Received: 20 December 2019 | Revised: 22 May 2020 | Accepted: 11 June 2020

DOI: 10.1002/gdj3.100

DATA PAPER

Geoscience Data Journal | RMetS | WILEY

CLASSnmat: A global night marine air temperature data set, 1880–2019

Richard C. Cornes[1] | **Elizabeth.C. Kent**[1] | **David.I. Berry**[1] | **John J. Kennedy**[2]

[1]National Oceanography Centre, Southampton, UK

[2]Met Office Hadley Centre, Exeter, UK

Correspondence
Richard C. Cornes, National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK.
Email: ricorne@noc.ac.uk

Funding information
NERC CLASS programme, Grant/Award Number: NE/R015953/1; NERC HOSTACE project, Grant/Award Number: NE/J020788/1; NERC GloSAT project, Grant/Award Number: NE/S015647/2; Copernicus Climate Change Service, Grant/Award Number: C3S_311a_Lot2; Met Office Hadley Centre Climate Programme funded by BEIS and Defra

Abstract

A new data set of Night Marine Air Temperature (NMAT) is presented that builds on the HadNMAT2 data set, which was released in 2013. In a similar manner to HadNMAT2, the new data set (CLASSnmat) provides uninterpolated, monthly global values at a 5° resolution back to 1880. In addition to being extended to the end of 2019, four main developments are made in CLASSnmat: (1) the NMAT values are extracted from the most recent version of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS Release 3) and a revised method of eliminating duplicated observations is used; (2) values of NMAT are adjusted to 2m and 20m heights in addition to the 10 m height used in HadNMAT2; (3) a refinement is made to the corrections necessary during World War 2, which uses more of the NMAT observations and hence results in a more extensive spatial coverage for this period than was possible in HadNMAT2; (4) an updated gridding method is used that allows for an improved propagation of uncertainty from the individual NMAT values through to the gridded estimates. In this paper, the method used to construct CLASSnmat (version 1.0.0.0) is described.

KEYWORDS

climate change, global surface temperature, marine, NMAT, observations

1 | INTRODUCTION

The interactions that occur at the ocean–atmosphere boundary act as a major control on the climate system and understanding these processes requires not only measurements of sea surface temperature (SST) but also marine air temperature (MAT). The ship-based observations contained in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman *et al*., 2017) provide a global data set of observations of MAT spanning the past 150 + years and these

Dataset

Identifer: uuid/5bbf48b128bd488dbb10a56111feb36a

Creator: Richard C. Cornes, Elizabeth C. Kent, David I. Berry and John J. Kennedy

Title: CLASSnmat: monthly, global, gridded night marine air temperature data

Publication year: 2020

Version: 1.0.0.0

Available at: https://catalogue.ceda.ac.uk/uuid/5bbf48b128bd488dbb10a56111feb36a

data have previously been used to construct several gridded MAT datasets.

The National Oceanography Centre (NOC) surface flux dataset v2.0 (Berry and Kent, 2009, 2011) provided daily fields of several variables including MAT back to 1973 at a 1° spatial resolution. An adjustment was applied to the MAT data used in that data set to correct for the bias associated with heating of the ship via solar radiation during daylight hours (Berry *et al*., 2004). Other gridded marine temperature data sets extend back to the mid-nineteenth century, for example the Global Ocean Surface Temperature Atlas (GOSTA) (Bottomley *et al*., 1990), the Met Office Historical Marine Air Temperature data set version 4 (MOHMAT4) (Parker *et al*., 1997) and the Hadley Centre Marine Air Temperature data set version 1 (HadMAT1) (Rayner *et al*., 2003). A correction for daytime heating biases is not applied to the data in these data sets, and only observations recorded when the effect of solar heating is minimal are used (one hour after sunset to one hour after sunrise) to reduce the heating bias. As a reflection of their primary role in quantifying long-term changes in temperature, these century-long data sets are generally constructed as monthly anomaly fields, relative to a given base period, with a spatial resolution of 5°, although climatology fields are also provided in MOHMAT4 and HadMAT1 to allow for the reproduction of gridded absolute values.

The HadNMAT2 data set (Kent *et al*., 2013) provided several improvements over earlier Night Marine Air Temperature (NMAT) data sets: both absolute and anomaly fields were generated back to 1880, uncertainty estimates were calculated and a more consistent method was developed to adjust the measurements to a common height above sea level. In addition, certain adjustments were refined to ensure that SST measurements were not used directly in the production of the data set. More recently, a global monthly data set of NMAT back to 1900 has been developed (UAHNMATv1) by Junod and Christy (2019). While that data set was developed using different processing procedures compared to HadNMAT2, the data sets are broadly comparable in terms of long-term trends and decadal variability at both the regional and global scales.

In this paper, an updated version of HadNMAT2 is described. The data set (CLASSnmat[1]) follows the approach of its predecessor in providing 5°×5° global fields of monthly mean NMAT back to 1880, but updates the data set to December 2019. Several other amendments and additions are made in this data set: the latest version of ICOADS and its near-real time update are used, release 3.0 (Freeman *et al*., 2017) and 3.0.2 respectively (*c.f*. the use of ICOADS release 2.5 in HadNMAT2); a new duplicate identification procedure is employed to quality control the data; a refined gridding and uncertainty calculation are used; NMAT values are adjusted to 2m and 20m height, in addition to the 10m as used in HadNMAT2; and a refinement is made to the corrections applied to the data during World War 2.

[1]The dataset has developed as part of the Climate-Linked Atlantic Sector Science (CLASS) project, which is funded by the Natural Environment Research Council (NERC, see acknowledgements). It is from the project acronym that the dataset name has been derived.

2 | DATA PROCESSING

2.1 | Pre-processing of the ICOADS data

Ship-based air temperature measurements obtained from ICOADS R3.0 (Freeman *et al*., 2017) were used for the period 1880–2014; thereafter, data from the near-real time update of ICOADS R3.0.2 were used. Henceforth, we will simply refer to these data as ICOADS. Ship-based observations were identified using the ICOADS platform type (PT) codes 0-5 and 9. Observations from specialist ship sources, such as research vessels, were excluded. Observations in Deck 245 (UK Royal Navy Ship's Logs 1938-47) (Brohan *et al*., 2009; Wilkinson *et al*., 2011) that were recorded onboard submarines during World War 2 were also excluded from further analysis pending further assessment of this unusual data source.

As with HadNMAT2 and its predecessors, only night-time observations are used in CLASSnmat. The night-time period is defined using the same approach as HadNMAT2, as starting one hour after sunset and finishing one hour after sunrise. Analysis of the daily cycle of MAT values has revealed that the data from Deck 781 (Chinese/ Global Ocean Data Archeology and Rescue [GODAR] Ships) over the period 1968–1993 appear to be given as local standard time. These times have been adjusted to UTC prior to the calculation of sunrise/sunset times. Solar elevation is calculated using the algorithm described by Spencer (1989) in a similar manner to previous NMAT data sets (MOHMAT4 and HadNMAT2).

Duplicated observations exist in ICOADS as a consequence of the multitude of data sources used in the construction of the data set. These values were identified and removed in HadNMAT2 using the duplicate elimination (dupelim) procedure described by Slutz *et al*. (1985). In CLASSnmat, a new method of identifying duplicate reports is used that is based on the ICOADS R3.0 'Total' files; these files include all available duplicates (Freeman *et al*., 2017). The new approach follows the ICOADS technique of identifying similar reports based on date, time, position and content, but extends this with a more rigorous approach to the comparison of ship identifier information (e.g. name, number or callsign; Carella *et al*., 2017) and incorporates checking for consistency in the timestamp and positioning of the ships using the

position-checking and date/time-checking components of the Met Office Quality Control suite (using an updated version of the procedures described in Rayner *et al*., 2006, see https://github.com/ET-NCMP/MarineQC). Details about the new duplicate identification scheme are provided in Kent *et al*. (2019a). Missing ship identifier information is completed where possible using the ship-tracking method devised by Carella *et al*. (2017).

The NMAT values that remained following application of the duplicate identification procedure were then subjected to further quality control using the Met Office Quality Control tests. The plausibility of the NMAT values relative to climatological normal values was assessed using pentad values calculated from HadNMAT2 over the period 1961-90 that had been disaggregated to a 1°x 1° grid-spacing from the original 5° x 5° resolution. NMAT values in a given grid cell and pentad that deviated more than |10°C| from the climatological mean were removed. NMAT values outside of the range [−80°C, 65°C] were also removed, and buddy-checking was carried out to ensure consistency amongst neighbouring values.

During the late nineteenth century, a warm bias exists in the NMAT observations recorded onboard ships passing through the Suez Canal (Bottomley *et al*., 1990; Kent *et al*., 2013). This bias has been attributed to the practice on certain ships of storing cargo on deck rather than in the hold. In the construction of HadNMAT2, Kent *et al*. (2013) noted that this bias extended farther than the Suez Canal region into the North Atlantic. This bias was resolved by removing NMAT values recorded in Deck 193 (Netherlands Marine) across this region during the period 1880–1893. This procedure was repeated in CLASSnmat.

