

# Quantifying Daytime Heating Biases in Marine Air Temperature Observations from Ships

THOMAS E. CROPPER<sup>1</sup>,<sup>a</sup> DAVID I. BERRY<sup>1</sup>,<sup>a,b</sup> RICHARD C. CORNES<sup>1</sup>,<sup>a</sup> AND ELIZABETH C. KENT<sup>1</sup>,<sup>a</sup>

<sup>a</sup> National Oceanography Centre, Southampton, United Kingdom

<sup>b</sup> World Meteorological Organization, Geneva, Switzerland

(Manuscript received 22 July 2022, in final form 6 January 2023, accepted 11 January 2023)

**ABSTRACT:** Marine air temperatures recorded on ships during the daytime are known to be biased warm on average due to energy storage by the superstructure of the vessels. This makes unadjusted daytime observations unsuitable for many applications including for the monitoring of long-term temperature change over the oceans. In this paper a physics-based approach is used to estimate this heating bias in ship observations from ICOADS. Under this approach, empirically determined coefficients represent the energy transfer terms of a heat budget model that quantifies the heating bias and is applied as a function of cloud cover and the relative wind speed over individual ships. The coefficients for each ship are derived from the anomalous diurnal heating relative to nighttime air temperature. Model coefficients, cloud cover, and relative wind speed are then used to estimate the heating bias ship by ship and generate nighttime-equivalent time series. A variety of methodological approaches were tested. Application of this method enables the inclusion of some daytime observations in climate records based on marine air temperatures, allowing an earlier start date and giving an increase in spatial coverage compared to existing records that exclude daytime observations.

**SIGNIFICANCE STATEMENT:** Currently, the longest available record of air temperature over the oceans starts in 1880. We present an approach that enables observations of air temperatures over the oceans to be used in the creation of long-term climate records that are presently excluded. We do this by estimating the biases inherent in daytime temperature reports from ships, and adjust for these biases by implementing a numerical heat-budget model. The adjustment can be applied to the variety of ship types present in observational archives. The resulting adjusted temperatures can be used to create a more spatially complete record over the oceans, that extends further back in time, potentially into the late eighteenth century.

**KEYWORDS:** Climate; Diurnal effects; Surface temperature; Data quality control; In situ atmospheric observations; Ship observations

## 1. Background and motivation

Marine air temperature (MAT) observations from ships form a long-term climate record used to construct gridded data products as either the principal data source (Berry and Kent 2009, 2011; Kent et al. 2013; Cornes et al. 2020; Junod and Christy 2020) or for bias adjustment of sea surface temperature (SST) products (Huang et al. 2017; Kennedy et al. 2019). These gridded products only use MAT observed during nighttime (NMAT) to exclude data affected by solar heating of the instrument and local ship environment during daytime (DMAT). Using only NMAT approximately halves the number of available observations and limits the temporal extent of any MAT-based dataset as early observations were often only recorded during the daytime (Fig. 1a). For example, two recently published NMAT datasets begin in 1880 (CLASSnmat; Cornes et al. 2020) and 1900 (UAHNMAT; Junod and Christy 2020). Extending the MAT record further back in time requires

bias adjustment of DMAT, and if this adjustment can be determined accurately the sampling and coverage of MAT will be improved throughout the record.

Global mean surface temperature (GMST) anomaly datasets, combining observations over land, ice, and ocean, have used SST in lieu of MAT for their ocean component (Lensen et al. 2019; Morice et al. 2021; Huang et al. 2020), including in the sixth Intergovernmental Panel on Climate Change Assessment Report (Gulev et al. 2021). GMST is used instead of global surface air temperature (GSAT) for three main reasons: there are more (all-hours) SST observations than NMAT; quantification of SST measurement bias and uncertainties is more mature than for MAT (Kennedy et al. 2019); and the belief that SST anomalies are more reliable than MAT at large spatial scales (Kent and Kennedy 2021). It was also asserted that large-scale anomalies of SST and MAT display similar variability and trends (Huang et al. 2017), although this is increasingly being questioned (e.g., Cowtan et al. 2015; Richardson et al. 2016; Rubino et al. 2020). Here we demonstrate a method to estimate the daytime heating biases in MAT observations on a ship-by-ship basis that can be applied throughout the observed record. The ultimate goal is to use these adjusted data to create a GSAT record based on air temperature over land, ice, and ocean. This will facilitate comparison of the observed surface temperature record

<sup>1</sup> Denotes content that is immediately available upon publication as open access.

Corresponding author: Thomas E. Cropper, thomas.cropper@noc.ac.uk

DOI: 10.1175/JTECH-D-22-0080.1

© 2023 American Meteorological Society. This published article is licensed under the terms of a Creative Commons Attribution 4.0 International (CC BY 4.0) License



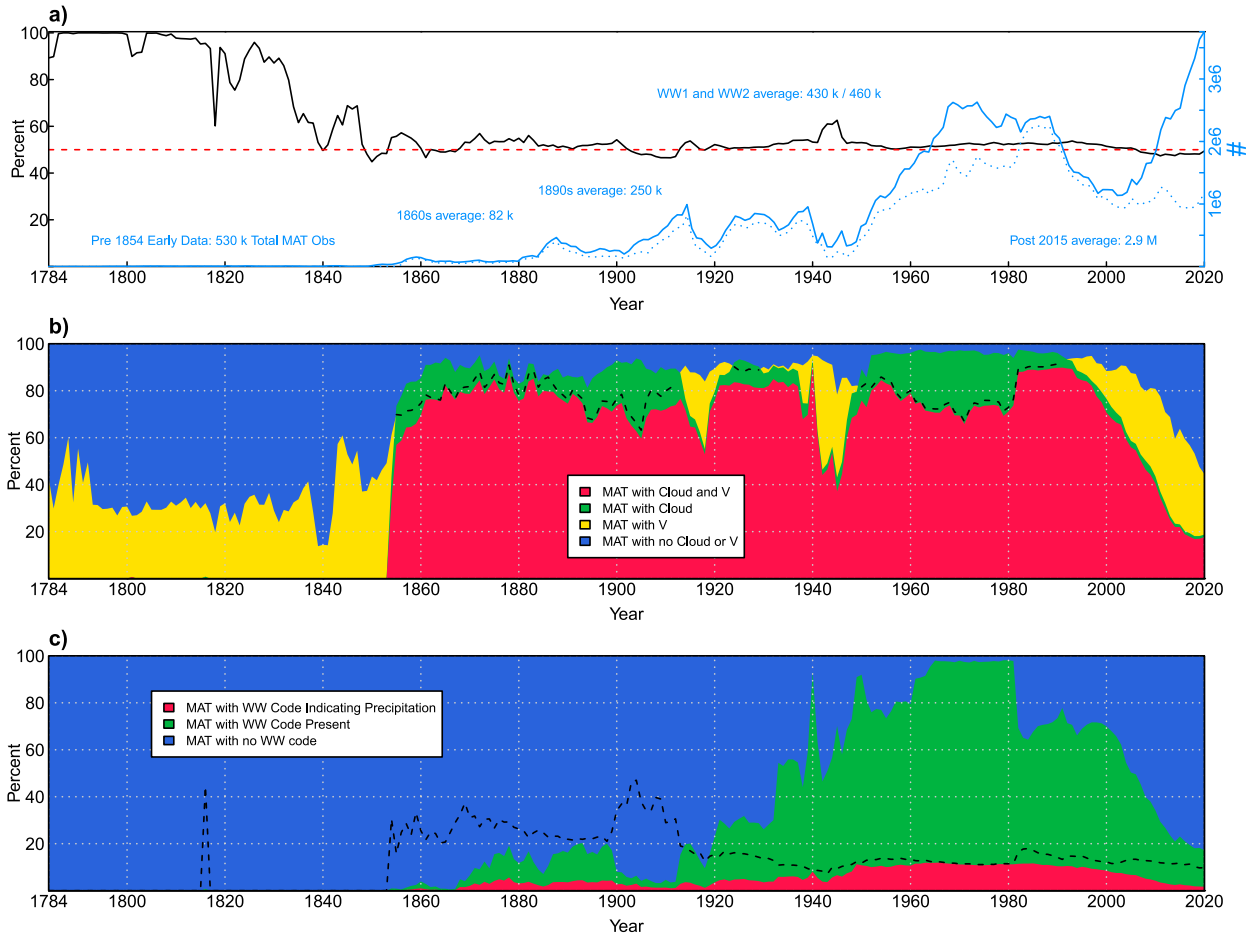


FIG. 1. Sampling characteristics of MAT observations from ship reports in ICOADS (Freeman et al. 2017). (a) The 1784–2020 percentage of MAT observations recorded annually during daytime (DMAT, black line, left-hand axis), the red dashed line indicates 50% daytime observations. The solid blue line (right-hand axis) is the annual total number of MAT observations, and the dotted blue line is the number of MAT observations associated with a ship track of 12 or more reports, with diurnal sampling, and including only observations with associated cloud and relative wind speed ( $V$ ) observations. Free text comments indicate the annual average number of MAT observations for select periods (or the total amount for 1784–1853). (b) Stacked plot of the percent of MAT observations with a corresponding cloud and/or  $V$  value. The red area indicates reports with both cloud and  $V$ ; the blue area indicates reports with neither. Reports with either cloud or  $V$ , but not both, are indicated in green and yellow, respectively. The dotted line indicates MAT with  $V$  but without cloud when the green color overrides. When yellow is visible, lack of cloud information is the major constraint on applying the heating bias model introduced in section 2a; when green is visible, lack of  $V$  is the constraint. (c) Stacked plot of the percentage of MAT observations with an associated present weather code (WW; green) and with WW code indicating precipitation (red). The dashed line shows the percentage of extant WW indicating precipitation.

with the output of climate models (Jones 2020), which most straightforwardly provide estimates of GSAT rather than GMST.

