

Toward Estimating Climatic Trends in SST. Part III: Systematic Biases

ELIZABETH C. KENT

National Oceanography Centre, Southampton, United Kingdom

ALEXEY KAPLAN

Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York

(Manuscript received 24 February 2004, in final form 24 July 2005)

ABSTRACT

A method is developed to quantify systematic errors in two types of sea surface temperature (SST) observations: bucket and engine-intake measurements. A simple linear model is proposed where the SST measured using a bucket is cooled or warmed by a fraction of the air–sea temperature difference and the SST measured using an engine intake has a constant bias. The model is applied to collocated nighttime observations made at moderate wind speeds, allowing the effects of solar radiation and strong vertical gradients in the upper ocean to be neglected. The analysis is complicated by large random errors in all of the variables used. To estimate coefficients in this model, a novel type of linear regression, where errors in two variables are correlated with each other, is introduced. Because of the uncertainty in a priori estimates of the error covariance matrix, a Bayesian analysis of the regression problem is developed, and maximum likelihood approximations to the posterior distributions of the model parameters are obtained.

Results show that the temperature change in bucket SST resulting from the air–sea temperature difference can be detected. The analysis suggests that bucket SST may be in error by a fraction from $0.12^\circ \pm 0.02^\circ$ to $0.16^\circ \pm 0.02^\circ\text{C}$ of the air–sea temperature difference. When this temperature change of the bucket SST is accounted for, a warm bias in engine-intake SST in the mid- to late 1970s and the 1980s was found to be smaller than that suggested by previous studies, ranging between $0.09^\circ \pm 0.06^\circ$ and $0.18^\circ \pm 0.05^\circ\text{C}$. For the early 1990s the model suggests that the engine-intake SSTs may have a cold bias of $-0.13^\circ \pm 0.07^\circ\text{C}$.

1. Introduction

It is well known that there are significant differences between in situ sea surface temperature (SST) measured by voluntary observing ships (VOSS) using the two most common methods: bucket and engine intake (Roll 1951; Saur 1963; Walden 1966; Crawford 1969; Tauber 1969; James and Fox 1972; Parker 1985; Folland et al. 1993, Kent et al. 1993; Folland and Parker 1995, hereafter FP95). Kent and Taylor (2006, hereafter Part I) summarize the results of these studies, some of which have never been published in the refereed literature. However, modeling these differences is difficult. For example, the bucket SST is expected to be affected by the effects of turbulent fluxes, which typically cool the

sample as it is hauled from the sea surface to the deck. To estimate the turbulent fluxes we need the air–sea temperature difference, which is poorly known due partly to large random errors in the temperature observations (Kent and Berry 2005) and partly to biases in the data (Berry et al. 2004; Part I). Any individual SST measurement is affected by the individual circumstances of the observation, for example, the bucket could have been left in the sun before the measurement was taken, or the engine-intake thermometer could be positioned close to the ships engines. However, by taking a large number of observations we hope to reduce random effects and determine important systematic errors.

Section 2 describes the data sources used in the study. In section 3 the model for systematic errors in the bucket and engine-intake measurements is presented and reduced to the special type of linear regression problem where errors in two variables are correlated with each other and a priori estimates of error covari-

Corresponding author address: Dr. Elizabeth C. Kent, National Oceanography Centre, European Way, Southampton SO14 3ZH, United Kingdom.
E-mail: Elizabeth.C.Kent@noc.soton.ac.uk

ance matrix elements have uncertainty in them. The solution to this regression problem is given in the appendix. Section 4 presents the results, which are discussed in detail in section 5. Conclusions are drawn in section 6.

2. Data

As in Part I and Kent and Challenor (2005, hereafter Part II), we use data from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Woodruff et al. 1998; Diaz et al. 2002) for the period of 1970–97 and metadata from the World Meteorological Organization’s “List of selected, supplementary and auxiliary ships” (e.g., WMO 1997; Kent et al. 2005, manuscript submitted to *J. Atmos. Oceanic Technol.*, hereafter KEN05). The use of this external metadata, along with that contained within the ICOADS, extends back to the 1970s—the period over which comparisons of SSTs with different measurement methods can be made (Part I). Bathymetry data from the ETOPOS dataset (National Geophysical Data Center 1993; see section 3d herein) were also used.

