadSST3 median in the South Atlantic in the mean anomaly map for 1895–1899 (not shown), but is colder than the HadSST3 median in the southern Pacific. However, there is substantial uncertainty in the mean anomalies for this period and region, shown by the lack of agreement among the data sets.

[60] There is evidence in most of the latitude bands that there may be a warm bias in the recently digitized measurements for the WW1 period. A previous disagreement between NMAT and SST data in the 1990s (Figure 14b) has been resolved, with HadNMAT2 now closer to the HadSST3 median than the earlier data sets.

[61] Figure 15 shows the difference between each of the datasets plotted in Figure 14 and the HadSST3 median. HadNMAT2 generally agrees well with the HadSST3 median, with best agreement in the tropics and the mid-northern latitudes (outside of the WW2 period). Here again we see the relatively large offset in the late-1940s and 1950s in the region 55°S to 15°S , with HadNMAT2 cooler than the HadSST3 median. The HadSST3 ensemble spread is large in this region for this period, but the differences are large and require further investigation. The adjustments applied to MOHMAT4 during WW2 appear to be the least effective (compared to the HadSST3 median) of the observation-only data sets.

[62] The air temperature anomalies are compared with anomalies from CRUTEM4 [Jones *et al.*, 2012] in those 5° grid boxes and months where both CRUTEM4 and MOHMAT4 are present (Figure 16). During WW2, extratropical NMAT anomalies from all data sets are warmer than collocated land anomalies in 5° grid boxes with both land and ocean. This suggests that there are biases in the daytime marine air temperatures during this period and further investigation is required. In the tropics (Figure 16c), the WW2 period NMAT anomalies are warmer relative to the land anomalies than either HadMAT1 or MOHMAT4, but observations are sparse in this region and period. There is a suggestion that land anomalies are warming slightly faster than marine anomalies from both HadMAT1 and HadNMAT2 after about 1980. Although the marine and land grid boxes compared are co-located, the relatively coarse 5° grid would permit differences between marine and land data to exist. However, restricting the comparison according to the fraction of land contained within each grid

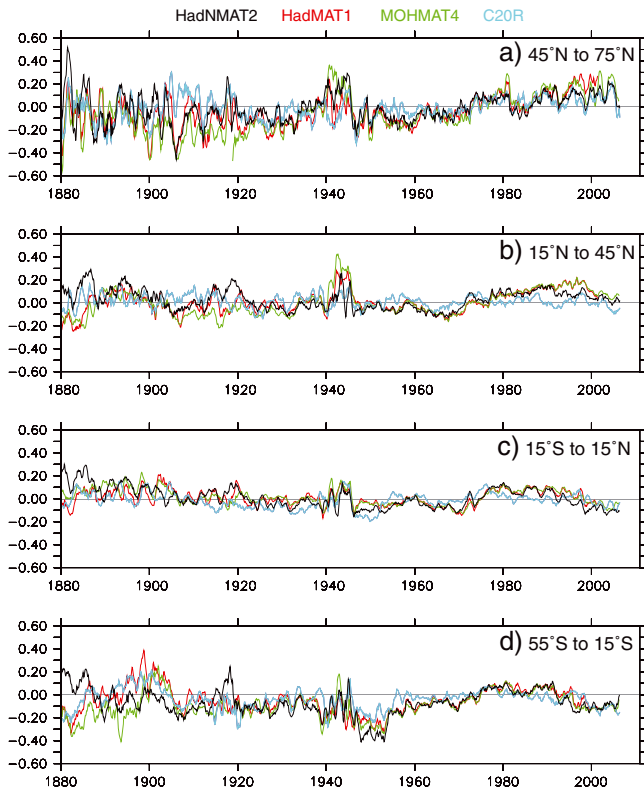


Figure 15. Air-sea anomaly temperature difference ($^{\circ}\text{C}$) 1880–2010. SST anomalies are from the HadSST3 median. Differences are calculated prior to averaging and are collocated. HadNMAT2–HadSST3 (black); HadMAT1–HadSST3 (red); MOHMAT4–HadSST3 (green); and C20R–HadSST3 masked for presence of HadNMAT2 (light blue). HadNMAT2 adjusted as in Figure 14. (a) 45°N to 75°N ; (b) 15°N to 45°N ; (c) 15°S to 15°N ; and (d) 55°S to 15°S .

box did not reduce the difference between the co-located land and marine anomalies.