## 2. Methods

### a. The Berry et al. (2004) model

Berry et al. (2004, hereafter BKT) developed a model to quantify heating-related biases in MAT, accounting for the energy accumulation and release by the superstructure of ships. The BKT model was developed and tested using temperature values recorded on board the Ocean Weather Ship *Cumulus* during 1988 and later used to examine exposure-

related bias on 17 ships extracted from the VOSCLIM database (Berry and Kent 2005). In the construction of the NOC Surface Flux and Marine Meteorological Dataset (Berry and Kent 2009, 2011) the BKT model was used to adjust the MAT observations obtained from the International Comprehensive Ocean–Atmosphere Dataset (ICOADS; Freeman et al. 2017) for the period 1973–2014. However, in order to simplify the calculations in that analysis, a fixed annual set of coefficients was applied across all ships. Here we develop coefficients ship by ship to give an adjustment for heating bias that reflects the characteristics of a particular ship.

We define several measures of temperature in Eqs. (1)–(3), which are illustrated in Fig. 2:  $T_{\text{air}}$  is the true air temperature;  $T_{\text{ship}}$  is the measured air temperature;  $T_{\text{nt}}$  is the background

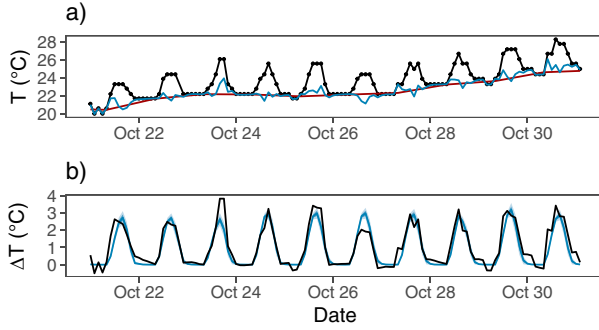


FIG. 2. (a)  $T_{\text{ship}}$  (black line, circles show individual observations),  $T_{\text{adj}}$  ( $T_{\text{ship}} - \Delta T_{\text{BKT}}$ ) (blue), and  $T_{\text{nt}}$  (red) for the ship *Raphael* during October 1884 and (b)  $\Delta T_{\text{diur}}$  (black) and  $\Delta T_{\text{BKT}}$  (blue, dark shading corresponds to  $\pm 1$  standard deviation of the  $\Delta T_{\text{BKT}}$  value from the 60-member ensemble and light shading corresponds to  $\pm 2$  standard deviations).

nighttime air temperature (see section 2b);  $\Delta T_{\text{err}}$  is the change in measured temperature due to ship heating;  $\Delta T_{\text{diur}}$  is an estimate of  $\Delta T_{\text{err}}$ ;  $\Delta T_{\text{BKT}}$  is the estimate of the temperature difference from the BKT model; and  $T_{\text{adj}}$  is the measured air temperature adjusted using the BKT model:

$$T_{\text{ship}} = T_{\text{air}} + \Delta T_{\text{err}} \approx T_{\text{air}} + \Delta T_{\text{diur}}, \quad (1)$$

$$\Delta T_{\text{diur}} = T_{\text{ship}} - T_{\text{nt}}, \quad (2)$$

$$T_{\text{adj}} = T_{\text{ship}} - \Delta T_{\text{BKT}}. \quad (3)$$

The BKT model relates  $T_{\text{air}}$  and  $T_{\text{ship}}$  (both measured in Kelvin) to the heat exchange, Eq. (4):

$$mc \frac{d(T_{\text{ship}})}{dt} = Q_{\text{SW}} + Q_{\text{LW}} + Q_{\text{Conv}} + Q_{\text{Cond}}. \quad (4)$$

In this equation  $m$  is the mass (kg) and  $c$  the specific heat capacity ( $\text{J kg}^{-1} \text{K}^{-1}$ ) of the sensor environment (that part of the ship that affects the measurement),  $t$  is time (s),  $Q_{\text{SW}}$  is the shortwave irradiance [Eq. (5)], and  $Q_{\text{Conv}}$  and  $Q_{\text{Cond}}$  are the rates of heat transfer between the ship and the atmosphere through convection and conduction [all in  $\text{W m}^{-2}$ , Eq. (6)]:

$$Q_{\text{SW}} = \alpha_s A_s R_{\text{top}} (a_i + b_i \sin \theta) \sin \theta, \quad (5)$$

$$Q_{\text{Conv}} + Q_{\text{Cond}} = (T_{\text{air}} - T_{\text{ship}}) A_c (h_\mu + h_o). \quad (6)$$

Here,  $\alpha_s$  is the solar absorptivity of the sensor environment,  $A_s$  is the surface area normal to the direction of the incoming direct solar radiation ( $\text{m}^2$ ),  $R_{\text{top}}$  is the solar radiation at the top of the atmosphere (we use  $1368 \text{ W m}^{-2}$ ),  $\theta$  is the solar elevation,  $a_i$  and  $b_i$  are cloud-cover-dependent coefficients (index  $i$  indicates categories of total cloud cover quantities by oktas; from Dobson and Smith 1988),  $h_\mu$  and  $h_o$  are the convective and conductive heat transfer coefficients ( $\text{W m}^{-2} \text{K}^{-1}$ ), and  $A_c$  is the surface area of the sensor environment ( $\text{m}^2$ ). Following Berry and Kent (2005) we exclude the small thermal

TABLE 1. Empirical coefficients;  $V$  is relative wind speed ( $\text{m s}^{-1}$ ).

Coefficient	Definition	Min	Max
$x_1$	$A_s \alpha_s / mc$	0.0001	0.1
$x_3 V^{x_4}$	$(A_c / mc) h_\mu$		
$x_3$		0.0001	10
$x_4$		-2	2
$x_5$	$(A_c / mc) h_o$	0.0001	10

energy transfer component ( $Q_{\text{LW}}$ ), shown by BKT to account for a maximum  $\sim 3\%$  of the estimated heating bias. Assuming  $d(T_{\text{air}})/dt \approx 0$ , Eq. (4) becomes

$$mc \frac{d\Delta T_{\text{err}}}{dt} + (h_\mu + h_o) A_c \Delta T_{\text{err}} = \alpha_s A_s R_{\text{top}} (a_i + b_i \sin \theta) \sin \theta. \quad (7)$$

Substituting the coefficients given in Table 1 into Eq. (7) gives

$$\frac{d\Delta T_{\text{err}}}{dt} + (x_3 V^{x_4} + x_5) \Delta T_{\text{err}} = x_1 [R_{\text{top}} (a_i + b_i \sin \theta) \sin \theta], \quad (8)$$

where  $V$  is relative wind speed ( $\text{m s}^{-1}$ ) and the empirical coefficients  $x_{1,3,4,5}$  represent terms of the energy budget model (Table 2). We have redefined the coefficients  $x_3$  and  $x_5$  to incorporate and exclude  $x_2$  (used in the original BKT definition), so cooling depends on  $(x_3 V^{x_4} + x_5)$  and heating on  $x_1$ . Expansion of the  $\sin \theta$  terms in Eq. (8) and further substitutions (Table 2) gives

$$\frac{d\Delta T_{\text{err}}}{dt} + h_1 (\Delta T_{\text{err}}) = h_2 + h_3 \cos(\phi) + h_4 \cos^2(\phi), \quad (9)$$

where  $\phi$  is hour angle in radians. The solution to Eq. (9), gives the value of the heating error at any time during daylight hours (for a full description of the solution, see BKT):

$$\begin{aligned} \Delta T_{\text{err,day}} = & \frac{h_2}{h_1} + \frac{h_3 \alpha}{h_1^2 + \alpha^2} + \left[ \sin(\phi) + \frac{h_1}{\alpha} \cos(\phi) \right] \\ & + \frac{4\alpha^2 h_4}{h_1^2 + 4\alpha^2} \left[ \frac{\cos(\phi) \sin(\phi)}{2\alpha} + \frac{h_1 \cos^2(\phi)}{4\alpha^2} + \frac{1}{2h_1} \right] \\ & + k_{\text{int}} \exp(-h_1 t), \end{aligned} \quad (10)$$

where  $\alpha$  is  $-\pi/12$ . The integrating factor  $k_{\text{int}}$  can be determined assuming the sensor environment is in equilibrium at sunrise ( $d\Delta T_{\text{diur}}/dt \sim 0$ ) such that

TABLE 2. Substitutions used in solving the BKT model; dec is the solar declination and the  $k$  terms use latitude in radians.