As a result of recent data-recovery efforts, ICOADS R3.0 contains more observations than R2.5 for certain periods (see Freeman *et al*., 2017). However, due to the stricter duplicate elimination procedure used in the new processing of the data, and the exclusion of data from specialist ships, this is not always reflected in the observation numbers in CLASSnmat. Indeed, for certain periods there are fewer observations in CLASSnmat compared to HadNMAT2 (Figure 1). However, since the new duplicate elimination scheme removes more of the unreliable values, the NMAT observations used in CLASSnmat are expected to be of a higher quality overall. The largest diminution in observations occurs after 1960 where *in extremis*, there are 20% fewer observations in CLASSnmat in terms of annual totals, although generally the reduction is much less. The reduction post-1960 is generally spatially consistent, whereas the increase in observation numbers before this date is often confined to certain ocean basins (Figure 2). During the two World Wars, there is a notable increase in the number of observations used in CLASSnmat. This reflects recent data-recovery efforts. In addition, the change to the correction of data during World War 2 period

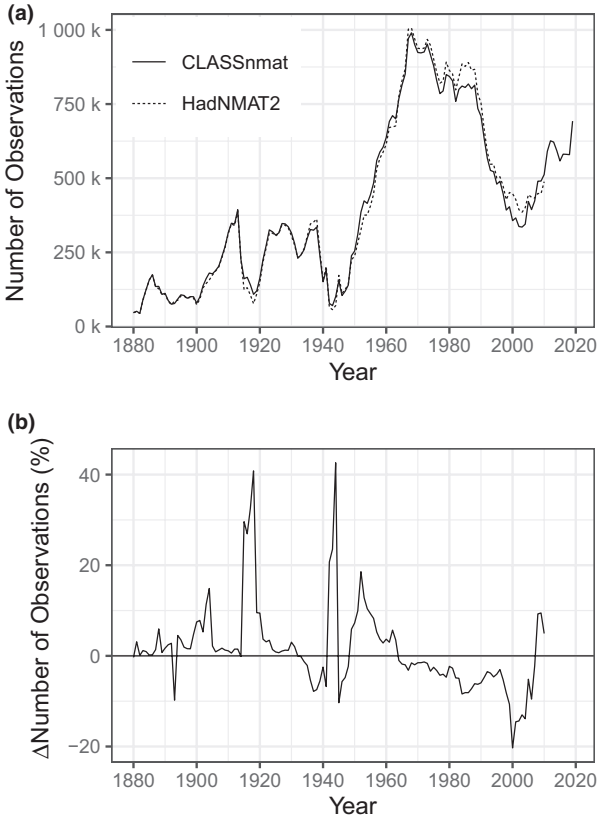

FIGURE 1 Annual total number of observations used globally in the CLASSnmat and HadNMAT2 data sets (a) and the difference in the annual totals (CLASSnmat-HadNMAT2) expressed as a percentage of HadNMAT2 totals (b)

(see Section 2.4) results in more values available for use in CLASSnmat, particularly across the Pacific Ocean. The increase in observations in this region is particularly important given the strong El Niño that occurred during 1939–1942 (Brönnimann, 2009).

2.2 | Adjusting NMAT to a common reference height

Over time, the observing height onboard ships has increased, and to avoid spurious trends in the final NMAT dataset, the temperature readings need to be adjusted to a common height. This requires information about the observation height of the thermometer as well as the lapse rate of the lower layer of the atmosphere.

Heights prior to 1970 were estimated as in HadNMAT2 using globally fixed values (see table 1 in Kent *et al*. (2013)). Information about the height of observation on Voluntary Observing Ships (VOS) for the period after 1970 was taken from the World Meteorological Organization's (WMO) Publication 47 (Pub. 47) (Kent *et al*., 2007). However, it is only after 2002 that this publication explicitly lists the height of the thermometer onboard a given ship. Before this time, the height

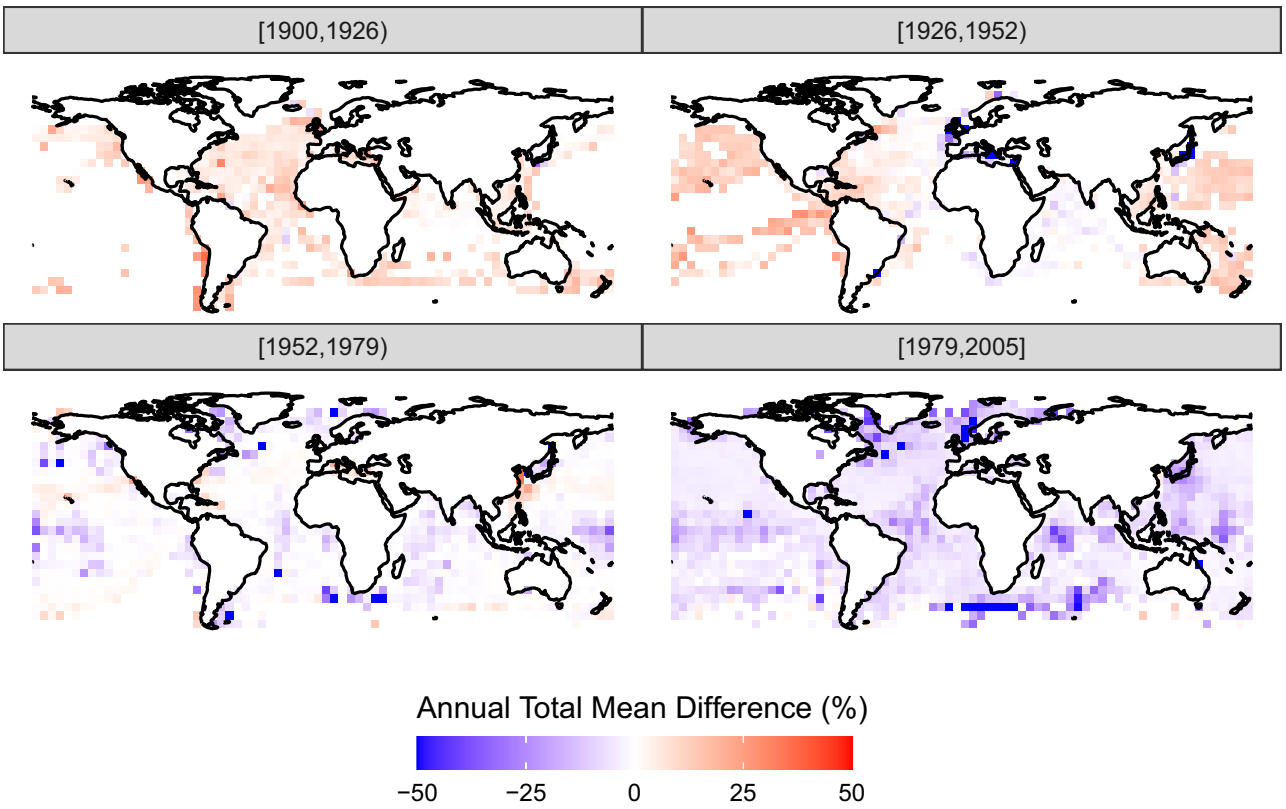

FIGURE 2 Mean differences (CLASSnmat minus HadNMAT2) in the total annual number of observations, calculated for consecutive 27-year periods (chosen to provide convenient segments for plotting) from 1900 to 2005 expressed as a percentage of HadNMAT2 totals

must be estimated from either the 'platform height' or the 'height of the barometer' (see Kent *et al*., 2013). The platform height, the barometer height and thermometer heights were used for the periods 1970–1994, 1995–2001 and 2002–2014, respectively. Rather than using the heights directly from Pub. 47, these heights were taken from the Berry and Kent (2006) version of the publication, which corrects for typographical errors, entry duplication and inconsistencies in the data.

For many ships, the observation height is not listed in Pub. 47 and following HadNMAT2 estimates of these heights were made by averaging known observation heights in each 5° grid cell for each month. These values were then smoothed over time in each grid cell with a triangular filter and these values were used to infill any missing heights for ships in the respective grid cell and month, under the assumption that ships in a given grid cell will be of a similar height.

The NMAT values are adjusted to the common reference height using Monin–Obukhov similarity theory with the parameterizations of Smith (1980) and Smith (1988). As with HadNMAT2, 5° area average, monthly climatological values of wind speed, humidity, sea-air temperature difference and mean sea-level pressure from the NOC surface flux dataset v2.0 (NOCSflux v2.0, Berry and Kent, 2011) were used to determine the stability of the lower atmosphere. In HadNMAT2, the NMAT values were adjusted to a height of 10m. In CLASSnmat, the data are also adjusted to 2m and 20m to allow comparison against other temperature data sets which are generated at different heights above sea level. The adjustments vary over time principally on account of the long-term changes in measurement height (Figure 3) and because of differences in regional sampling (Figure 4).