[63] Figure 17 shows 5 year averages of anomalies for HadNMAT2, MOHMAT4, HadMAT1, HadCRUT4, and C20R for four example periods. The increased sampling in HadNMAT2 over MOHMAT4 is clear in each of the four periods. In particular, our ability to study the average marine air temperature over the period 1890–1894 has been revolutionized by the recent digitization of measurements for this period. Anomalies from HadNMAT2 are closer to those from the HadSST3 median and C20R than either MOHMAT4 or HadMAT1 anomalies; the agreement in 1970–1974 is particularly close. In fact, the El Niño-like pattern in 1940–1944 looks more coherent in HadNMAT2 than in the HadSST3 median; unfortunately, this map also exhibits unrealistic-looking warm ship tracks in the Southern Hemisphere. One exception to the generally improved agreement is the period 1915–1919, which is relatively warm in HadNMAT2 in the Pacific Ocean: see also Figure 13d.

6.4. Global Temperature Time Series and Its Uncertainty

[64] Figure 18 shows global anomalies from HadNMAT2 and SST from the HadSST3 ensemble median. Uncertainties

in the global values are also plotted. For NMAT, these uncertainties are a combination of those due to the bias adjustment and those due to uncorrelated measurement and sampling. For HadSST3, two ranges of uncertainty are plotted; the narrower band is equivalent to that plotted for HadNMAT2. The wider band also includes uncertainty due to incomplete global coverage and to correlated measurement and sampling. As noted in section 5.2, these uncertainties also affect NMAT but have not so far been included in the uncertainty calculation for HadNMAT2. We also note that SST is more widely sampled throughout the period than NMAT, initially due to the exclusion of daytime MAT, and in the later part of the record because of the availability of SST measurements from drifting buoys.

[65] The broad agreement between SST and NMAT anomalies, in the context of their uncertainty estimates is good. However there are some periods where there are interesting differences between these two different, and largely independent, measures of marine surface temperature. In the 1880s, HadNMAT2 is warmer than the HadSST3 median. This is likely to be a problem with the NMAT measurements, a continuation of the problem with the early instrumental record identified by previous authors (including *Bottomley et al.* [1990]). Higher temperatures are seen in NMAT during WW2 and to a lesser extent in SST. These are again due to