3. Analysis methods

a. Errors in VOS SST and their effect on analysis

VOS weather reports are known to contain errors, both systematic (e.g., James and Fox 1972; Kent et al. 1993; Kent and Taylor 1997; Berry et al. 2004) and random (e.g., Kent et al. 1999; Part II; Kent and Berry 2005). This has two implications for the analysis of biases. First, a large volume of data needs to be analyzed to reduce the random component of the error, because for individual observations the random error is likely to be much larger than the bias. Walden (1966) noted “the difference in measurement of the water temperatures by the two methods is completely concealed by many . . . large errors.” Second, care must be taken to ensure that random or systematic errors do not create artificial dependencies between analyzed variables. Note that ships recruited by a particular country are likely to have been issued with the same instruments and instructions on how to take the observations. Disentangling physical variability and systematic biases are especially difficult when biases are parameterized as functions of variables such as latent or sensible heat flux, which are expressed as combinations of basic variables. To avoid these problems, here we propose an analysis method in which errors in the different types of observations are parameterized in a very simple form. The model used neglects the influence of wind speed, evaporation, and solar ra-

diation. It focuses on the biases in bucket SST measurements resulting from air–sea temperature differences and allows for a simple offset in engine-intake SSTs.

b. Physical effects responsible for bias in VOS sea temperatures, model design, and sample selection

Part I summarizes the main sources of measurement error in VOS SST. Here we wish to develop a model that can account for these errors yet remain simple. We only consider nighttime observations and therefore neglect SST measurement errors related to solar radiation (which are likely to mainly affect bucket-measured SST). The most important remaining factors are heat exchange with the atmosphere for the bucket SST and a potential bias in engine-intake SST (Part I). Therefore, our simple model includes the direct heat loss from the bucket, parameterized using the air–sea temperature difference. This direct heat loss in reality depends on the relative wind speed at the time of sampling (while the bucket and sample are hauled to the deck), and a reduced relative wind speed during measurement if the observation was made in a sheltered location. Heat loss by evaporation is not explicitly included; however, any element of evaporative heat loss that is correlated with the air–sea temperature difference can be implicitly included by an increase in the model coefficients.

There are several reasons keeping the model simple. To include further processes in the analysis requires more variables to be present within each report. For example, to represent more completely the effects of heat loss one needs to know the SST, air temperature, relative wind speed (i.e., the ship speed, its course, and the true wind speed and direction), and humidity. The number of applicable reports, containing all required variables, is small. The random error entering the analysis in a complicated model is large, because more variables are used.

There is a trade-off between the noise introduced into the analysis by a poor representation of the physics and that introduced by increased model complexity and reduced amount of data. When a simple model, like the one used in this study, gives physically reasonable results when applied to large quantities of data, a case of reasonable balance between complexity and error has been made. The application of these results to larger datasets, however, should be treated with caution, because of the approximations, assumptions, and exclusions made in their calculation.

We have not applied the model to daytime data for several reasons. Daytime air temperatures are contaminated by solar radiation effects (e.g., Berry et al. 2004). In fact, these air temperatures might need special cor-

rection prior to their use in any SST correction scheme. The water in the bucket experiences the true air temperature immediately after sampling when it is remote from the ship. However, once the bucket is hauled from the sea to the deck the water in the bucket is affected by the ship-modified air temperature. Restricting the analysis to nighttime data makes the differences between the true and ship-modified air temperature smaller.