Parameter	Substitution
$h_1$	$x_3 V^{x_4} + x_5$
$h_2$	$x_1 R_{\text{top}} (ak_1 + ak_1^2)$
$h_3$	$x_1 R_{\text{top}} (ak_2 + 2bk_1 k_2)$
$h_4$	$x_1 R_{\text{top}} (bk_2^2)$
$k_1$	$\sin[\text{lat} \sin(\text{dec})]$
$k_2$	$\cos[\text{lat} \cos(\text{dec})]$

$$k_{\text{int}} = - \left\{ \frac{h_2}{h_1} + \frac{4\alpha^2 h_4}{4\alpha^2 + h_1^2} \left[ \frac{\cos(\phi_{\text{sr}}) \sin(\phi_{\text{sr}})}{2\alpha} + \frac{h_1 \cos^2(\phi_{\text{sr}})}{4\alpha^2} \right. \right. \\ \left. \left. + \frac{1}{2h_1} \right] + \frac{\alpha h_3}{\alpha^2 + h_1^2} \left[ \sin(\phi_{\text{sr}}) + \frac{h_1}{\alpha} \cos(\phi_{\text{sr}}) \right] \right\}, \quad (11)$$

where  $\phi_{\text{sr}}$  is the hour angle at the time of sunrise. At night the heating error is

$$\Delta T_{\text{err,night}} = (T_{\text{err,ss}}) \exp(-h_1 \tau_{\text{ss}}), \quad (12)$$

where  $\tau_{\text{ss}}$  is the time elapsed since sunset.

Using Eq. (10) (daytime) or Eq. (12) (nighttime)  $\Delta T_{\text{BKT}}$  at any location and time can be calculated using the coefficients  $x_{1,3,4,5}$  along with cloud cover and  $V$ .

### b. Estimation of the temperature error due to ship heating

Both the true diurnal variation of  $T_{\text{air}}$  and the heating error are poorly known. *BKT* estimated the heating error ( $\Delta T_{\text{err}}$ ) in two ways: as the MAT anomaly from the local midnight to sunrise mean and as the  $T_{\text{ship}} - \text{SST}$  difference. The former is likely to overestimate  $\Delta T_{\text{err}}$  as it incorporates the true diurnal cycle of  $T_{\text{air}}$ , while the latter is likely to be an underestimate.

Given the difficulty of making an adjustment that accounts for the real diurnal cycle of  $T_{\text{air}}$ , a pragmatic approach was taken to estimate  $\Delta T_{\text{diur}}$  [Eq. (2)], and hence the adjustment [ $\Delta T_{\text{BKT}}$ , from Eqs. (10) and (12)] relative to an estimated background nighttime temperature ( $T_{\text{nt}}$ ). First, an estimate of the expected diurnal SST anomaly associated with every value of  $T_{\text{ship}}$  was calculated as a function of cloud cover and wind speed following *Morak-Bozzo et al. (2016)* and subtracted from  $T_{\text{ship}}$ . Nighttime values were then calculated using the normal definition of 1 h after sunset to 1 h after sunrise (*Bottomley et al. 1990*). These nighttime averages were assigned to the time of each sunrise and were linearly interpolated over the 24-h period for each ship ( $T_{\text{nt}}$ ). This approach allows the construction of climate records from a combination of adjusted all-hours MAT with unadjusted NMAT.

### c. Solar parameterization

The *BKT* model uses cloud-cover-dependent coefficients to estimate solar radiation based on location, date, and time. *BKT* used coefficients from the *Dobson and Smith (1988)* okta model, which were derived from a limited geographical region. Using the same okta model as *Dobson and Smith (1988)*, we generate a set of updated coefficients. To do this, we used data from the Surface Solar Radiation dataset–Heliosat version 2.1 (*Pfeifroth et al. 2019*), which covers most of the Atlantic (65°S–65°N, 65°W–65°E). We collocate the 30-min sampling interval of satellite instantaneous incoming solar radiation values with ICOADS cloud observations for the period 1983–2017 and use this information to generate updated okta model coefficients ([https://git.noc.ac.uk/glosat\\_tc/okta\\_model](https://git.noc.ac.uk/glosat_tc/okta_model)). The resulting coefficients produce a less peaked solar cycle than the original *Dobson and Smith* coefficients and reduce the overall RMSE of estimated to satellite incoming solar radiation by ~10% for data not included in the fit. The  $a_i$  and  $b_i$  terms

[Eq. (5)] become  $a_{i,\text{lat}}$  and  $b_{i,\text{lat}}$  as specific coefficients are available for 10° latitudinal bands. Presently the *BKT* model implementation requires the solar parameterization to be in the same form as the okta model, precluding the use of, for example, the parameterization of *Aleksandrova et al. (2007)*.

Other than adjusting nine oktas to eight when the ICOADS present weather code indicates precipitation (*Aleksandrova et al. 2018*), we do not make any adjustments to the ICOADS cloud record. Considering long-term temporal trends, biases likely remain due to heterogeneous recording practices and conversions across the diversity of ICOADS source data. For example, cloud observations pre-1949 (when cloud recording changed from tenths to oktas) may be biased low due to being double adjusted if the original observation was in oktas (*Gulev and Aleksandrova 2020*).

### d. Optimization

The optimization selects values for the  $x$  coefficients that minimize the difference between  $\Delta T_{\text{diur}}$  and  $\Delta T_{\text{BKT}}$  using Eqs. (10) and (12), and using several different cost functions. Coefficients are derived for selected individual ships, but could also be applied across a group of ships thought to have similar  $\Delta T_{\text{diur}}$  characteristics.

The solution uses the L-BFGS-B (*Byrd et al. 1995*) solver in R (an option in the *optim* function; *R Core Team 2019*) with lower and upper coefficient limits from *Table 1*. We minimize six different cost functions to evaluate the *BKT* model solutions. Each cost function tests different aspects of the goodness of fit and the spread across the cost functions is wider giving more realistic estimates of fit uncertainty:

- 1) The residual root-mean-square error (RMSE),  $\text{RMSE} = \sqrt{(1/n) \sum_{k=1}^n (\Delta T_{\text{diur}_k} - \Delta T_{\text{BKT}_k})^2}$ , the RMSE gives the simplest measure of the fit.
- 2) Weighted RMSE ( $\text{RMSE}_w$ ) only using MAT observation times between 3 and 8 h after sunrise.  $\text{RMSE}_w$  gives weight only to hours where  $\Delta T_{\text{BKT}}$  values are expected to be largest.
- 3)  $\text{RMSE}_{V_2}$ , where the RMSE is calculated from bin means of data in  $2 \text{ m s}^{-1} V$  and local-hour intervals.
- 4)  $\text{RMSE}_{V_5}$ , as  $\text{RMSE}_{V_2}$  but with  $5 \text{ m s}^{-1} V$  and 2-hourly intervals. Both  $\text{RMSE}_{V_2}$  and  $\text{RMSE}_{V_5}$  are designed to greater weight the importance of minimizing the  $(\Delta T_{\text{diur}} - \Delta T_{\text{BKT}})$  residual through the day and across values of  $V$ .
- 5)  $\text{RMSE}_{\text{DW}} = (1 - \lambda)\text{RMSE} + \lambda(|\text{DW} - 2|)$ , where DW is the Durbin–Watson statistic and  $\lambda$  is a scaling factor that we set to 0.3.  $\text{RMSE}_{\text{DW}}$  is used to down-weight solutions where the residual displays autocorrelation.
- 6)  $\text{RMSE}_{\text{KS}} = (1 - \lambda)\text{RMSE} + \lambda(\text{KS})$ , where KS is the Kolmogorov–Smirnov statistic. This cost function gives greater weight to solutions where the cumulative sums of daytime values of  $\Delta T_{\text{diur}}$  and  $\Delta T_{\text{BKT}}$  are small.

An ensemble of these cost functions is used to test different aspects of the structure of the residual ( $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$ ) to ensure a reasonable fit throughout the day and across all cloud-cover and relative wind speed combinations. Avoiding unphysical starting coefficient combinations improves efficiency and helps to

TABLE 3. Sixteen ships selected from ICOADS to illustrate the results of fitting the BKT model. Deck refers to the original source data collection in ICOADS. Metadata contain information that could be readily obtained via an Internet search of the original call sign or name of the ship. The ship U.S. Navy 12388 samples at 0800, 1200, and 2000 local hour, a common feature of currently available WWII-era ships.

Ship	Year	Hourly sampling frequency	Deck	Metadata
USS <i>Constitution</i>	1854	2	721	Sail, wood
USS <i>Despatch</i>	1858	2	701	Screw steamer
USS <i>Merrimac</i>	1858	2	721	Steam frigate
<i>Mary</i>	1884	2	704	Unknown
<i>Panay of Salem</i>	1884	2	704	Sail
<i>Raphael</i>	1884	2	704	Sail, wood
<i>Chosen Maru</i>	1916	4	762	Cargo, steel, screw steamer
<i>Kanagawa Maru</i>	1916	4	762, 706	Passenger, steel, screw steamer
U.S. Navy 12388	1942	3 times daily	195	Unknown U.S. naval ship
U.S. hourlies –129	1955	1	116, 117	Unknown
Merchant Marine 0805	1955	3	116, 117	Unknown
<i>Kajtum</i>	1975	6	927	Cargo ship
<i>Westfalen</i>	1995	3	926, 892, 888	Passenger/cargo ship
<i>Cape Azalea</i>	2014	1	992	Bulk carrier
<i>Polar Resolution</i>	2014	1	992	Oil tanker
<i>Alliance St. Louis</i>	2020	1	798, 992	Vehicle carrier

avoid local minima so we use a pool of  $\sim 350$  precalculated sets of starting coefficients to initialize the fit. For each ship, we randomly select 10 sets of starting coefficients and 5 subsets of 70% of available days. This gives an ensemble of 300 sets of coefficients (10 starting values, 5 data subsets, and 6 cost functions), and any convergence failures are rerun until there are 50 sets of coefficients per cost function. Unless otherwise stated, hereafter the  $\Delta T_{\text{BKT}}$  value is the ensemble mean taken from 60 realizations of the  $\Delta T_{\text{BKT}}$  using the 10 best-fit time series from each of the 6 cost functions.