2.3 | Uncertainty estimates in the MAT height adjustments

The uncertainty estimates for the height corrections are a combination of uncertainties in the stability estimates (σ_s) and the height estimates (σ_h). Each of these uncertainties consists of random ($\sigma_{s,r}$ and $\sigma_{h,r}$) and systematic components ($\sigma_{s,s}$ and $\sigma_{h,s}$). The uncertainty in the height estimates arises from variations in the heights across different ships around a mean value ($\sigma_{h,r}$) and the uncertainty in the mean itself ($\sigma_{h,s}$). The stability-related uncertainties are taken as arising from a combination of uncertainty in the estimates of stability used for individual height adjustments ($\sigma_{s,r}$) and from a pervasive uncertainty in the stability estimates ($\sigma_{s,s}$). Figure 5 demonstrates the magnitude of the height adjustments when the temperature values measured at a range of heights are adjusted to the three CLASSnmat reference heights. The mean adjustment increases as the observation moves away from the reference height. The

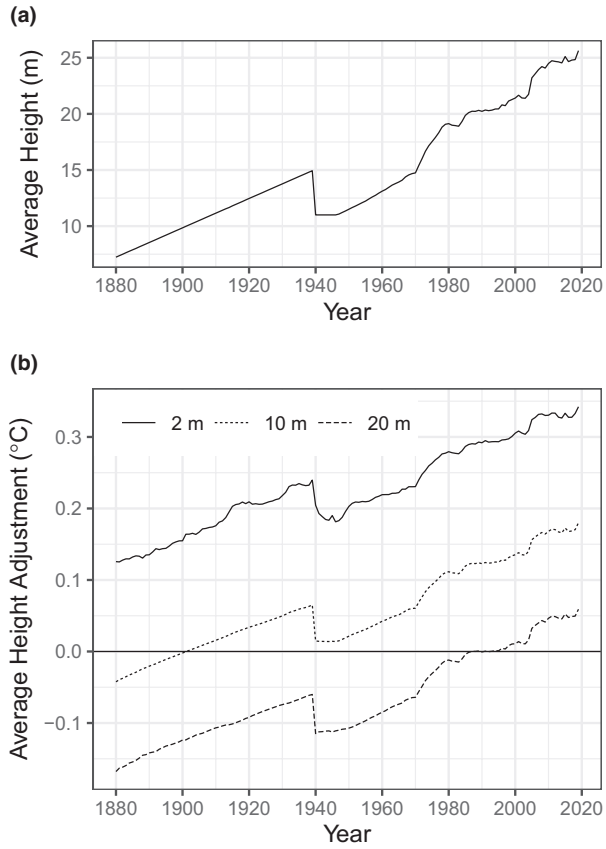

FIGURE 3 The annual global average ship heights (a) and the temperature values adjustments (b) for the three reference levels. The lines in (b) represent the average adjustment to the three reference levels from the respective ship heights for each year on a global average basis

uncertainty in the adjustment also scales as the observation–reference height difference increases, but in practice, this will vary geographically depending on the mean state of the lower boundary stability. The random component in the height estimates (σh,r) was set to zero where the height was known. In the cases where estimated heights were used, σh,r was determined using a kernel density estimation (KDE, e.g. Scott, 1992) of the distribution of known heights in a given 5° grid cell and in a 30-month range centred on the month of interest. KDE is particularly useful in this context as the distribution of heights can be highly irregular and hence the fitting of the non-parametric KDE to the distribution is preferable to any parametric alternatives. The KDE produces a smoothed estimate of the distribution, which alleviates the problem of selecting a suitable bin-width, which was the case in HadNMAT2, where σh,r was estimated by sampling from an empirical histogram.

Taking the probability distribution function (f), the KDE function ($\hat{f}$) for each distribution of heights (x) was calculated as

$$\hat{f}_K(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - x_i\right) \tag{1}$$

where n is the number of heights and K is a subjectively chosen kernel function (Scott and Sain, 2005). The KDE is therefore formed as the sum of kernels centred on each data value and the calculation proceeds through the use of the scaled kernel [$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$]. Although the form of K is a subjective choice, it has less effect on the KDE than the choice of bandwidth (h), which defines the degree of smoothing of a discrete point in the distribution. In this analysis, a Gaussian kernel function was used and h was chosen after the method described by Scott (1992), where for n sampled heights with an inter-quartile range iqr and standard deviation σ, $h = min(\sigma, \mathrm{iqr}/1.34)\cdot 1.06\, n^{\frac{1}{5}}$. This distribution was sampled 5,000 times to generate a range of heights that honours the underlying distribution of known heights subject to the smoothness of the fitted KDE. The heights used in the KDE were initially log-transformed to ensure that all height samples were greater than zero once back-transformed. A temperature adjustment was calculated using this sample of heights, along with the climatological stability parameters for the respective grid cell and climatological month; the standard deviation across these adjustments was taken as $\sigma_{h,r}$.

The systematic height uncertainty component ($\sigma_{h,s}$) was calculated following the technique used in HadNMAT2. A sample of heights was generated from a normal distribution with a mean taken as the grid cell mean and with a fixed standard deviation (σ) depending on the time period under consideration: 2σ = 2m for the period 1880–2014 and 2σ = 4m after 2014. This increased uncertainty for the later period reflects the fact that recorded heights end in 2014 (*c.f.* 2007 in HadNMAT2).

The random contribution to the stability uncertainty ($\sigma_{s,r}$) was estimated from the 1° × 1° longitude/latitude daily air–sea temperature and wind speed values taken from the NOCflux v2 dataset. Since these variables are correlated, a random sample of air–sea temperature differences and wind speed values were extracted from a bivariate KDE using the distribution of high-resolution values in each 5° grid cell over the period 1970–2015. As an extension to the univariate KDE used for the calculation of random height uncertainty, the bivariate KDE takes the form.

$$\hat{f}(x, \mathrm{H}) = \frac{1}{n}\sum_{i=1}^{n} K_{\boldsymbol{H}}\left(x - x_i\right) \tag{2}$$

where in this case, f is a bivariate distribution, x consists of two vectors (the air–sea temperature and wind speed values) and the bandwidth (**H**) is a positive definite, symmetric matrix that acts as a covariance matrix. The bivariate distribution was sampled 5,000 times and stability parameters were calculated

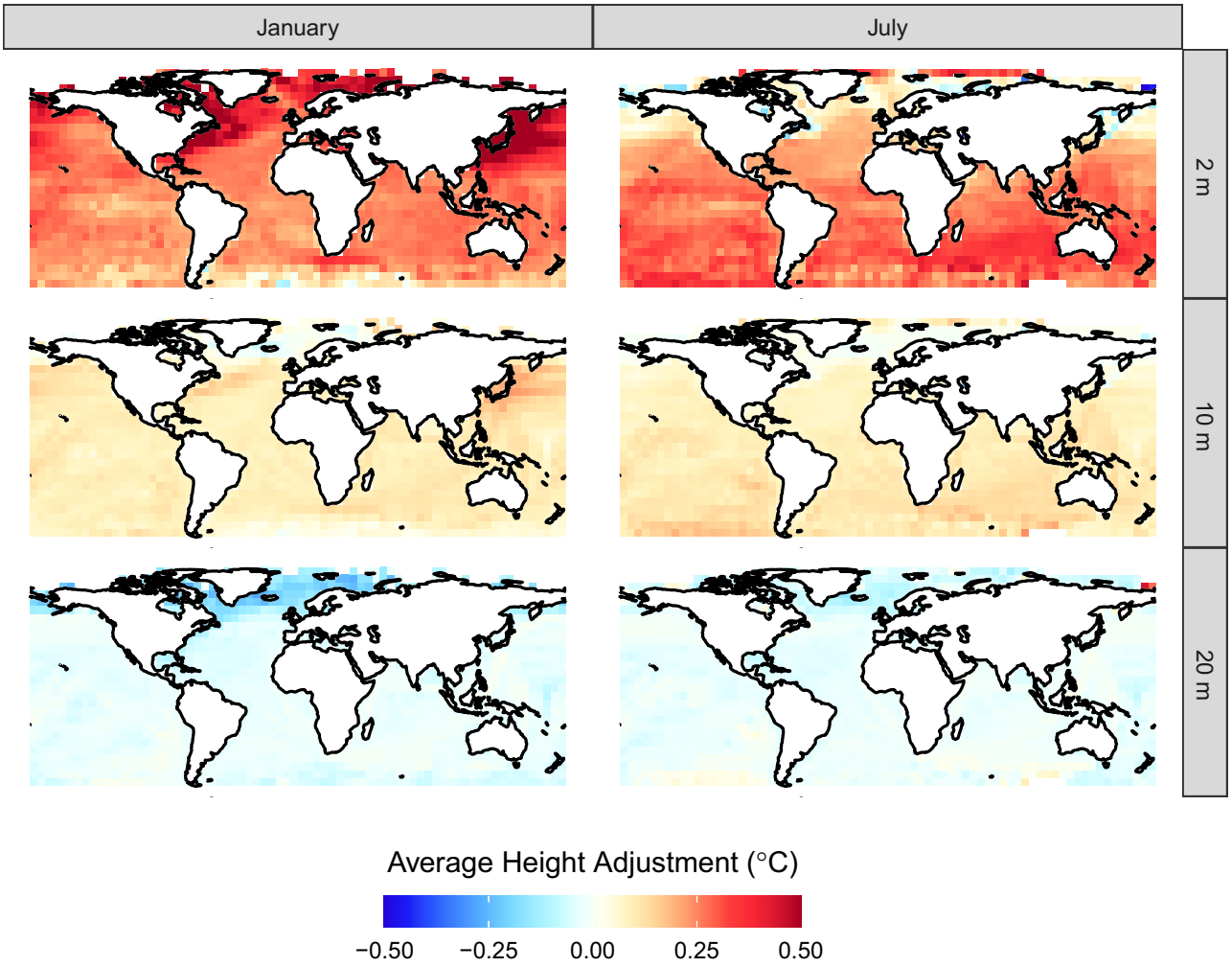

FIGURE 4 The average height adjustments for the three reference levels for January and July over the period 1961–1990. The values are restricted to the range −0.5–0.5°C for plotting purposes

using these absolute-value samples along with mean values of sea-level pressure and relative humidity from the respective grid cell, which were also derived from NOCflux v2.0. Examples of the bivariate distribution and the sampled values are shown in Figure 6 for two grid cells: one grid cell situated over the Grand Banks of Newfoundland (47.5°W; 47.5°N) and one cell situated off the coast of Peru (82.5°W; 12.5°S). This figure demonstrates the flexibility of the KDE in forming a bivariate distribution for these contrasting data series.