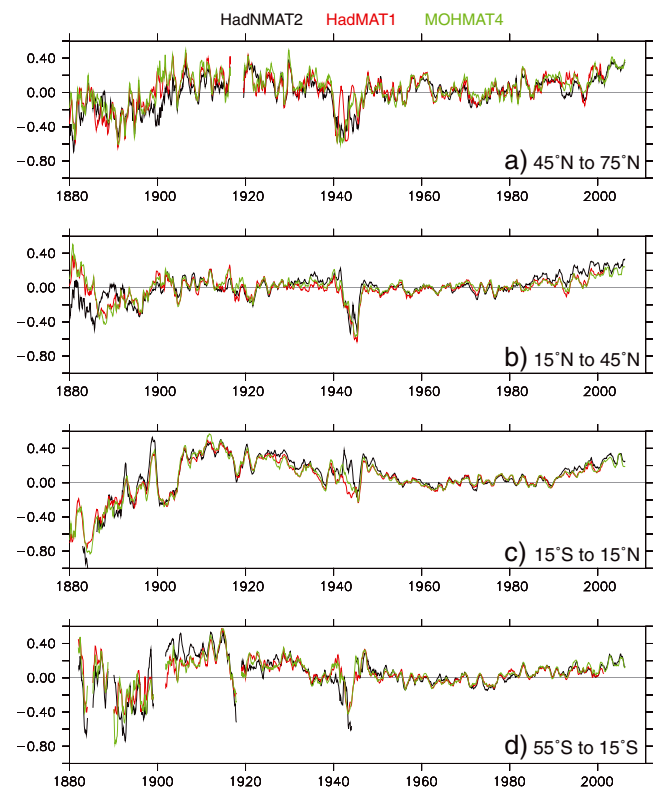


Figure 16. Land-ocean anomaly differences ($^{\circ}\text{C}$) 1880–2010. Land anomalies are from CRUTEM4. CRUTEM4–HadNMAT2 (black); CRUTEM4–HadMAT1 (red); and CRUTEM4–MOHMAT4 (green). HadNMAT2 adjusted as in Figure 14. Comparisons masked for presence of MOHMAT4. (a) 45°N to 75°N ; (b) 15°N to 45°N ; (c) 15°S to 15°N ; and (d) 55°S to 15°S .