### 3. Results

#### a. Fitting to individual ships

To illustrate the application of the BKT model, we show results from 16 ships covering different time periods, sampling frequencies, and original input sources (Table 3). The data for these ships were obtained from the ICOADS (Freeman et al. 2017) archive: release 3.0.0 up to 2014 and release 3.0.1 thereafter. Quality checking has been applied to the data prior to model fitting (appendix). The reports from these 16 ships contain all of the variables required to fit the adjustment model, and all have reported data over at least 150 days. Collectively, these ships provide a global sample of data between 60°S and 60°N, with 64% of observations in the tropics (30°N–30°S), 26% in the Northern Hemisphere, and 10% in the Southern Hemisphere. Longitudinally, there are 31% of observations in the Atlantic Ocean, 29% in the Pacific Ocean, 18% in the Indian Ocean, 17% in the South China Sea and adjoining gulfs/seas, with 3% in the Mediterranean Sea and remainder (2%) of observations from minor ocean basins.

Figure 2 shows the diurnal adjustment for the ship *Raphael* during October 1884. Figure 2a shows  $T_{\text{ship}}$ ,  $T_{\text{nt}}$ , and  $T_{\text{adj}}$ . Figure 2b shows the estimates of  $\Delta T_{\text{diur}}$  and  $\Delta T_{\text{BKT}}$ .

Figure 3 shows the mean  $\Delta T_{\text{diur}}$ ,  $\Delta T_{\text{BKT}}$ , and residuals ( $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$ ) using the best-fit set of coefficients for each cost function (i.e., six lines) for the ship *Mary* (Figs. 3a–d), split across local hour of the day, cloud cover, 2 m  $\text{s}^{-1}$  intervals of  $V$ , and 10° latitude bins. Following BKT we use a target accuracy of  $\pm 0.2^\circ\text{C}$ . Figure 3 shows that across the input parameters of the BKT model (time/position, cloud cover, and  $V$ ), the heating bias is removed, with bin-mean residuals that are generally within  $\pm 0.2^\circ\text{C}$ , and the bin-mean local-hour average residuals are always within the  $\pm 0.2^\circ\text{C}$  target. However, for this ship the BKT model appears to underadjust for clear skies (0 okta) and a  $V$  of 22–24  $\text{m s}^{-1}$ , although these bins are poorly sampled (14 observations for 0 oktas and 45 and 16 observations for the 22 and 24  $\text{m s}^{-1}$  bins, respectively).

Figures 4a–c display the  $\Delta T_{\text{diur}}$ ,  $\Delta T_{\text{BKT}}$ , and residuals ( $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$ ) across all 16 ships as a function of the number of hours since sunrise.  $\Delta T_{\text{diur}}$  can be  $< 0^\circ\text{C}$ , as MAT values close to sunrise will be cooler than the nighttime mean MAT.  $\Delta T_{\text{BKT}}$  is always above  $0^\circ\text{C}$ , and this is reflected in the negative residuals for hours 0–1 and  $\geq 18$ . Aside from the 28–34, 46, and 50  $\text{m s}^{-1}$  wind speed bins and 60°N latitude bin, the bin-mean residuals (Figs. 4c–f) are all within  $\pm 0.2^\circ\text{C}$ . The pattern of a relative  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  underadjustment for hours 3–5 and 9–14 (Fig. 4c) appears consistent regardless of whether a single cost function is used or a different sample of ships is selected (not shown). Possible causes are inaccurate estimates of solar radiation [Eq. (5)] or systematic errors in our estimate of  $\Delta T_{\text{diur}}$ .

The mean overall  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  residual for each individual ship is always within  $\pm 0.2^\circ\text{C}$ , with 11 out of 16 ships within  $\pm 0.05^\circ\text{C}$ . The largest residual ( $0.14^\circ\text{C}$ ) is found for the U.S. Navy 12388 ship. The WWII period is one of the more difficult periods to apply the BKT model correction, due to the limited number of observations from which determine  $T_{\text{nt}}$ , as well as the occurrence of and a warm bias in nighttime observations over 1942–46 (Cornes et al. 2020).

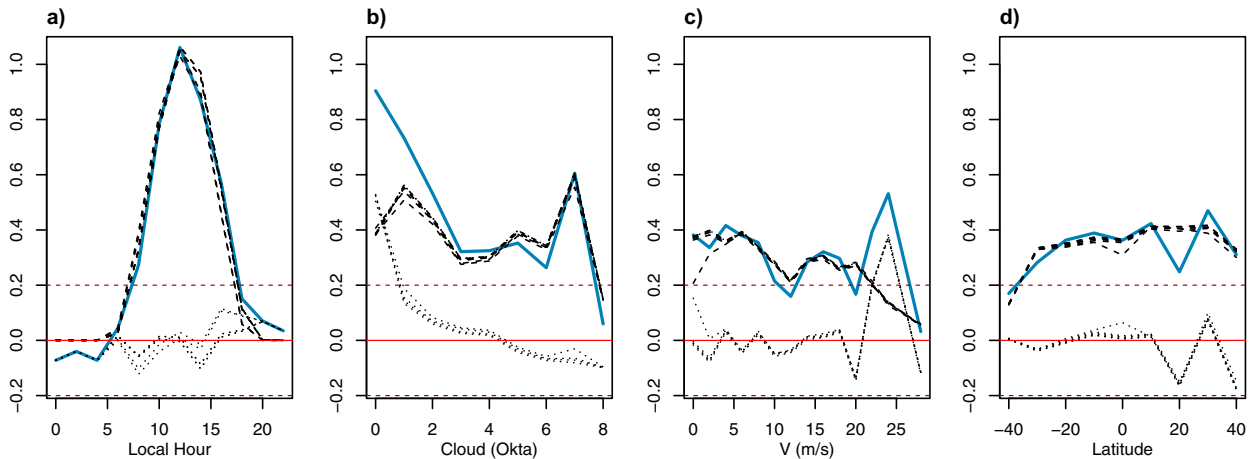


FIG. 3. The mean  $\Delta T_{\text{diur}}$  (solid blue line),  $\Delta T_{\text{BKT}}$  (dashed lines), and residual (dotted lines) for the ship *Mary* (Table 3) grouped by (a) local hour (every 2 h), (b) cloud cover (one okta intervals), (c)  $2 \text{ m s}^{-1}$  intervals of  $V$ , and (d)  $10^\circ$  latitude bins. Individual dashed and dotted lines represent the best-fitting  $\Delta T_{\text{BKT}}$  from each of the six cost functions described in section 2d. The horizontal red lines indicate  $0^\circ$  and  $\pm 0.2^\circ\text{C}$  limits. Each bin contains at least 100 observations, except for 0 and 7 oktas, 18 and  $>20 \text{ m s}^{-1}$ , and  $40^\circ$  latitude.

To determine the relative improvement of MAT data after applying the BKT model adjustment,  $\Delta T_{\text{diur}}$  and  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  should be compared. First, it is clear that the spread of  $\Delta T_{\text{diur}}$  values (Fig. 4a) is greater than the spread for both  $\Delta T_{\text{BKT}}$  (Fig. 4b) and  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$

(Fig. 4c), as expected. The RMSE reduction ( $\Delta T_{\text{diur}}$  cf.  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$ ) ranges from 15% (U.S. Navy 12388,  $1.53^\circ\text{--}1.35^\circ\text{C}$ ) to 53% (*Kajtum*,  $2.45^\circ\text{--}1.13^\circ\text{C}$ ), with a mean of 28% across all 16 ships. The RMSE reduction significantly correlates ( $r = 0.92$ ) with the magnitude of  $\Delta T_{\text{diur}}$ .