The estimate of random stability-related uncertainty ($\sigma_{s,r}$) was calculated from the sample of stability parameters in the respective month and 5° grid cell along with the recorded or estimated observation height to construct a sample of 5,000 height corrections; the standard deviation across these height adjustments is taken as $\sigma_{s,r}$. The systematic stability uncertainty ($\sigma_{s,s}$) was calculated in a similar manner but rather than sampling from the bivariate distribution, a sample of 5,000 values was drawn that had a mean equal to the climatological value of the variables in the respective month and grid cell and a fixed standard deviation for the wind speed, air temperature and SST variables (see table 2 in Kent *et al.* (2013)).

The sample size of 5,000 used to sample the distributions above follows the example of HadNMAT2. To test whether this is an optimal value, considering the significant computing time required when large sample sizes are used, uncertainty estimates were calculated using a range of different sample sizes from 30 to 10,000 for the year 2000. The results from that test are plotted in Figure 7, in which the uncertainty estimates are expressed as proportions of the median across the nine samples for the respective grid cell and month of the year. These results indicate that the uncertainty estimates generally stabilize at around sample sizes of 2000, with 5,000 being a reasonable value.

2.4 | Corrections applied to the data during the Second World War

The NMAT observations recorded during the Second World War contain warm biases on account of different recording

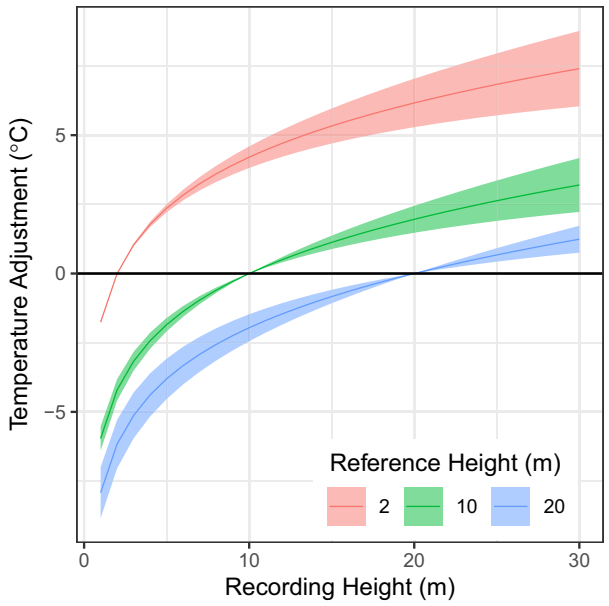

FIGURE 5 The magnitude of temperature adjustment relative to the height at which the temperature measurement was taken. These values are generated by applying the temperature correction for a range of stability parameters (Monin–Obukhov lengths [L] from 100 to 1,000 and the temperature scale [t*] fixed at 1) and recording heights (1–30 m), and correcting to 2, 10 and 20m reference heights. The coloured lines represent the mean ($\bar{x}$) and $\bar{x} \pm 2\sigma$ values

practices used during wartime operations. In HadNMAT2, NMAT values from Deck 195 (U.S. Navy Ship Logs) were excluded due to the presence of a large warm bias relative to the data from other decks but the data from the remaining decks also suggested a warm bias during the night-time period. This artefact has been attributed to the practice of taking measurements of NMAT in overly sheltered locations—possibly indoors—to avoid showing a light on deck during wartime. In HadNMAT2 during the period January 1942 to February 1946, night-time anomalies of MAT were replaced with daytime anomalies that had been adjusted so that the average difference (DMAT-NMAT) over that period matched the average difference over the period 1947–1956.

The air temperature data during the Second World War are mostly contained in three ICOADS decks: 195, 204 and 245. The data from Deck 245 (UK Royal Navy Ships) are relatively newly digitized values and provide additional observations for the period 1938–1947 (Brohan *et al*., 2009; Wilkinson *et al*., 2011); during the Second World War, these data are distributed globally in similar areas to the US Navy (Deck 195) data. However, as with the US Navy data these Royal Navy data also show a warm bias, albeit at a slightly lower magnitude than the US data. In contrast, the data from Deck 204 do not show such a bias. These observations were also recorded onboard UK Royal Navy ships and were extracted from log books into the UK Meteorological Office Main Marine Data Bank (MDB) and were merged into COADS version 1c, released in February 2001. These log books were kept on the senior ship of the squadron and were recorded by qualified Meteorological Officers from the Instructor Branch of the Royal Navy using precision instruments (Rhodes, 1994). A limitation of these data, however, is that they were generally recorded three times a day at 0800, 1200 and 1800 local time, with fewer observations taken during the hours of darkness.

In order to use the data from Decks 195 and 245 in CLASSnmat, a correction was developed for the NMAT readings recorded during the period January 1942 to December 1945. Night-time was defined in this correction as the period between sunset and sunrise, that is, without the one-hour offset that is normally used. This definition was also used in HadNMAT2 for the Second World War correction and gives better results than using the normal night-time definition and probably relates to the watch-time used during this period relative to the sunset/sunrise times.

A calibration for NMAT from each ship in decks 195 and 245 was derived by comparing the data against the values from all other decks. This was determined by first calculating anomalies relative to the period 1961–1990 for each observation. Individual observation anomalies excluding Decks 195 and 245 were gridded at 5° monthly resolution. The biases in individual observation anomalies from Decks 195 and 245 were calculated taking the difference of each observation anomaly in those decks from the gridded anomalies from the other decks. The average of these values for each ship track was taken as the correction value for a given track and the standard deviation as the uncertainty in the correction (σ_{ww}). This correction assumes that the correction was also valid for locations where there were no observations from the data from other decks, that is, a consistent recording practice was used for the duration of a ship's voyage. The uncertainty is therefore considered to be correlated within the observations from a given track. Tracks were only used where there were more than 20 values available for calibration, where the absolute mean difference was less than 5°C and standard deviation was less than 5°C, under the assumption that data with a larger mean or spread were unreliable; the values from these ships were not used. This correction significantly reduces the warm bias evident in the NMAT values during the Second World War (Figure 8).

2.5 | The gridding method and uncertainty calculation

The calculation of the monthly mean 5° grid cell values and the associated estimation of uncertainty takes the approach that has been widely employed in other marine data sets (Berry and Kent (2017) and which builds on the optimal averaging technique described by Kagan *et al*. (1997). The NMAT observations are extremely heterogenous, with observations clustered in space and time, and the residual errors from the data may be correlated across observations made