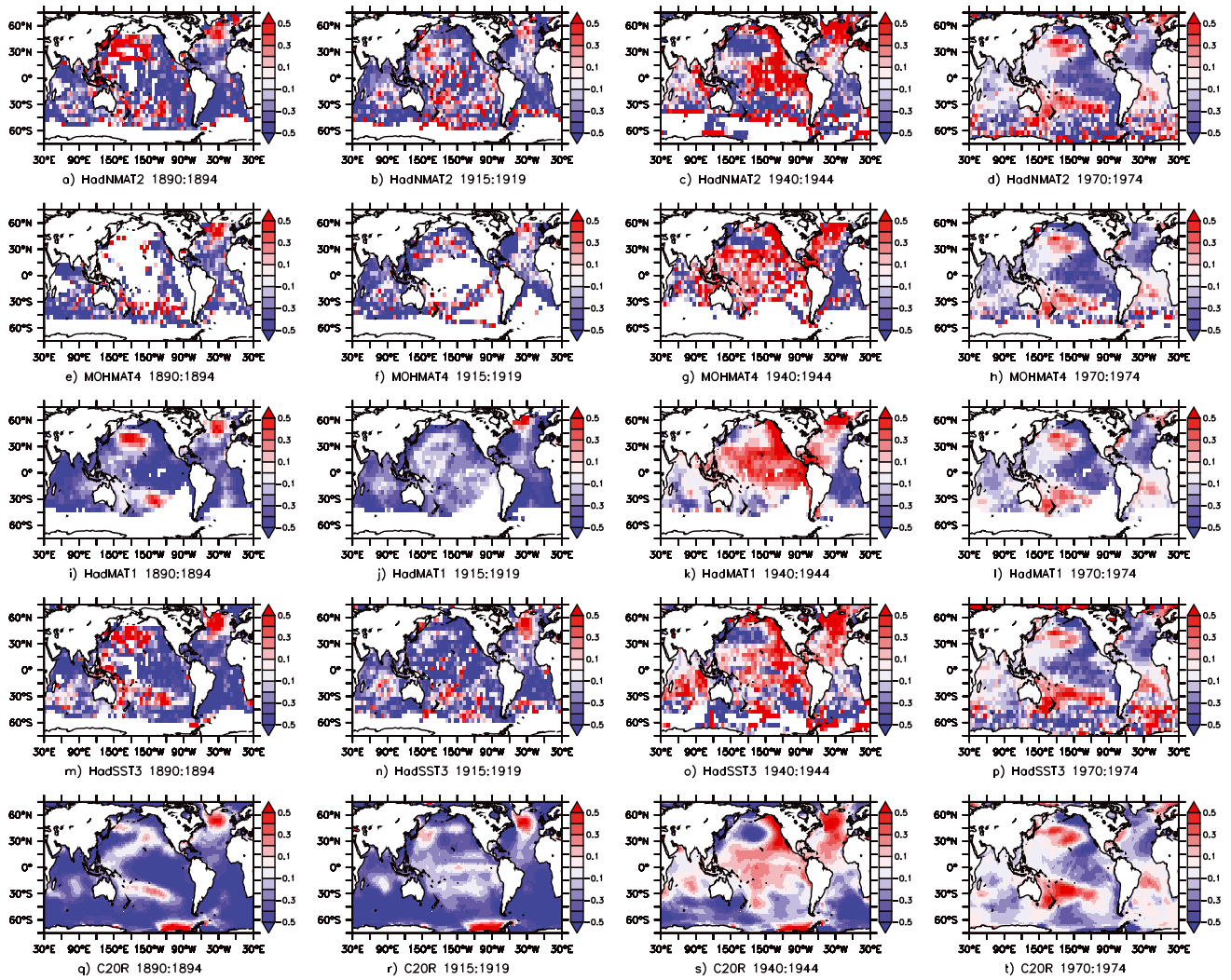


Figure 17. Five-year mean surface temperature anomalies ($^{\circ}\text{C}$, relative to 1961–1990). First column: 1890–1894; second column: 1915–1919; third column: 1940–1944; and fourth column: 1970–1974. First row: (a–d) HadNMAT2; second row: (e–h) MOHMAT4; third row: (i–l) HadMAT1; fourth row: (m–p) HadSST3 median; and fifth row: (q–t) C20R.

known problems with changing observing practice during wartime and are less prevalent in the collocated land air temperatures (e.g., Figure 16). The decades surrounding WW2 show NMAT anomalies consistently lower than SST anomalies by an amount that the uncertainty estimates suggest is unlikely to be realistic. We stress however that this comparison is not collocated and that NMAT uncertainties are expected to be too small. Differences in the mean height adjustment applied to HadNMAT2 compared to HadMAT1 are small in this period (Figure 4, offset line). Relative to previous adjustments they will have acted to slightly raise NMAT relative to SST prior to WW2 and to lower NMAT relative to SST between the end of WW2 and about 1970. The maximum impact of the change in height adjustment on the global mean around the WW2 period is of order 0.05°C . The relatively cold NMAT in the decades surrounding WW2 may be due to unresolved biases in either HadNMAT2 or in HadSST3 and will be the subject of further investigations.

[66] There are also differences between HadNMAT2 and HadSST3 in the more recent record. We note that after about

1996, there is excellent agreement in the global time series of HadSST3 and independent SST measurements from the Along Track Scanning Radiometer series of satellites [Merchant *et al.*, 2012]. Differences in geographical coverage

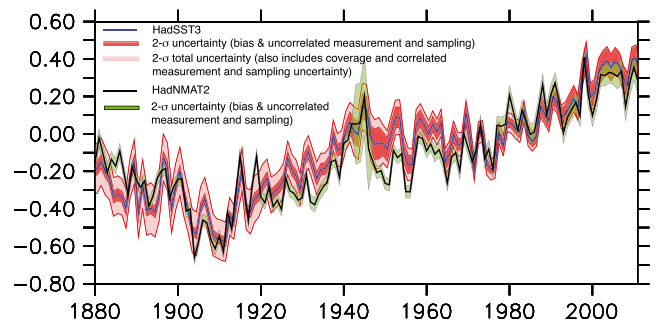


Figure 18. Global annual average timeseries of HadNMAT2 and HadSST3 median anomalies ($^{\circ}\text{C}$, relative to 1961–1990) and their estimated uncertainties, 1880–2010.

of SST and NMAT measurements made in situ may play a role here.

7. Key Results and Remaining Issues

[67] The HadNMAT2 is an overall improvement over MOHMAT4 and HadMAT1, due in large part to the additional observations that have recently become available in ICOADS Release 2.5. The adjustment for variations in observation height has been performed more rigorously than for MOHMAT4 and the uncertainty due to the adjustment process estimated for the first time. This suggests that an interpolated NMAT data set based on HadNMAT2 would be an improvement over HadMAT1.