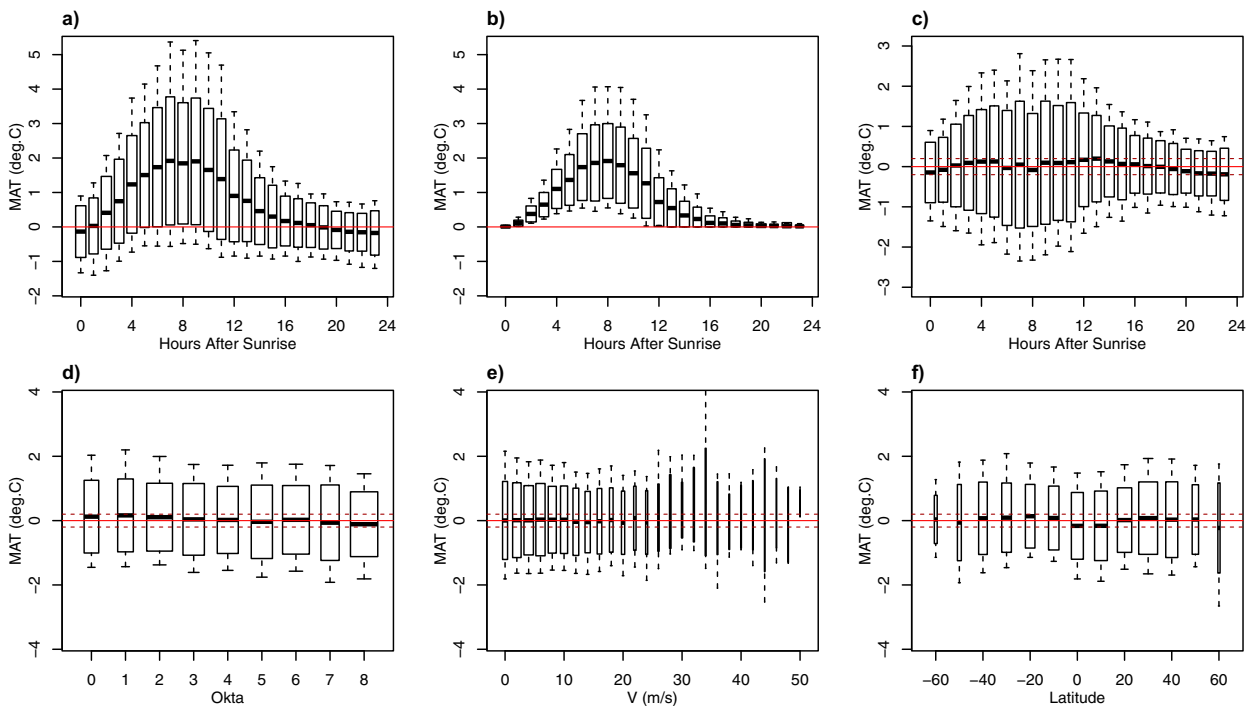


FIG. 4. Boxplots displaying the bin mean (solid line), bin mean  $\pm 1$  standard deviation (box limits), and 5th and 95th percentiles (whiskers) for (a)  $\Delta T_{\text{diur}}$  and (b)  $\Delta T_{\text{BKT}}$  as grouped by the number of hours after sunrise. (c)–(f) The  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  residual when grouped by (c) the number of hours after sunrise, (d) cloud cover, (e)  $2 \text{ m s}^{-1}$  intervals of  $V$ , and (f)  $10^\circ$  latitude bins. All 16 ships from Table 3 are included and the  $\Delta T_{\text{BKT}}$  is taken as the ensemble mean across 60 realizations of the  $\Delta T_{\text{BKT}}$  (the 10 best-fit realizations from each of the 6 cost functions described in section 2d). The horizontal solid red and dark-red dashed lines indicate zero and  $\pm 0.2^\circ\text{C}$  limits, respectively. The box widths correspond to the square root of the sample size in each bin.

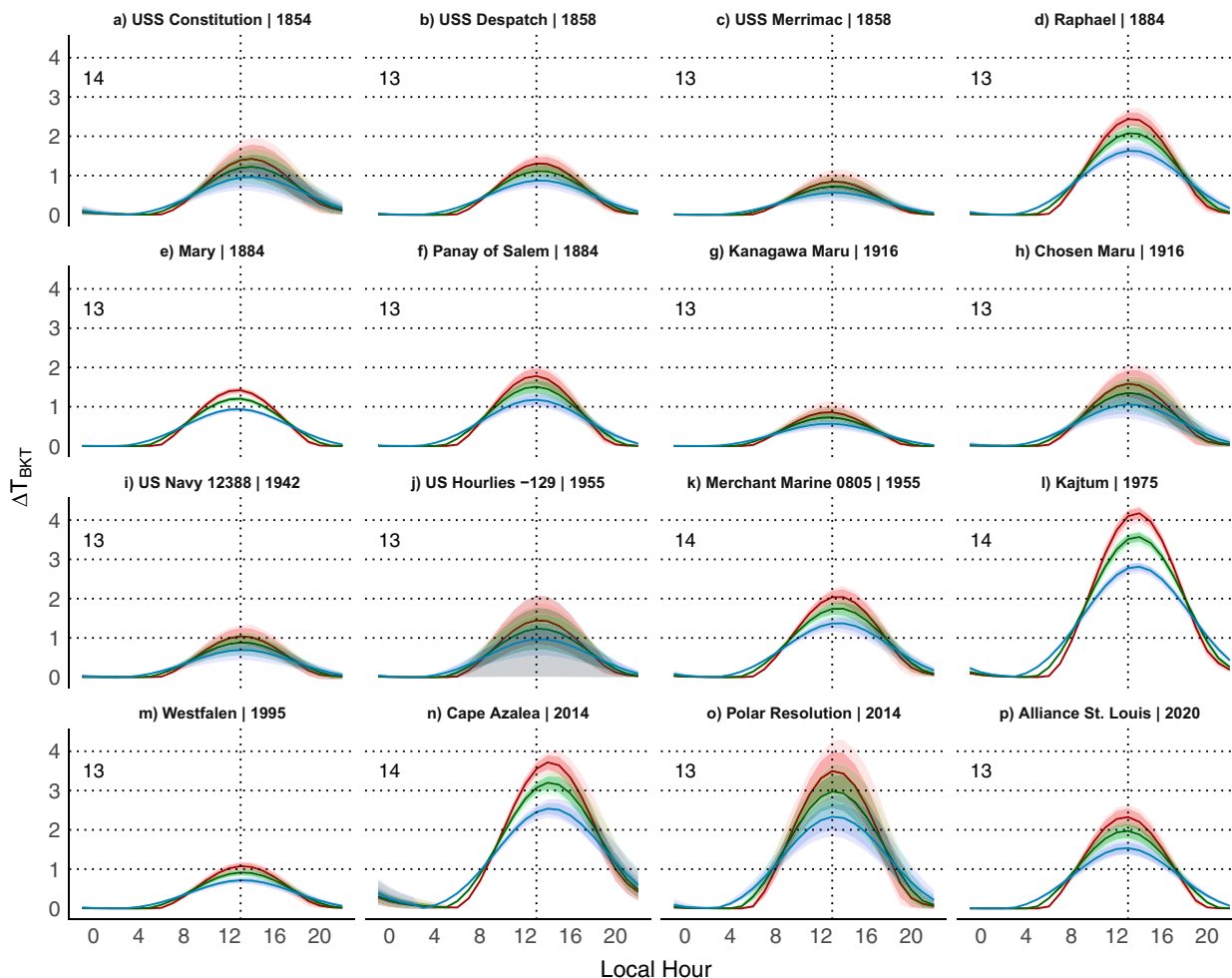


FIG. 5. The mean (solid line), standard deviation range (darker shading), and 5th–95th-percentile range (lighter shading) for  $\Delta T_{\text{BKT}}$  for the 16 different ships (Table 3) under fixed environmental conditions of  $15 \text{ m s}^{-1}$   $V$ , four oktas cloud cover,  $-20^\circ$  longitude, Julian day 150, and variable latitudes  $25^\circ\text{N}$  (red),  $50^\circ\text{N}$  (green), and  $65^\circ\text{N}$  (blue). The vertical line is at 1300 local time and the number in the upper left of each panel indicates the peak heating hour at  $25^\circ\text{N}$ .

It is not expected that  $\Delta T_{\text{BKT}}$  will exactly match  $\Delta T_{\text{diur}}$ . Residuals will include the effects of any model misfit, errors in  $V$ , cloud cover, or the parameterization of solar radiation and other nonsystematic differences such as weather effects. The magnitude and variability of the residuals, and the percentage changes, will depend on the relative sizes of the adjustment required and these other factors.

Figure 5 illustrates values of  $\Delta T_{\text{BKT}}$  under fixed environmental conditions and for selected latitudes for each ship, using the 60 ensemble member BKT model coefficients for each ship. Under these conditions,  $\Delta T_{\text{BKT}}$  in terms of amplitude and timing is similar for some ships (e.g., the pairing of the USS *Merrimac* and *Kanagawa Maru*), and different for others. To adjust the *Kajtum* using the coefficients generated for the *Chosen Maru* would leave the *Kajtum* still retaining a large MAT diurnal cycle, whereas the inverse operation would generate a physically unrealistic diurnal cycle for the *Chosen Maru*. Uncertainties across the ships are largest

around the peak heating hours, and the uncertainty range across different ships will relate to the magnitude of the  $\Delta T_{\text{diur}}$  and the environmental conditions, which will depend on the region in which the ship was operating. Figure 5 illustrates the importance of obtaining a BKT model solution for individual ships, but also suggests that coefficients can be estimated for groups of similar ships (see sections 3c and 3e).

If a ship contains observations where it was not possible to determine  $T_{\text{nt}}$ , but there are sufficient  $T_{\text{nt}}$  observations for that ship to fit the BKT model, then every observation with a corresponding cloud and  $V$  can be adjusted since sets of BKT model coefficients can be determined.