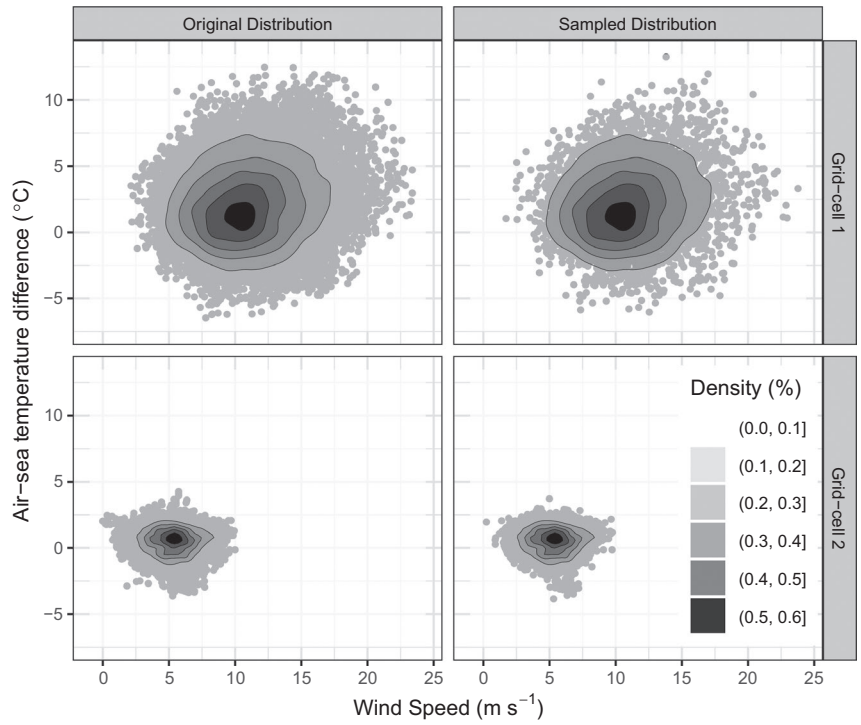

FIGURE 6 Examples of the distribution of air–sea temperature and wind speed distributions from two grid cells in January. Grid cell 1 is situated at 47.5°W, 47.5°N (Grand Banks of Newfoundland) and grid cell 2 at 82.5°W, 12.5°S (off Peru).The contours are repeated across the horizontal but the grey dots represent either the original distribution or those sampled from the KDE

onboard the same ship (Kent *et al*., 2019b). As a result, simply taking the arithmetic mean of all values (t_{obs}) in a given grid cell and month would likely lead to a value that is biased towards the point locations with the greatest number of points and towards the ships contributing the most observations. A weighted mean that takes into account the correlation between observations is therefore calculated:

$$\tilde{T} = \sum_{i=1}^{n} w_i t_{obs,i} \tag{3}$$

where the weights are given by w_i. The expected mean-squared error in the calculated areal mean ($\tilde{T}$) compared to the 'true' areal mean ($\overline{T}$) is taken as:

$$E\{(\tilde{T}-\overline{T})^2\} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \left(\Delta_i\Delta_j + \alpha_i\alpha_j + \beta_i\beta_j\right) \tag{4}$$

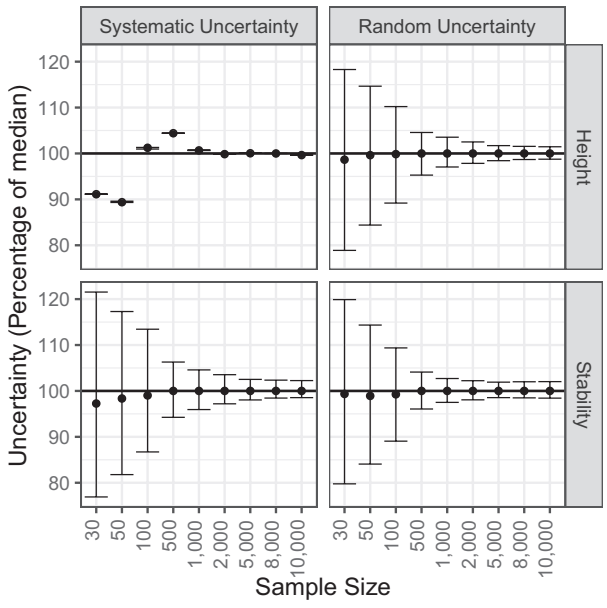

FIGURE 7 The effect of different sample sizes on the uncertainty estimates. Uncertainty values calculated for the year 2000 are expressed as a proportion of the average across the nine sample sizes for the respective grid cell and month of the year. The error ranges indicate the median and 10th and 90th percentiles across all grid cells/months

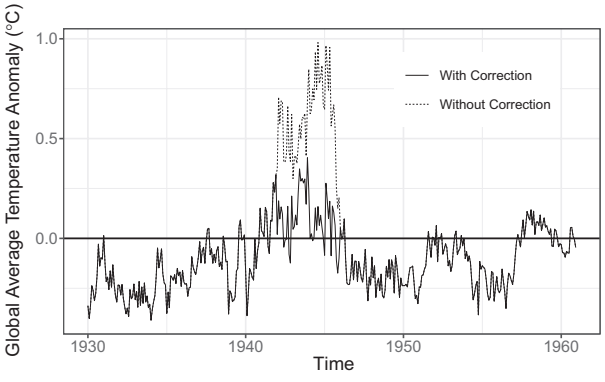

FIGURE 8 Global average monthly anomalies (weighted by the cosine of latitude and relative to 1961–1990 averages) from January 1930 to December 1960 derived using the gridded data both before and after application of the Second World War correction

where three uncertainty types are defined for observations i and j: uncorrelated (α), systematic (β) and sampling uncertainties (Δ). In the case of the uncorrelated uncertainty, a covariance matrix α is constructed as $\alpha_{ii}=\sigma^2_{r,i}$ and $\alpha_{ij}=0$ as a reflection of the random nature of the errors between observations. In contrast, the matrix β needs to reflect the perfect correlation across observations from the same ship track and the zero relationship otherwise. Hence taking $\boldsymbol{x}=\begin{matrix}\sigma_{st,1}\\ \vdots \\ \sigma_{st,n}\end{matrix}$, $B=\boldsymbol{x}^T\boldsymbol{x}$ and $A_{ij}=1$ where the systematic uncertainties are from the same ship track and zero otherwise, $\beta=AB$. While the new ship-tracking technique assigns more ships to a track than was the case with HadNMAT2, there are cases where the tracking fails to attribute a ship to a particular track. In these cases, a covariance value of $A_{ij}=0.25$ is used, which reflects this ambiguity in the covariance of the two observations.

Assuming that the individual uncertainty terms are independent of each other, the systematic height correction (stability and height-related) uncertainties for each observation were combined in quadrature with the estimated uncertainty in the World War 2 correction (σ_{ww}, where applicable) and a systematic observational error (0.6°C) taken from Berry and Kent (2017), giving $\sigma_{st}=\sqrt{\sigma^2_{s,s}+\sigma^2_{h,s}+\sigma^2_{ww}+0.6^2}$. In a similar manner for the random component but with a larger observational error estimate (0.9°C), $\sigma_r=\sqrt{\sigma^2_{s,r}+\sigma^2_{h,r}+0.9^2}$; in this case $\sigma_{ww}=0$.

In a manner that is the basis of the widely used kriging equations, the covariance term for sampling uncertainty (Δ) is calculated assuming an exponential correlation in the observations over space and time as $r=c\cdot\exp\left(-\left[\left(\frac{x}{x_0}\right)^2+\left(\frac{t}{t_0}\right)^2\right]\right)$, where c is the sill, which quantifies the variance in NMAT at the characteristic length scale (x_0), x is the geographical distance between the two points, t is the absolute time difference between the observations and t_0 is the characteristic temporal scale. The covariance matrix is constructed as $\Delta_{ii}=c$ along the diagonal and $\Delta_{ij}=r$ for off-diagonal elements. The value of x_0 was fixed at 200km, which is considered a reasonable value given the average spatial autocorrelation of MAT, and $t_0=2$ days. The value of the sill (c) was calculated for each 5° grid cell from the variance across the 1° daily mean air temperature values from the ERA5 reanalysis (Copernicus Climate Change Service, 2017) in each month over the period 1981–2010; the use of this climatological period results from the data only currently being available from 1979 onwards. These values provide a measure of the climatological variance of MAT in each grid cell (Figure 9). Although the reanalysis data set assimilates the same ship observations, alongside many other in situ data and satellite observations, the data were used because they provide spatially complete fields of the climatological temperature variance in each 5° grid cell.

Expressing equation 4 in matrix form and adding the Lagrange parameter (μ) to ensure that the weights sum to one we have.

$$E\{(\tilde{T}-\overline{T})\}=\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w}+\mu\mathbf{1}\boldsymbol{w}-\mu \tag{5}$$

Taking the first derivative of this expression and setting to zero, we derive

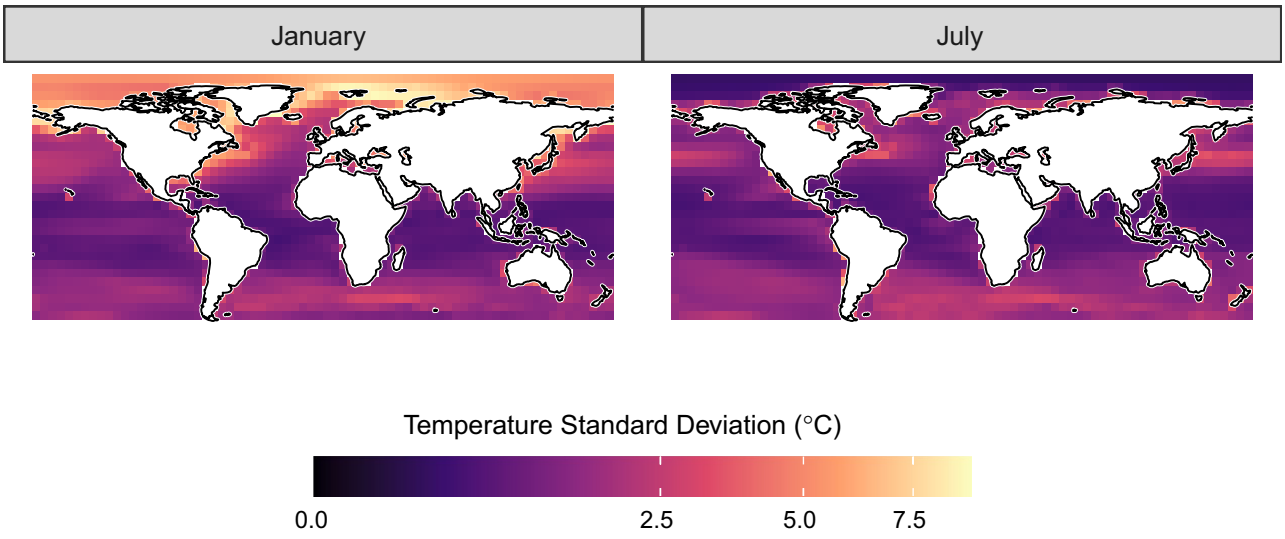

FIGURE 9 Standard deviation of 2m temperature derived from the ERA5 reanalysis data set for January and July. A square-root transformation is used on the colour scale in this figure