During the day there may be large vertical gradients in ocean temperature between the relatively shallow bucket SST sample (typically taken at depths less than 1 m) and the depth of the engine-intake SST (typically between 1 and 4 m for small vessels such as fishing vessels, research ships, and Coast Guard ships, and between 5 and 15 m for larger ships; KEN05). We would expect these gradients to be largest in the daytime and at low wind speeds, so we have excluded these data from the analysis. Even larger daytime errors can occur if the bucket or sample are directly heated by solar radiation.

c. Formulation of the model

Let T represent the true value of the SST, unaffected by observational errors. Introduce T_b , T_e , and T_a as observations of bucket water temperature, engine-intake water temperature, and air temperature, respectively. The two SST measurements (T_b and T_e) differ because of a combination of random and systematic errors. We assume T_b is in error by a fraction α of the air–sea temperature difference and T_e is on average in error by a constant bias β :

$$T_b - T = \alpha(T_a - T), \tag{1}$$

$$T_e - T = \beta. \tag{2}$$

These equations are too simple to be precisely true, but we expect the errors in the model to be overwhelmed by the random observational errors affecting the actual measurements of T_b , T_e , and T_a , and thus we do not include explicit model error terms in (1) and (2).

The implicit assumption of the model (1)–(2) is that α and β do not vary significantly within the sample. Values of α falling outside the range of 0–1 are non-physical. The value $\alpha = 0$ means that the bucket temperature is unaffected by heat exchange with the air, and $\alpha = 1$ means that T_b and T_a are equal within their range of uncertainty. Physically β can take either sign; positive β means that the engine-intake water is warmer than the SST of the ocean.

By eliminating T between Eqs. (1) and (2), we obtain

$$T_b - T_e = \alpha(T_a - T_e) + (\alpha - 1)\beta. \tag{3}$$

Introducing

$$x = T_a - T_e, \quad y = T_b - T_e, \quad \gamma = (\alpha - 1)\beta, \tag{4}$$

we can rewrite (3) as a linear regression model

$$y = \alpha x + \gamma. \tag{5}$$

Given observations (x^o, y^o) for many pairs (x, y) , the parameters α and γ can be determined. Individual observations are affected by observational errors:

$$x^o = x + \varepsilon_x, \quad y^o = y + \varepsilon_y. \tag{6}$$

The regression results depend on the assumptions that are made about the observational errors ε_x and ε_y . Traditional types of regression account for cases where one of the variables is error free (ordinary linear regression) or where the error variances in each direction are equal (known as orthogonal regression). More advanced regression techniques include modifications to orthogonal regression where the variances of the error in each direction are unequal but their ratio is fixed (e.g., Draper and Smith 1998; Kent et al. 1998; Thomas et al. 2005) or for the case where the error variances can vary between individual points but are known (Boggs and Rogers 1990; Press et al. 1992; Deming 1964). However, none of these regression methods allows for correlation between ε_x and ε_y . In published studies such errors are assumed independent. In our case, this correlation is inherent because the observed values of both x and y contain T_e and its associated error. If the observational errors in T_b , T_e , and T_a are ε_b , ε_e , and ε_a , respectively, from (4) and (6) we obtain

$$\varepsilon_x = \varepsilon_a - \varepsilon_e, \quad \varepsilon_y = \varepsilon_b - \varepsilon_e.$$

We assume ε_b , ε_e , and ε_a are uncorrelated with each other and have variances c_b , c_e , and c_a , respectively. Thus,

$$\langle \varepsilon_x^2 \rangle = c_a + c_e, \quad \langle \varepsilon_y^2 \rangle = c_b + c_e, \quad \langle \varepsilon_x \varepsilon_y \rangle = c_e. \tag{7}$$

Furthermore, the parameters c_b , c_e , and c_a in our regression problem (5)–(7) are not known precisely; they are estimated as intercepts in a semivariogram regression (Part II). Therefore, they can be naturally modeled by normal distributions

$$c_b \sim \mathcal{N}(c_b^o, s_b), \quad c_e \sim \mathcal{N}(c_e^o, s_e), \quad c_a \sim \mathcal{N}(c_a^o, s_a), \tag{8}$$

where c_b^o , c_e^o , and c_a^o are estimates of the semivariogram intercepts for bucket, engine-intake, and air temperature measurements, respectively, which were calculated a priori using the technique of Part II. The standard errors s_b , s_e , and s_a , respectively, are calculated as part of the semivariogram analysis.