#### b. Estimating missing cloud and $V$ values

Depending on the observation source, MAT will not always be accompanied by cloud and wind observations. Figure 1 shows the proportion of potentially adjustable MAT observations using the BKT model has been decreasing since a sustained peak in

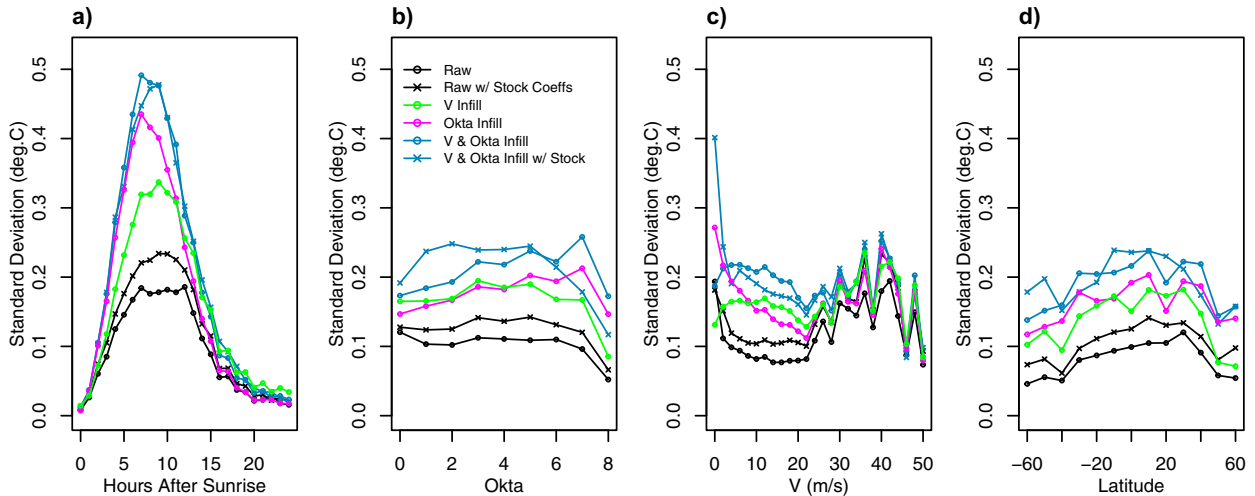


FIG 6. (a)–(d) The bin-mean  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  residual and (e)–(h) the bin mean of the 60-member standard deviation of the  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  residual. Six different approaches to defining the  $\Delta T_{\text{BKT}}$  were used: the “normal” approach using raw observational data (black line with circles), using infilled cloud (magenta line),  $V$  (green line), both cloud and  $V$  (blue line with circles) alongside fitting the  $\Delta T_{\text{BKT}}$  using “stock” coefficient combinations for raw observational data (black line with crosses) and infilled cloud and  $V$  (blue line with crosses).

the 1980s, likely due to increasing contributions from automatic weather stations in ICOADS in the modern era (Freeman et al. 2017).

As a means to examine the impact of infilling data on the BKT model adjustment (explored in section 3d), we generate the empirical histogram of clouds on a  $1^\circ$  spatial grid at monthly resolution (using ICOADS data from 1961 to 1990). We can then sample cloud cover values from this climatological histogram to generate ensembles of cloud cover estimates for MAT with missing cloud cover, which will vary across a  $1^\circ$  grid and month. Similarly for  $V$ , we sample wind speed (ws) values from the Rayleigh distribution, with the scale parameter set as  $\overline{\text{ws}}/\sqrt{\pi}$ , and direction sampled from the climatological distribution of the 16-point compass direction. To generate  $V$ , we further add a random directional component from the uniform distribution ( $\pm 22.49^\circ$ ) to the coarse-resolution wind direction and then use the observed ship speed and value to recalculate a sampled  $V$ .

### c. Bulk application of the BKT model using “stock” coefficient combinations

The optimization of model coefficients is computationally intensive and impractical for application to every ship in ICOADS. To avoid this the optimization was applied to over 10000 ships in ICOADS during the period 1854–2020, generating a collection of “stock” coefficients (without using infilled cloud and  $V$ ). Many of these coefficient combinations were similar, so we reduced the number of stock coefficients. We first reduced the number of the coefficients by removing duplicate values across the four  $x$  coefficients when rounding to two significant figures. We then calculated the hourly BKT adjustment value for a selection of spatial locations, environmental conditions, and days, and removed coefficient combinations

resulting in the same hourly rounded ( $0.1^\circ\text{C}$ ) values of the  $\Delta T_{\text{BKT}}$  throughout the day. This results in a stock coefficient dataset of 2500 coefficients, suitable for adjustment of data from widely differing ships. The 2500 different possible  $\Delta T_{\text{BKT}}$  values can be calculated and the coefficients selected using the same set of cost functions used for optimization.  $\Delta T_{\text{BKT}}$  values can be determined following the same approach in section 2d allowing efficient adjustment of large datasets.

### d. BKT adjustment and uncertainty using “stock” coefficients and climatological infilling

The impact of infilling missing cloud and  $V$  values (section 3b) and fitting the model using a pool of stock coefficients in lieu of running the optimization (section 3c) is shown in Fig. 6. Figure 6a shows that the mean uncertainty value (defined as one standard deviation of the 60-member  $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$  ensemble spread) is at a minimum when using raw observation data and fitting via optimization, with largest uncertainty values during the peak heating hours of 6–12 h after sunrise. The uncertainty increases slightly when using raw observation data and the stock coefficients (black line with crosses), and further increases when infilling  $V$  and cloud cover (green and magenta lines). The greatest increase in uncertainty comes from replacing observation data with climatological infilling of both  $V$  and cloud. Using either optimized (section 2b) or stock coefficients (section 3c) when infilling both variables makes little difference (both blue lines). The greater increase in uncertainty when infilling cloud only (magenta line) as opposed to  $V$  only (green line) is logical in the context of the BKT model [Eq. (10)] as the okta value scales the incoming solar radiation, and that sets the initial magnitude of  $\Delta T_{\text{BKT}}$ . This pattern typically holds true when assessing the uncertainty change against bins of cloud cover,  $V$ , and latitude, though some bin values



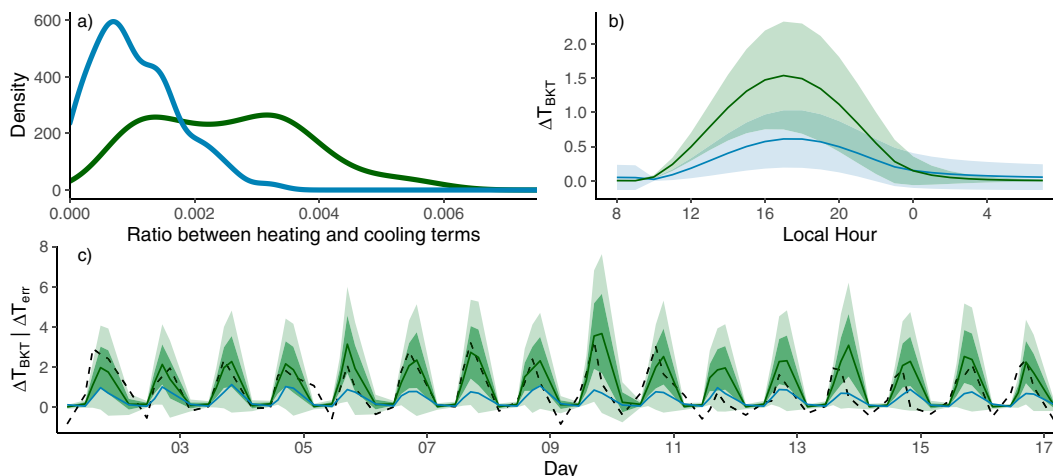


FIG. 7. (a) Density plot of the ratio between the BKT model heating and cooling terms (i.e.,  $x_1/x_3 V^{x_4} + x_5$ ) for each cost function as selected by either the 83-member ensemble using the 1854–70 grouping (blue) or the 59-member ensemble using pre-1854 ships (green). (b) The magnitude and uncertainty of  $\Delta T_{\text{BKT}}$  under the same fixed environmental conditions as in Fig. 5 at 25°N. (c) The  $\Delta T_{\text{diur}}$  time series (black dashed line) for the ship HMS *Favorite* during December 1831, with  $T_{\text{BKT}}$  for the pre-1854 ensemble (solid green line) where shading intensity corresponds to  $\pm 1$ –2 standard deviations. The  $T_{\text{BKT}}$  for the 1854–70 grouping is shown without shading for comparison (solid blue line).

differ. Climatological infilling of both parameters typically doubles the uncertainty compared to using raw data and optimizing (Fig. 6). The relatively minor increase in uncertainty when using stock coefficients and raw data gives us confidence in the en masse application of the BKT model using this approach.

#### e. Application to pre-1854 ships

If it is not possible to generate any  $T_{\text{nt}}$  values for a ship,  $\Delta T_{\text{diur}}$  cannot be estimated [Eq. (2)] and the BKT model cannot be fit using the methodology we outline in section 2d. A  $\Delta T_{\text{BKT}}$  value can be determined using stock coefficients, but the chosen sets of coefficients have to be determined via analog, based on an expected  $\Delta T_{\text{diur}}$  profile for the particular ship.

Before ca. 1854 there are increasingly fewer NMAT observations (Fig. 1a), and ships that do sample the diurnal cycle are unlikely to have cloud and  $V$  observations available (Fig. 1b).

Stock coefficients enable an estimation of the  $\Delta T_{\text{BKT}}$  to be made without a  $\Delta T_{\text{diur}}$  target. As the accuracy of the adjustment cannot be directly assessed this way, the quality of the adjustment will be based on the efficacy of the grouping of the ships. For example, we can expect that most ships pre-1854 are wooden-hulled sailing ships, with nonstandard observing practices (i.e., differences between countries and individual ships). It would therefore be desirable to obtain a set of coefficients that have been successfully applied to analog ships during the following years. Good analogs are difficult to derive during the early 1850s as the global shipping fleet transitioned from sail to steam. However, the 1853 Brussels Marine Conference (Maury 1853) led to an increased standardization of measuring practices and hence the metadata in ICOADS/digitized records could enable selection of ships in the decades following 1854 that are the most appropriate counterparts to the assumptions listed above.