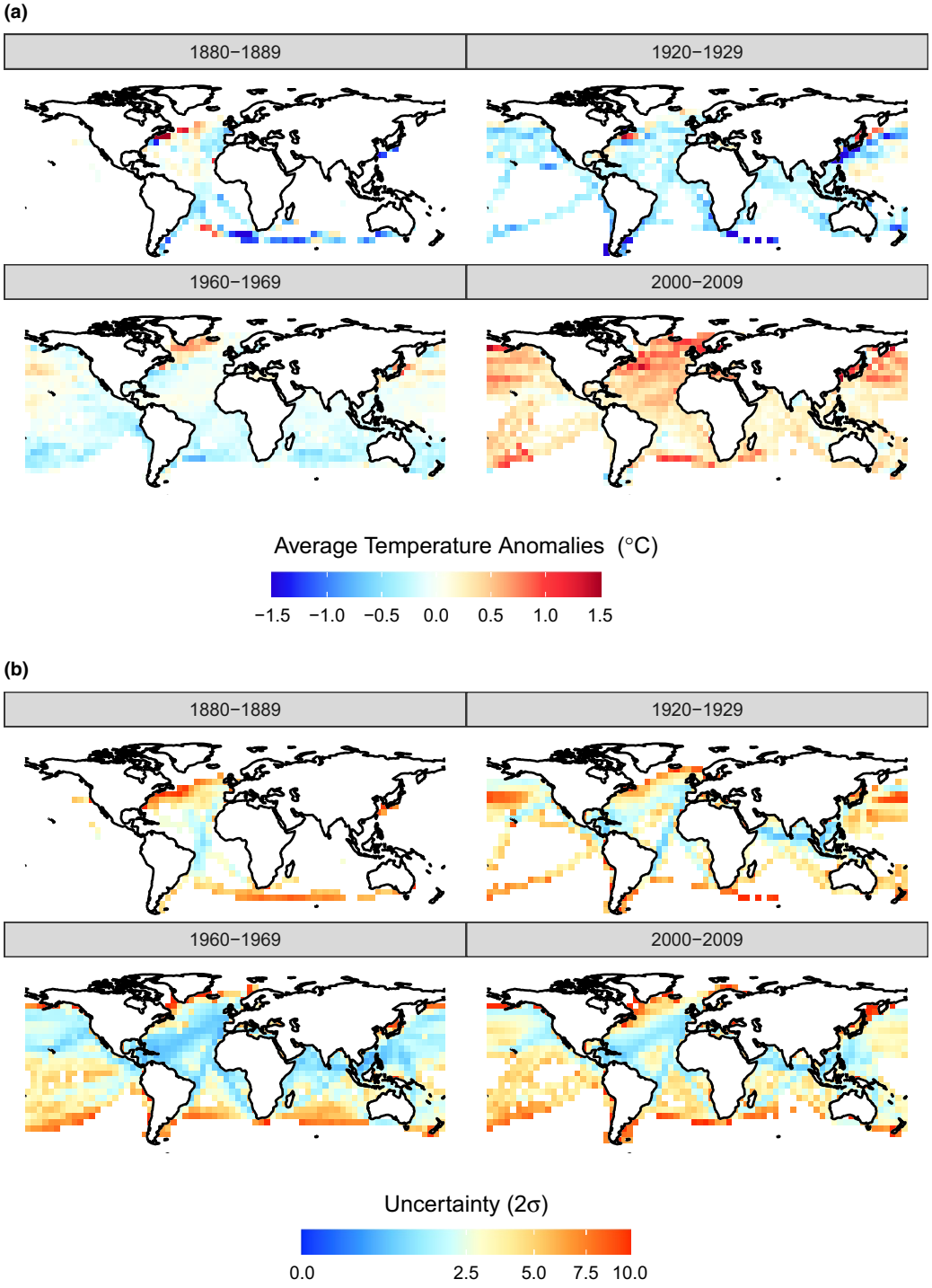

FIGURE 10 CLASSnmat grid cell average annual anomalies (with respect to the 1961-90 base period) calculated for four decades (a) and the associated uncertainties (b). Note the square-root scaling used in the legend in (b). grid cells with fewer than 96 complete months in the decade are marked as missing

$$\begin{bmatrix} C_{1,1} & \cdots & C_{1,n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n,1} & \cdots & C_{n,n} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

and hence the weights are obtained as.

$$\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{1,1} & \cdots & C_{1,n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n,1} & \cdots & C_{n,n} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (7)$$

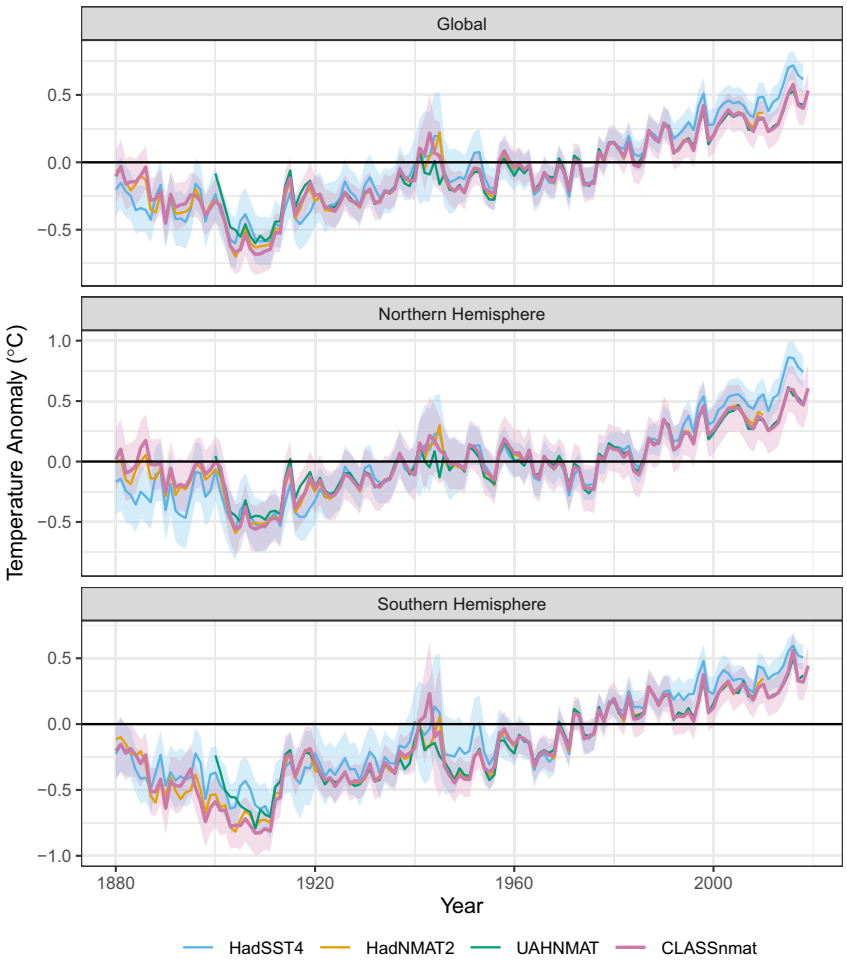

FIGURE 11 Annual average global anomaly time series (with respect to the 1961-1990 base period) calculated from CLASSnmat, HadNMAT2, HadSST4 and UAHNMATv1 across three regions. The pink (blue) shading indicates the 2 σ uncertainty range of the CLASSnmat (HadSST4) series

The final gridded values are taken as the weighted average of the NMAT values in each grid cell/month (Equation 3) with variance $\boldsymbol{w}^T\boldsymbol{C}\boldsymbol{w}$.

3 | COMPARISON AGAINST OTHER GRIDDED DATASETS

As indicated in Figure 10, the uncertainty in the gridded values is clearly related to the sampling density, with greatly reduced uncertainty in the gridded fields in areas and periods with a relatively high density of ship data. While on the whole the level of uncertainty decreases over time, with increased data coverage, the uncertainty across the Southern Ocean remains high throughout the period. The uncertainty values are calculated in that figure assuming that the correlated uncertainty components are perfectly correlated over time in each grid cell but that the sampling and uncorrelated components are not correlated.