[68] It is possible though that the adjustments applied to the data after WW2 are not applicable to the data in the region 15 to 55°S, since there is a relative cool bias of about 0.4°C here during the mid-1940s to mid-1950s, as compared to the HadSST3 ensemble median.

[69] The HadNMAT2, unlike MOHMAT4 and HadMAT1 is not dependent on time-varying SST for any adjustment, although at the cost of a shorter data set. The requirement for the Suez adjustment was removed by the exclusion of observations rather than using SST anomalies. WW2 biases in NMAT are adjusted using daytime marine air temperature anomalies, as in previous data sets. The adjustment appears to have slightly better results than that used in MOHMAT4 and is applied over a shorter period. However, comparisons with collocated land anomalies suggest that HadNMAT2 remains too warm during WW2. Further investigation of the daytime marine air temperatures is therefore required. Additionally, our analysis suggests that the data prior to 1886 are also erroneously warm and should not be relied upon.

[70] The new observations for the First World War period may also contain relatively small warm biases and should be further investigated.

[71] More generally, the new observations validate the HadMAT1 reconstruction approach to some extent. The differences between HadNMAT2 and MOHMAT4 are correlated with the differences between HadMAT1 and MOHMAT4, showing that where the reconstruction has altered data, this is subsequently confirmed (on average) when new observations have become available.

[72] Improved methods for estimating measurement and sampling uncertainty have recently been developed [Kennedy *et al.*, 2011a] and future versions of HadNMAT will use this methodology. The likely impact of this change has been estimated by examining the impact of the changed methodology on estimates of uncertainty for SST.

[73] The lack of call signs in ICOADS after 2007 has increased the uncertainty in HadNMAT2 for this period. Also, metadata from Pub. 47 are not presently up to date. The lack of both call signs and metadata requires urgent resolution to ensure that marine surface datasets can be properly adjusted and their uncertainty estimated.

[74] These shortcomings notwithstanding, HadNMAT2 provides valuable independent corroboration of the other components of the surface temperature observing system.

[75] **Acknowledgments.** This work was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) and received funding through the National Oceanography Centre's National Capability programme. HadNMAT2, HadMAT1, MOHMAT4, HadSST3,

and CRUTEM4 are all available from the Met Office Hadley Centre data server at <http://www.metoffice.gov.uk/hadobs/>. International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 2.5, Individual Observations are available as dataset ds540.0 published by the CISL Data Support Section at the National Center for Atmospheric Research, Boulder, CO, available online at <http://dss.ucar.edu/datasets/ds540.0/>. Scott Woodruff, Sandy Lubker, Steven Worley, and Eric Freeman have all provided help and expert advice when needed. The authors wish to acknowledge use of the Ferret program for analysis and graphics in this paper. Ferret is a product of NOAA's Pacific Marine Environmental Laboratory (information is available at <http://ferret.pmel.noaa.gov/Ferret/>). 20th Century Reanalysis V2 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at <http://www.esrl.noaa.gov/psd/>. Support for the Twentieth Century Reanalysis Project data set is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office. We wish to acknowledge the earlier, unpublished work of Julian Hill, Jen Hardwick, and Simon Tett which helped in this study.