We trial two attempts to adjust the pre-1854 data. First, data from all ships between 1854 and 1870 are used as the *pre-1854 analog period*. Each stock coefficient combination is given an identification number, and the number of times each set of stock coefficient occurs within the 60-member ensemble for a ship in the 1854–70 period occurs is counted. From this, a break in the most frequently occurring coefficients was identified at  $n = 83$ , which generated an ensemble of 83 different realizations from the stock coefficients that are then applied to the pre-1854 ships. The mean of the  $\Delta T_{\text{BKT}}$  is determined from these 83 sets of coefficients and the standard deviation of the  $\Delta T_{\text{BKT}}$  values becomes the uncertainty range. Second, for pre-1854 ships with over 50  $\Delta T_{\text{diur}}$  observations, we generate cloud and  $V$  values as in section 3b, which enables the BKT model to be fit as in section 3c, resulting in a 59 member ensemble size.

Figure 7a presents a density plot of the ratio between the heating and cooling terms of the BKT model. This allows a broad approximation of the exposure and heating bias of a ship. The 1854–70 ensemble is characterized by most ships' ratio being below  $\sim 0.002$ , which after Fig. 4 in Berry and Kent (2005), is an appropriate range for *good* ships with a low heating bias. The pre-1854 ensemble distribution is more uniformly spread, indicative of more ships with larger heating biases. This is reflected in Fig. 7b, using the same fixed environmental conditions as Fig. 5; the  $\Delta T_{\text{BKT}}$  is shown to be larger (and less certain) for the pre-1854 ensemble. The impact on a MAT time series using either ensemble is shown in Fig. 7c, alongside the  $\Delta T_{\text{diur}}$ . For the HMS *Favorite* during December 1831, it is clear that the pre-1854 ensemble (Fig. 7c) captures the evolution of the  $\Delta T_{\text{diur}}$  more appropriately than the 1854–70 ensemble (Fig. 7d) as the  $\Delta T_{\text{diur}}$  often falls out of the uncertainty range for the latter.

En masse application of the pre-1854 ensemble of coefficients to the pre-1854 data would result in larger values of

$\Delta T_{\text{BKT}}$  as opposed to using the 1854–70 ensemble. The purpose of the comparison here is not to identify the better overall choice, but to highlight that it is possible to achieve sensible heating bias adjustments to the early data. Rather than a broadscale adjustment, specific **BKT** model coefficient groupings could be made for different ICOADS decks or source IDs, and as for newly digitized data as they become available. Utilizing this analog approach to the heating bias adjustment is not limited to pre-1854, and could be used throughout the full ICOADS period.

#### 4. Summary and discussion

##### a. General application of the *BKT* model

In this paper we have extended the method developed by **BKT** for the correction of diurnal heating biases in ship-based air temperature measurements. From our estimate of this heating bias, we are able to generate MAT time series for individual ships,  $T_{\text{adj}}$  (Fig. 2), that substantially reduces the  $\Delta T_{\text{err}}$ , leaving a mean residual within  $\pm 0.2^\circ\text{C}$  (Fig. 4). Results focus on a sample of 16 ships, but the approach is applicable to all ship-based observations in ICOADS and ultimately will be used in the construction of improved estimates of global surface air temperature trends.

Our  $\Delta T_{\text{diur}}$  estimate [Eq. (2)], based on the difference between MAT and the underlying NMAT trend, minus the climatological SST cycle from buoys as defined in Morak-Bozzo et al. (2016), is likely an overestimate of the true heating bias. The heating bias is difficult to disentangle from the true diurnal cycle as both depend on the incoming solar radiation.

Application of the **BKT** model requires observations to be part of a ship-track time series, either through an extant identifier or after application of a tracking methodology (Carella et al. 2017), to enable  $\Delta T_{\text{diur}}$  to be calculated. This can be ameliorated by improved tracking methods or ensuring ship identify information is preserved in metadata records as they are stored/digitized.

For ships that lack accompanying cloud and  $V$  it is possible to estimate the MAT daytime bias (Fig. 6) by infilling these variables. The uncertainty in  $\Delta T_{\text{BKT}}$  inflates to account for the infilling.

It is possible to achieve a removal of the daytime heating bias for ships without sampling across the diurnal cycle; this is required for temporal extension of the MAT record further back in time than ca. 1854. For example, the English East Indian Company ships (ICOADS Deck 248) mostly report a single daily observation at local noon, which makes determining a nighttime value and  $\Delta T_{\text{diur}}$  estimate impossible. However, in this paper we have demonstrated that if a sufficient number of analog ships can be identified, which are able to be adjusted, the most commonly occurring **BKT** model coefficients used in the adjustment of these ships can be used to generate an ensemble of  $\Delta T_{\text{BKT}}$  for these older ships (Fig. 7), enabling a backward temporal extension of the MAT record. Here, an outline for choosing analog ships was made, but this can be refined in the future as data from newly recovered

sources are digitized, and/or metadata tied to existing observations are utilized.

##### b. Data issues and quality control

Relatively strict quality control procedures have been applied (appendix) to ensure the analysis uses data that accurately portray the measured diurnal cycle.

The diurnal-cycle-based quality control routine (appendix) identified data from a number of ships in the 1880s that passed the climatology-based QC checks but that had a  $\sim 12$ -h offset. Without removal or adjustment these data would adversely affect NMAT datasets. Furthermore we were able to identify ships suspected of making measurements in cabins, by analysis of the peak hour of  $\Delta T_{\text{diur}}$  (appendix). Overall, this shows that there is still much to be learned about MAT observations and diurnal-cycle-based assessments are likely to remain a useful tool in improving the long-term records (Cornes et al. 2020; Chan and Huybers 2021). A further unresolved issue is whether some reported ICOADS wind and direction values are the true wind and direction, or relative values uncorrected for ship trajectory (Gulev 1999).

##### c. Precipitation and weather codes

The presence of precipitation invalidates the energy transfer assumptions of the **BKT** model. When the recording of the present weather (WW) code is systematically high ( $>95\%$  during the 1960–70s, green color in Fig. 1c), the percent of WW observations indicating precipitation is  $\sim 10\%$  (red color in Fig. 1c). As the WW code is not always recorded with every MAT observation, it may not be possible to identify all observations that may have been affected by precipitation. Further work is required to better identify affected observations and to understand the impact of precipitation on the heating bias.

##### d. Systematic structure in diurnal residuals

The approach outlined here, across the 16 analyzed ships, reduces the mean hourly local time error in all-hours observations ( $\Delta T_{\text{err}} \approx \Delta T_{\text{diur}}$ ) to within  $\pm 0.2^\circ\text{C}$  ( $\Delta T_{\text{diur}} - \Delta T_{\text{BKT}}$ ). However, a systematic diurnal structure remains in the residuals of the **BKT** model adjustment (Fig. 4c). Further reduction in these residuals is likely to require an improved analysis method. Examples of possible improvements might be better estimates of incoming solar radiation, potentially including a diffuse term; reinstatement of the original  $x_6$  thermal transfer term (which would add dewpoint temperature as a data requirement in applying the **BKT** model); or explicitly estimating, or fitting the true diurnal cycle of MAT.

##### e. The need for more complete data and metadata

The value in the recovery and digitization of MAT data, in terms of the marine contribution to extending the global temperature record, cannot be overstated. While much work has been done in extracting historical observations from available archives, e.g., García-Herrera et al. (2005), extra value can be prescribed to MAT observations that cover the full diurnal cycle and have concomitant cloud and wind speed observations, particularly for pre-1854. Cloud cover is an essential climate

variable, but the advent of automated measurements has resulted in major drop in ship-based cloud observations in ICOADS since the peak in the 1980s (Kent et al. 2019).

*Acknowledgments.* This work was funded by the Natural Environment Research Council (NERC) (GloSAT NE/S015647/2), making use of software and methods developed under HOSTACE (NE/J020788/1), CLASS (NE/R015953/1), and Copernicus Climate Change Service Contract C3S\_311a\_Lot2.

*Data availability statement.* ICOADS data are available from <https://rda.ucar.edu/datasets/ds548.0/>. Heliosat data (DOI: 10.5676/EUM\_SAF\_CM/SARAH/V002) were acquired from <https://wui.cmsaf.eu/safira/action/viewProduktHome>. The code and data to generate the figures alongside an example of running the BKT model fit are available at <https://git.noc.ac.uk/thocro/heating-bias-paper-figures>.