In terms of global and hemispheric averages, the interannual and longer-timescale variability in anomalies (relative to the 1961–1990 base period) from the three NMAT data sets (CLASSnmat, HadNMAT2 and UAHNMATv1) are broadly comparable, although the different adjustments made during World War 2 lead to noticeable differences between the data sets, particularly UAHNMATv1, which is cooler than the other series (Figures 11 and 12). The NMAT data sets all show a differential trend relative to HadSST4 (Kennedy *et al*., 2019), which is particularly apparent after ca. 1990.

The uncertainty ranges in the large-scale averages in CLASSnmat and HadSST4 are of the same order of magnitude over the 1880–2018 period, which would be expected given the similar processing methods and origin of the data (Figure 11). The uncertainty ranges are slightly larger overall in CLASSnmat, and this is principally due to larger correlated uncertainty and because of the additional climatology uncertainty component (see Section 4 below). The increased correlated uncertainty arises from the higher proportion of ships with ship IDs in CLASSnmat; this is most apparent in the Northern Hemisphere.

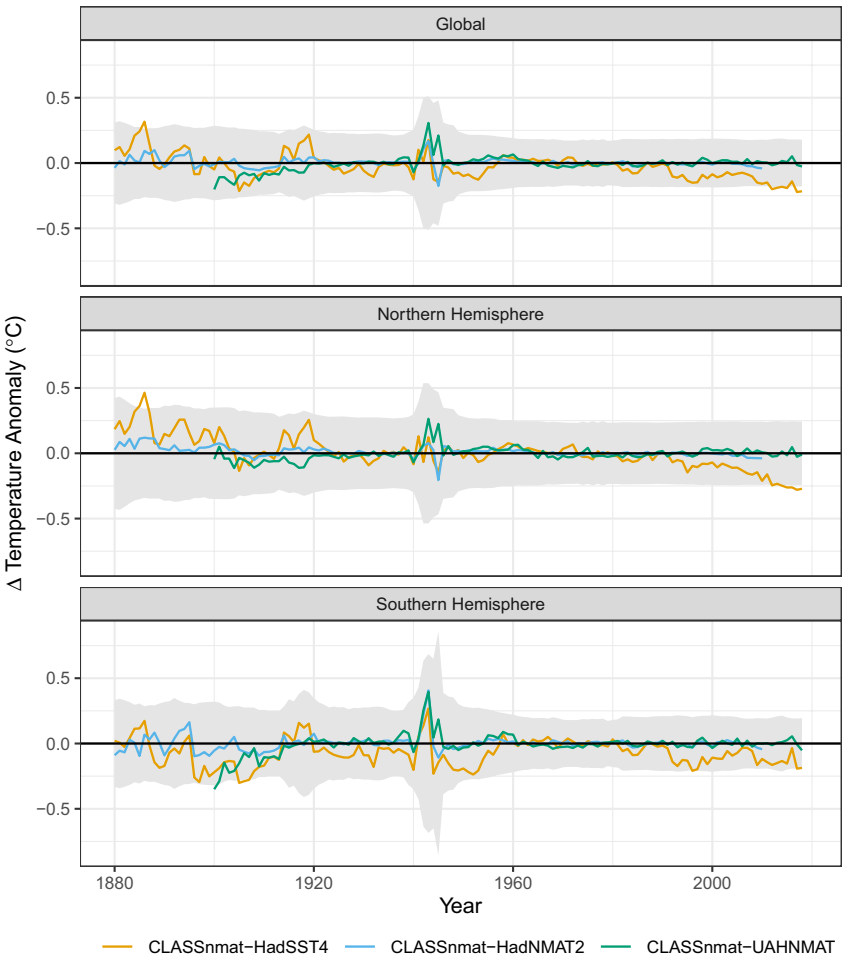

FIGURE 12 Differences in the CLASSnmat annual anomaly series (with respect to the 1961-1990 base period) relative to the HadNMAT2, HadSST4 and UAHNMATv1 series. The grey shading indicates the combined (2 σ) uncertainty from HadSST4 and CLASSnmat, in which the correlated uncertainty components are assumed to perfectly correlated and the other components are uncorrelated

ACKNOWLEDGMENTS

This work was funded by NERC under the CLASS programme (NE/R015953/1) and under the GloSAT project (NE/S015647/2). John Kennedy was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. The duplicate identification procedure was developed under the NERC HOSTACE project (NE/J020788/1) with additional funding from the Copernicus Climate Change Service (Contract C3S_311a_Lot2). The ERA5 reanalysis data were obtained from the Copernicus Climate Change Service Data Store (https://cds.climate.copernicus.eu/); the JRA-55 reanalysis data from the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory; and the HadSST4 and HatNMAT2 data were from the Met Office Hadley Centre observation data sets repository (https://www.metoffice.gov.uk/hadobs/). We thank Robert Junod (UAH) for providing the UAHNMATv1 data. The ICOADS data are gratefully acknowledged (Research Data Archive *et al.*, 2016). We thank the anonymous reviewers for their constructive comments. CLASSnmat is produced using the JASMIN/CEDA high-performance computing cluster (https://www.ceda.ac.uk/services/jasmin/).

DATA AVAILABILITY

The CLASSnmat data are available from the Centre for Environmental Data Analysis (CEDA) archive. Each file contains the gridded NMAT values, the corresponding uncertainty values and the anomalies relative to three different base periods (1961–1990, 1971–2000 and 1981–2010). In addition, monthly and annual averages of NMAT anomalies (with respect to the three climatological periods) are provided in CLASSnmat for four regions: globally, for the northern and southern hemispheres and for the tropics (30°S-30°N). The construction of the anomaly fields and large-scale averages and accompanying uncertainty values are described below.

CALCULATION OF ANOMALY FIELDS

Anomaly values have been calculated at each grid cell by subtracting the climatological monthly average at each grid cell. These climate normals were obtained by calculating the means across each month of the year over the given year range. The values (y) were then fitted with a second-order

harmonic model using least-squares regression to produce the climate normal in order to produce a smoothed annual cycle and reduce the effects of sampling variability (see Wilks, 2019):

$$y(t) = A_0 + A_1\cos(\omega t) + B_1\sin(\omega t) + A_2\cos(\omega t) + B_2\sin(\omega t) \quad (8)$$

where $\frac{2\pi}{\omega} = 12$ months.

Uncertainty in the climatology values is estimated by generating 200 random draws from the annual cycle model at each grid cell. The realizations are conditional upon the model covariates and as such provide a way of taking into account the temporal correlation in the climatology uncertainty which is important in the calculation of the annual regional averages (described below in Section 4.2). The climatology uncertainty in the monthly gridded data is calculated as the standard deviation across the 200 realizations for the respective month of the year (σ_{norm}). The 1-σ uncertainties are given as σ_{mean} for the absolute NMAT values and $\sigma_{\text{anom}} = \sqrt{\sigma^2_{\text{mean}} + \sigma^2_{\text{norm}}}$ for the anomaly values, which takes the combined effect of uncertainty in the grid cell NMAT values and the uncertainty in the climate normals. Separate files are provided that contain the climate normals for each base period and the 200 climatology realizations.

LARGE-SCALE MONTHLY AND ANNUAL NMAT AVERAGES

The large-scale averages are calculated using area-weighted, gridded 10m NMAT anomalies and are accompanied by uncertainty estimates, which have been calculated following the approach described for the HadSST3 and HadSST4 data sets (Kennedy *et al*., 2011; Kennedy *et al*., 2019).

The uncertainty values for the monthly averages make use of a covariance matrix $C_{(p,q)}$, which takes into account the systematic uncertainty (σ_{st}) that results from a given ship traversing grid cells p and q:

$$\boldsymbol{C}_{(p,q)} = \sum_{h=1}^{m} \boldsymbol{w}(p)_h \boldsymbol{w}(q)_h \boldsymbol{\sigma}_{st\,h}(p) \boldsymbol{\sigma}_{st\,h}(q)_h \quad (9)$$

where the weights (w) are as described in Section 2.5, and m is the number of individual ships in a given month. The random, sampling and climatology uncertainties, taken as being uncorrelated across grid cells (in addition to the systematic uncertainty), form the diagonal elements of the matrix as:

$$\boldsymbol{C}_{(p,p)} = \boldsymbol{\sigma}^2_{st}(p) + \boldsymbol{\sigma}^2_{r}(p) + \boldsymbol{\sigma}^2_{sp}(p) + \boldsymbol{\sigma}^2_{clim}(p) \quad (10)$$

where $\boldsymbol{\sigma}^2_{clim}$ is the variance across the 200 climatology ensemble members, from which the monthly grid cell means have been subtracted. The monthly uncertainty estimate for the area average is $\sigma^2_{\text{month}} = \boldsymbol{aCa}^{\boldsymbol{T}}$ where **a** are the grid cell area weights. To this is added a coverage uncertainty component that estimates the uncertainty related to the fact that there are missing grid cells in the area average (see Kennedy *et al*., 2011).