References

- Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Brönnimann (2011), The international atmospheric circulation reconstructions over the Earth (ACRE) initiative, *Bull. Amer. Meteor. Soc.*, 92, 1421–1425, doi:10.1175/2011BAMS3218.1.
- Berry, D. I., and E. C. Kent (2005), The effect of instrument exposure on marine air temperatures: An assessment using VOSclim data, *Int. J. Climatol.*, 25(7), 1007–1022, doi:10.1002/joc.1178.
- Berry, D. I., and E. C. Kent (2006), WMO publication 47: Consistency checking and gap filling, version 1.0, documentation for metadata database, 77 pp. [Available from <http://eprints.soton.ac.uk/341419/>].
- Berry, D. I., and E. C. Kent (2009), A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates, *Bull. Amer. Meteor. Soc.*, 90(5), 645–656, doi:10.1175/2008BAMS2639.1.
- Berry, D. I., and E. C. Kent (2011), Air-sea fluxes from ICOADS: The construction of a new gridded dataset with uncertainty estimates, *Int. J. Climatol.*, 31, 987–1001, doi:10.1002/joc.2059.
- Berry, D. I., E. C. Kent, and P. K. Taylor (2004), An analytical model of heating errors in marine air temperatures from ships, *J. Atmos. Oceanic Technol.*, 21(8), 1198–1215, doi:10.1175/JCLI1520-0426(2004)021%3C1198:AAMOHE%3E2.0.CO;2.
- Board of Trade (1857), Meteorological observations: Directions for keeping the weather book, abstract of log, or meteorological register, *First Number of Meteorological Papers published by authority of the Board of Trade*, London, Printed by George Edward Eyre and William Spottiswoode for Her Majesty's Stationery Office, pp. 56–63.
- Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker (1990), Global ocean surface temperature atlas "GOSTA". Meteorological Office, Bracknell, U.K. and the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. 20 pp. and 313 plates.
- Boyer, T. P., J. I. Antonov, O. K. Baranova, H. E. Garcia, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, D. Seidov, I. V. Smolyar, and M. M. Zweng (2009), World ocean database 2009, vol. 1, Introduction, NOAA Atlas NESDIS, vol. 66, edited by S. Levitus, 219 pp., NOAA, Silver Spring, Md.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, 111, D12106, doi:10.1029/2005JD006548.
- Brohan, P., R. Allan, J. E. Freeman, A. M. Waple, D. Wheeler, C. Wilkinson, S. Woodruff (2009), Marine observations of old weather, *Bull. Amer. Meteor. Soc.*, 90, 219–230, doi:10.1175/2008BAMS2522.1.
- Chenoweth, M. (2000), A new methodology for homogenization of 19th century marine air temperature data, *J. Geophys. Res.*, 105, 29,145–29,154.
- Compo, G. P., et al. (2011), The twentieth century reanalysis project, *Quarterly J. Roy. Meteorol. Soc.*, 137, 1–28, doi:10.1002/qj.776.
- Dixon, W. J. (1960), Simplified estimation from censored normal samples, *The Annals of Mathematical Statistics*, 31, 385–391.
- Dyer, A. J. (1974), A review of flux-profile relations, *Bound.-Layer Meteorol.*, 1, 363–372.
- Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young (1996), Bulk parameterization of air-sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment, *J. Geophys. Res.*, 101(C2), 3747–3764, doi:10.1029/95JC03205.
- Haimberger, L., C. Tavalato, and S. Sperka (2012), Homogenization of the global radiosonde dataset through combined comparison with reanalysis background series and neighboring stations, *J. Climate*, 25, 8108–8131; View, doi:10.1175/JCLI-D-11-00668.1

- Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.
- Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto (2005), Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection, *Int. J. Climatol.*, **25**, 865–879, doi:10.1002/joc.1169.
- Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010, *J. Geophys. Res.*, **117**, D05127, doi:10.1029/2011JD017139.
- Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan (1998), Analyses of global sea surface temperature 1856–1991, *J. Geophys. Res.*, **103**, 18,567–18,589.
- Kennedy J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011a), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: Measurement and sampling errors, *J. Geophys. Res.*, **116**, D14103, doi:10.1029/2010JD015218.
- Kennedy J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011b), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: Biases and homogenization, *J. Geophys. Res.*, **116**, D14104, doi:10.1029/2010JD015220.
- Kent, E. C., S. D. Woodruff, and D. I. Berry (2007), Metadata from WMO publication No. 47 and an assessment of voluntary observing ship observation heights in ICOADS, *J. Atmos. and Oceanic Technol.*, **24**(2), 214–234.
- Large, W. G., and S. Pond (1982), Sensible and latent heat flux measurements over the ocean, *J. Phys. Oceanogr.*, **12**, 464–482.
- Lindsay, W. S. (1876), History of Merchant Shipping and Ancient Commerce, Vol. 4, Sampson Low, Marston, Low and Searle, London, U.K., 666 pp.
- Marshall, M. W. (1989), Ocean Traders. From the Portuguese Discoveries to the Present Day, B. T. Batsford Ltd., London, UK, 192 pp.
- Maury, M. F. (1854), “Maritime conference held at Brussels for devising a uniform system of meteorological observations at sea, August and September, 1853”, in Explanations and Sailing Directions to Accompany the Wind and Current Charts, 6th Ed., E. C. and J. Biddle, Philadelphia, pp. 54–96. Available from: <http://archive.org/details/explanationssail00-maur>, accessed 16 November 2012.
- Menne, M. J., C. N. Williams, and R. S. Vose (2009), The U.S. historical climatology network monthly temperature data, version 2, *Bull. Amer. Meteorol. Soc.*, **90**, 993–1007, doi:10.1175/2008BAMS2613.1.
- Embury, O. et al. (2012), A twenty-year independent record of sea surface temperature for climate from along track scanning radiometers, *J. Geophys. Res.*, **117**, C12013, doi:10.1029/2012JC008400.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset, *J. Geophys. Res.*, **117**, D08101, doi:10.1029/2011JD017187.
- Parker, D. E., C. K. Folland, and M. Jackson (1995), Marine surface temperature: Observed variations and data requirements, *Clim. Chang.*, **31**, 559–600.
- Parker, D. E., R. B. Hackett, and M. Jackson (1997), Marine air temperatures: The New MOHMAT4 datasets, Hadley Centre Internal Note No. 82, 18pp + figs.
- 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 19th century, *J. Geophys. Res.*, **108**(D14), 4407, doi:10.1029/2002JD002670.
- Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. Ansell and S. F. B. Tett (2006), Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 data set, *J. Climate*, **19**(3), 446–469, doi:10.1175/JCLI3637.1.
- Smith, S. D. (1980), Wind stress and heat flux over the ocean in gale force winds, *J. Physical Oceanography*, **10**, 709–726.
- Smith, S. D. (1988), Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature, *J. Geophys. Res.*, **93**(C12), 15,467–15,472.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA’s historical merged land-ocean surface temperature analysis (1880–2006), *J. Climate*, **21**, 2,283–2,296, 10.1175/2007JCLI2100.1.
- Thompson, D. W. J., J. J. Kennedy, J. M. Wallace, and P. D. Jones (2008), A large discontinuity in the mid-twentieth century in observed global-mean surface temperature, *Nature*, **453**, 646–649, doi:10.1038/nature06982.
- Thorne, P. W., D. E. Parker, S. F. B. Tett, P. D. Jones, M. McCarthy, H. Coleman, and P. Brohan (2005), Revisiting radiosonde upper-air temperatures from 1958 to 2002, *J. Geophys. Res.*, **110**, D18105, doi:10.1029/2004JD005753.
- Wilkinson, C., S. D. Woodruff, P. Brohan, S. Claesson, E. Freeman, F. Koek, S. Lubker, C. Marzin, and D. Wheeler (2011), Recovery of logbooks and international marine data: The RECLAIM project, *Int. J. Climatol.*, **31**: 968–979, doi:10.1002/joc.2102.
- Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley (1998), COADS release 2 data and metadata enhancements for improvements of marine surface flux fields, *Phys. Chem. Earth*, **23**, 517–526.
- Woodruff, S. D., H. F. Diaz, S. J. Worley, R. W. Reynolds, and S. J. Lubker (2005), Early ship observational data and ICOADS, *Clim. Chang.*, **73**, 169–194, doi:10.1007/s10584-005-3456-3.
- Woodruff, S. D. (2011), ICOADS release 2.5 and data characteristics, *Int. J. Climatol.*, **31**, 951–967, doi:10.1002/joc.2103.
- 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. Climatol.*, **25**, 823–842, doi:10.1002/joc.1166.