## APPENDIX

### Data Selection and Quality Control

MAT values that fail quality control (QC) checks are excluded from this analysis. The QC routine includes the following:

- 1) The QC flag from CLASSmat (Cornes et al. 2020), which is a combination of an initial application of the Met Office QC checks (<https://github.com/ET-NCMP/MarineQC>) and the outliers identified in the processing of CLASSmat.
- 2) QC checks as applied to the Shipboard Automated Meteorological and Oceanographic System data described in Smith et al. (2018). We also remove any observations in a track where the MAT deviates from the running median by  $\pm 15^\circ\text{C}$  (window length depending on sampling frequency). Individual ship tracks are split into segments if gaps in MAT exceed two days. Observations in each segment where the MAT is  $\geq 4$  standard deviations above or below the mean for that segment, and the absolute lag<sub>1</sub> MAT exceeds the mean value of absolute lag<sub>1</sub> differences are removed. Additionally, observations with  $\Delta T_{\text{diur}}$  values  $\geq -5^\circ$  and  $\leq 15^\circ\text{C}$  are removed.
- 3) All precipitation-flagged observations identified from the ICOADS present weather code following Petty (1995) are excluded.
- 4) We identify and exclude observations from ships where a one-tailed *t* test indicates a statistically significant warmer night than day. This test is applied both monthly and annually, removing months and years that fail either of the tests. Ships with less than 5 days with corresponding day and night observation values are not assessed.
- 5) Observations from ships during the period 1854–94 with mean peak heating hour  $\leq 2$  or  $\geq 13$  h after sunrise or between 2200 and 0600 local time are excluded. This identifies observations that may have been taken in cabins during rough conditions—predominantly at nighttime, in colder months in the higher latitudes, as investigated by Chenoweth (2000).

## REFERENCES

- Aleksandrova, M. P., S. K. Gulev, and A. V. Sinitsyn, 2007: An improvement of parametrization of short-wave radiation at the sea surface on the basis of direct measurements in the Atlantic. *Russ. Meteor. Hydrol.*, **32**, 245–251, <https://doi.org/10.3103/S1068373907040048>.
- , —, and K. Belyaev, 2018: Probability distribution for the visually observed fractional cloud cover over the ocean. *J. Climate*, **31**, 3207–3232, <https://doi.org/10.1175/JCLI-D-17-0317.1>.
- 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**, 1007–1022, <https://doi.org/10.1002/joc.1178>.
- , and —, 2009: A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Amer. Meteor. Soc.*, **90**, 645–656, <https://doi.org/10.1175/2008BAMS2639.1>.
- , and —, 2011: Air–sea fluxes from ICOADS: The construction of a new gridded dataset with uncertainty estimates. *Int. J. Climatol.*, **31**, 987–1001, <https://doi.org/10.1002/joc.2059>.
- , —, and P. K. Taylor, 2004: An analytical model of heating errors in marine air temperatures from ships. *J. Atmos. Oceanic Technol.*, **21**, 1198–1215, [https://doi.org/10.1175/1520-0426\(2004\)021<1198:AAMOHE>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<1198:AAMOHE>2.0.CO;2).
- Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: *Global Ocean Surface Temperature Atlas ‘GOSTA’*. Meteorological Office–Massachusetts Institute of Technology, 20 pp.
- Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, 1995: A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, **16**, 1190–1208, <https://doi.org/10.1137/0916069>.
- Carella, G., E. C. Kent, and D. I. Berry, 2017: A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.*, **37**, 2233–2247, <https://doi.org/10.1002/joc.4492>.
- Chan, D., and P. Huybers, 2021: Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II. *J. Climate*, **34**, 4585–4602, <https://doi.org/10.1175/JCLI-D-20-0907.1>.
- Chenoweth, M., 2000: A new methodology for homogenization of 19th century marine air temperature data. *J. Geophys. Res.*, **105**, 29 145–29 154, <https://doi.org/10.1029/2000JD900050>.
- Cornes, R. C., E. C. Kent, D. I. Berry, and J. J. Kennedy, 2020: CLASSmat: A global night marine air temperature data set, 1880–2019. *Geosci. Data J.*, **7**, 170–184, <https://doi.org/10.1002/gdj3.100>.
- Cowtan, K., and Coauthors, 2015: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. *Geophys. Res. Lett.*, **42**, 6526–6534, <https://doi.org/10.1002/2015GL064888>.
- Dobson, F. W., and S. D. Smith, 1988: Bulk models of solar radiation at sea. *Quart. J. Roy. Meteor. Soc.*, **114**, 165–182, <https://doi.org/10.1002/qj.49711447909>.
- Freeman, E., and Coauthors, 2017: ICOADS release 3.0: A major update to the historical marine climate record. *Int. J. Climatol.*, **37**, 2211–2232, <https://doi.org/10.1002/joc.4775>.
- García-Herrera, R., G. P. Können, D. A. Wheeler, M. R. Prieto, P. D. Jones, and F. B. Koek, 2005: CLIWOC: A climatological database for the world’s oceans 1750–1854. *Climatic Change*, **73**, 1–12, <https://doi.org/10.1007/s10584-005-6952-6>.
- Gulev, S. K., 1999: Comparison of COADS release 1a winds with instrumental measurements in the northwest Atlantic. *J.*

- Atmos. Oceanic Technol.*, **16**, 133–145, [https://dx.doi.org/10.1175/1520-0426\(1999\)016<0133:COCRWW>2.0.CO;2](https://dx.doi.org/10.1175/1520-0426(1999)016<0133:COCRWW>2.0.CO;2).
- , and M. Aleksandrova, 2020: Homogenizing visually observed cloud cover over global oceans with implications for reconstructions of radiative fluxes at sea surface. *EGU General Assembly 2020*, Online, EGU, Abstract EGU2020-12826, <https://doi.org/10.5194/egusphere-egu2020-12826>.
- , and Coauthors, 2021: Changing state of the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 287–422.
- Huang, B., and Coauthors, 2017: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179–8205, <https://doi.org/10.1175/JCLI-D-16-0836.1>.
- , and Coauthors, 2020: Uncertainty estimates for sea surface temperature and land surface air temperature in NOAA Global-Temp version 5. *J. Climate*, **33**, 1351–1379, <https://doi.org/10.1175/JCLI-D-19-0395.1>.
- Jones, G. S., 2020: “Apples and oranges”: On comparing simulated historic near-surface temperature changes with observations. *Quart. J. Roy. Meteor. Soc.*, **146**, 3747–3771, <https://doi.org/10.1002/qj.3871>.
- Junod, R. A., and J. R. Christy, 2020: A new compilation of globally gridded night-time marine air temperatures: The UAHNMAV1 dataset. *Int. J. Climatol.*, **40**, 2609–2623, <https://dx.doi.org/10.1002/joc.6354>.
- Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick, 2019: An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. *J. Geophys. Res. Atmos.*, **124**, 7719–7763, <https://doi.org/10.1029/2018JD029867>.
- Kent, E. C., and J. J. Kennedy, 2021: Historical estimates of surface marine temperatures. *Annu. Rev. Mar. Sci.*, **13**, 283–311, <https://doi.org/10.1146/annurev-marine-042120-111807>.
- , 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 HadNMAV2 data set. *J. Geophys. Res. Atmos.*, **118**, 1281–1298, <https://dx.doi.org/10.1002/jgrd.50152>.
- , and Coauthors, 2019: Observing requirements for long-term climate records at the ocean surface. *Front. Mar. Sci.*, **6**, 441, <https://doi.org/10.3389/fmars.2019.00441>.
- Lessen, N. J. L., G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, and D. Zys, 2019: Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.*, **124**, 6307–6326, <https://doi.org/10.1029/2018JD029522>.
- Maury, M. F., 1853: *Maritime Conference Held at Brussels for Devising an Uniform System of Meteorological Observations at Sea, August and September 1853*. Maury, 125 pp.
- Morak-Bozzo, S., C. J. Merchant, E. C. Kent, D. I. Berry, and G. Carella, 2016: Climatological diurnal variability in sea surface temperature characterized from drifting buoy data. *Geosci. Data J.*, **3**, 20–28, <https://doi.org/10.1002/gdj3.35>.
- Morice, C. P., and Coauthors, 2021: An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *J. Geophys. Res. Atmos.*, **126**, e2019JD032361, <https://doi.org/10.1029/2019JD032361>.
- Petty, G. W., 1995: Frequencies and characteristics of global oceanic precipitation from shipboard present-weather reports. *Bull. Amer. Meteor. Soc.*, **76**, 1593–1616, [https://doi.org/10.1175/1520-0477\(1995\)076<1593:FACOGO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<1593:FACOGO>2.0.CO;2).
- Pfeifroth, U., S. Kothe, J. Trentmann, R. Hollmann, P. Fuchs, J. Kaiser, and M. Werscheck, 2019: Surface Radiation data set—Heliosat (SARAH)—Edition 2.1. CM SAF, accessed 8 March 2022, [https://doi.org/10.5676/EUM\\_SAF\\_CM/SARAH/V002](https://doi.org/10.5676/EUM_SAF_CM/SARAH/V002).
- R Core Team, 2019: R: A language and environment for statistical computing. R Foundation for Statistical Computing, <https://www.R-project.org/>.
- Richardson, M., K. Cowtan, E. Hawkins, and M. B. Stolpe, 2016: Reconciled climate response estimates from climate models and the energy budget of Earth. *Nat. Climate Change*, **6**, 931–935, <https://doi.org/10.1038/nclimate3066>.
- Rubino, A., D. Zanchettin, F. D. Rovere, and M. J. McPhaden, 2020: On the interchangeability of sea-surface and near-surface air temperature anomalies in climatologies. *Sci. Rep.*, **10**, 7433, <https://doi.org/10.1038/s41598-020-64167-1>.
- Smith, S. R., K. Briggs, M. A. Bourassa, J. Elya, and C. R. Paver, 2018: Shipboard automated meteorological and oceanographic system data archive: 2005–2017. *Geosci. Data J.*, **5**, 73–86, <https://doi.org/10.1002/gdj3.59>.