The coverage uncertainty is calculated using the 2m monthly average temperature values from the JRA-55 reanalysis data set (Kobayashi *et al*., 2015), which have been regridded using bilinear interpolation to match the resolution of CLASSnmat. This reanalysis data set was chosen because it covers a relatively long time period (1958–2019). The monthly anomaly values from JRA-55 were first area-weighted averaged over the respective region using all grid cells for the given month of the year; this was repeated for the respective months over the 1958–2019 period. The grid cells were then sub-sampled to match the coverage of the CLASSnmat data for the target month. The standard deviation across the full and reduced samples over the period 1958–2019 provides the estimate of coverage uncertainty for the given month, and this was added in quadrature to the other uncertainty components.

The calculation of the full space-time covariance matrix is computationally prohibitive for the annual average anomalies and we therefore used the simplification described by Kennedy *et al*. (2011, taken from their equations 23–25). To that estimate is added the climatology uncertainty (calculated across the climatology ensemble) and the coverage uncertainty (calculated as for the monthly averages except that annual averaged from the JRA-55 data are used). As with HadSST4, correlated uncertainty values are calculated only using data that have ship IDs. This is also the case with the monthly uncertainty estimates, and while this will lead to an underestimation in the correlated uncertainty, this will be less in CLASSnmat than HadSST4 as there are fewer observations with missing IDs due to the use of the new ship-tracking method (see Section 2.1).

ORCID

Richard C. Cornes https://orcid.org/0000-0002-7688-4485
Elizabeth.C. Kent https://orcid.org/0000-0002-6209-4247
David.I. Berry https://orcid.org/0000-0002-3862-3479
John J. Kennedy https://orcid.org/0000-0002-6841-7289

REFERENCES

Berry, D.I. and Kent, E.C. (2006). WMO publication 47: Consistency checking and gapfilling, version 1.0, documentation for metadata data-base. 77 pp. Available at http://eprints.soton.ac.uk/341419/

Berry, D.I. and Kent, E.C. (2009) A new air-sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bulletin of the American Meteorological Society*, 90(5), 645–656. https://doi.org/10.1175/2008bams2639.1

Berry, D.I. and Kent, E.C. (2011) Air-sea fluxes from ICOADS: the construction of a new gridded dataset with uncertainty estimates. *International Journal of Climatology*, 31(7), 987–1001. https://doi.org/10.1002/joc.2059

Berry, D.I. and Kent, E.C. (2017) Assessing the health of thein situ-global surface marine climate observing system. *International Journal of Climatology*, 37(5), 2248–2259. https://doi.org/10.1002/joc.4914

Berry, D.I., Kent, E.C. and Taylor, P.K. (2004) An analytical model of heating errors in marine air temperatures from ships. *Journal of Atmospheric and Oceanic Technology*, 21(8), 1198–1215. https://doi.org/10.1175/1520-0426(2004)021%3C1198:AAMOHE%3E2.0.CO;2

Bottomley, M., Folland, C., Hsiung, J., Newell, R. and Parker, D. (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.

Brohan, P., Allan, R., Freeman, J.E., Waple, A.M., Wheeler, D., Wilkinson, C. *et al*. (2009) Marine observations of old weather. *Bulletin of the American Meteorological Society*, 90(2), 219–230. https://doi.org/10.1175/2008bams2522.1

Brönnimann, S. (2009) Early twentieth-century warming. *Nature Geoscience*, 2(11), 735–736. https://doi.org/10.1038/ngeo670

Carella, G., Kent, E.C. and Berry, D.I. (2017) A probabilistic approach to ship voyage reconstruction in ICOADS. *International Journal of Climatology*, 37(5), 2233–2247. https://doi.org/10.1002/joc.4492

Copernicus Climate Change Service (2017) ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), https://cds.climate.copernicus.eu/

Freeman, E., Woodruff, S.D., Worley, S.J., Lubker, S.J., Kent, E.C., Angel, W.E. *et al*. (2017) ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 37(5), 2211–2232. https://doi.org/10.1002/joc.4775

Junod, R.A. and Christy, J.R. (2019) A new compilation of globally gridded night-time marine air temperatures: the UAHNMATv1 dataset. *International Journal of Climatology*, 40(5), 2609–2623. https://doi.org/10.1002/joc.6354

Kagan, R.L.Z., Gandin, L.S. and Smith, T.M. (1997). *Averaging of Meteorological Fields*. Dordrecht ; Boston: Kluwer Academic Publishers. ix, 279 p.

Kennedy, J.J., Rayner, N.A., Smith, R.O., Parker, D.E. and Saunby, M. (2011) Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *Journal of Geophysical Research*, 116(D14), D14104. https://doi.org/10.1029/2010jd015220

Kennedy, J.J., Rayner, N.A., Atkinson, C.P. and Killick, R.E. (2019) An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set. *Journal of Geophysical Research: Atmospheres*, 124(14), 7719–7763. https://doi.org/10.1029/2018jd029867

Kent, E.C., Woodruff, S.D. and Berry, D.I. (2007) Metadata from WMO Publication No. 47 and an assessment of voluntary observing ship observation heights in ICOADS. *Journal of Atmospheric and Oceanic Technology*, 24(2), 214–234. https://doi.org/10.1175/jtech1949.1

Kent, E.C., Rayner, N.A., Berry, D.I., Saunby, M., Moat, B.I., Kennedy, J.J. *et al*. (2013) Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *Journal of Geophysical Research: Atmospheres*, 118(3), 1281–1298. https://doi.org/10.1002/jgrd.50152

Kent, E.C., Rayner, N.A., Berry, D.I., Eastman, R., Grigorieva, V.G., Huang, B. *et al*. (2019) Observing requirements for long-term climate records at the ocean surface. *Frontiers in Marine Science*, 6, 28, Article Unsp 441. https://doi.org/10.3389/fmars.2019.00441.

Kent, E.C., Berry, D.I. and Perez Gonzalez, I. (2019a) Copernicus Climate Change Service Global Land and Marine Observations Database - Documentation for marine duplicate identification and linking of platform identifiers (ECMWF Copernicus Report, Issue. https://figshare.com/account/articles/11342390

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H. *et al*. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Parker, D.E., Hackett, R.B. and Jackson, M. (1997). Marine air temperatures: The New MOHMAT4 datasets. Hadley Centre Internal Note No. 82. 18pp + figs.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P. *et al*. (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14), https://doi.org/10.1029/2002jd002670

Rayner, N.A., Brohan, P., Parker, D.E., Folland, C.K., Kennedy, J.J., Vanicek, M. *et al*. (2006) Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *Journal of Climate*, 19(3), 446–469. https://doi.org/10.1175/jcli3637.1

Research Data Archive, Computational and Information Systems Laboratory, National Center for Atmospheric Research, University Corporation for Atmospheric Research, P. S. D., & Earth System Research Laboratory. (2016). *International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 3, Individual Observations Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory*. Boulder, CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6ZS2TR3

Rhodes, M.H. (1994) Additional Marine Log Book Data, 1936-48. [UK Met Office report contracted (together with two accompanying reports covering 1856-99 and 1911-20) to examine the archives "to see whether all data had been transferred to the Marine Data Bank at the Meteorological Office."]. https://icoads.noaa.gov/reclaim/pdf/rhodes_1936-48.pdf

Scott, D.W. (1992) *Multivariate density estimation: theory, practice and visualization*. New York, Chichester: Wiley. xii, 317 p.

Scott, D.W. and Sain, S.R. (2005). 9 - Multidimensional density estimation. In Rao, C.R., Wegman, E.J. and Solka, J.L. (Eds.), *Handbook of Statistics* (Vol. 24, pp. 229–261). Elsevier. https://doi.org/10.1016/S0169-7161(04)24009-3

Slutz, R.J., Lubker, S.J., Hiscox, J.D., Woodruff, S.D., Jenne, R.L., Joseph, R.H.*et al*. (1985). Comprehensive ocean-atmosphere dataset: Release 1. NTIS PB86-105723. 268.

Smith, S.D. (1980) Wind stress and heat flux over the ocean in gale force winds. *Journal of Physical Oceanography*, 10(5), 709–726. https://doi.org/10.1175/1520-0485(1980)010<0709:Wsahfo>2.0.Co;2

Smith, S.D. (1988) Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature. *Journal of Geophysical Research*, 93(C12), https://doi.org/10.1029/JC093iC12p15467

Spencer, J.W. (1989) Comments on the astronomical almanac's algorithm for approximate solar position (1950–2050). *Solar Energy*, 42(4), 1950–2050. https://doi.org/10.1016/0038-092X(89)90039-X

Wilkinson, C., Woodruff, S.D., Brohan, P., Claesson, S., Freeman, E., Koek, F. *et al*. (2011) Recovery of logbooks and international marine data: the RECLAIM project. *International Journal of Climatology*, 31(7), 968–979. https://doi.org/10.1002/joc.2102

Wilks, D.S. (2019) *Statistical Methods in the Atmospheric Sciences*. Amsterdam: Elsevier.

How to cite this article: Cornes RC, Kent E, Berry D, Kennedy JJ. CLASSnmat: A global night marine air temperature data set, 1880–2019. *Geosci Data J*. 2020;00:1–15. https://doi.org/10.1002/gdj3.100