The correlation between errors in x and y affects the regression results dramatically and does not seem to

have been discussed in the literature. Uncertainty in the error variance parameters requires a Bayesian approach, that is, a study of the joint distribution of data and parameters and obtaining parameter distributions conditional on the data. We therefore present the Bayesian analysis of the problem (4)–(8) in the appendix, along with an algorithm for computing a maximum likelihood approximation to the posterior probability density distribution $P(\alpha, \beta)$ for the coefficients α and β in the model (1)–(2), conditional on all available data.

d. Application to ICOADS

To implement the model described in section 3c, a dataset of collocated bucket and engine-intake SSTs taken along with air temperatures was required. For this we used ICOADS to generate datasets of paired reports within 100-km spatial separation and taken at the same reporting hour. Our analysis only considers reports from the North Atlantic located between 20° and 50°N. ETOPO5 bathymetry (National Geophysical Data Center 1993) is used to ensure that the data were confined to the Atlantic. To improve the consistency of the dataset, only reports from selected countries were included in the analysis (the Netherlands, Norway, the United States, the United Kingdom, France, Denmark, Hong Kong, New Zealand, Canada, Belgium, South Africa, Japan, Sweden, Germany, Israel, the former Union of Soviet Socialist Republics, Russia, and Singapore). Any data for which no country could be identified, or reported by a country whose SSTs were known to be erratic or were too few to analyze, were excluded. We include only nighttime data and exclude very high and very low wind speeds from our analysis. We therefore attempt to minimize the effects of direct solar heating and of the vertical stratification of the surface ocean, which are not included in the model. It should be noted that we have used the air temperature from the ship in the pair reporting bucket SST rather than that reporting engine-intake SST, because ships using buckets to measure SST tend to report better-quality air temperatures. Air temperature records on these ships were better correlated with nearby engine-intake SST measurements than were the air temperatures from those same ships. This is thought to be because of the typical quality of measurements made by VOS recruited by different countries. For example, in the Voluntary Observing Ships' Special Observing Project for the North Atlantic (VSOP-NA) (Kent and Taylor 1991) the screens on VOS recruited by the United States tended to be more poorly exposed than those on VOS recruited by the European contributors. Excepting France, all of the European VOSs tended to report bucket SST. So the VSOP-NA ships reporting

engine-intake SST (typically from the United States) tended to have poorly exposed screens, which lead to larger air temperature errors (Berry and Kent 2005).

In addition to this dataset two other paired datasets of single type (bucket–bucket and engine intake–engine intake) were also generated to allow the calculation of random error variances c_b^o , c_e^o , and c_a^o and their respective uncertainties s_b , s_e , and s_a using the semivariogram method (Lindau 1995; Kent et al. 1999; Part II). These estimates and their uncertainties were used in our procedure via the definitions (8) and (7) in order to find the values of parameters α and β of the models (1)–(2), along with their uncertainties.

4. Results

The formalism of section 3c and the solution algorithm (see the appendix) were applied to triplets of observations T_b , T_e , and T_a from paired reports grouped in 5-yr periods between 1975 and 1994. Outside this period there were not enough paired reports for calculations to be made. Maximum likelihood approximations of the joint posterior distribution function (PDF) were calculated on a 121×121 grid covering ranges of α from 0.01 to 0.5 and β from -0.3 to 0.3 . The results are summarized in Table 1, where statistical parameters for α and β correspond to their marginal distributions calculated using Eqs. (A7) and (A8).

We select the most data-rich period of 1980–84 in order to illustrate the results in more detail. Figure 1 shows isolines for the joint posterior PDF of α and β , computed using Eq. (A6), as well as individual distributions for α and β [Eqs. (A7) and (A8)]. Figure 2 shows a scatter diagram of SST difference (bucket – engine intake) versus air–sea temperature difference using collocated data for the same period (1980–84). The scatter in the data is large and the most obvious tendency in the data is due to the strong intervariable error correlation. This is expected, because both variables contain the engine-intake SST, which has large observational errors (Part II). The data therefore are scattered along the 1:1 line of Fig. 2. Our analysis however reveals another data connection, which cannot be explained by the error structure of the data, and is represented by the regression line (5). This line with parameters corresponding to the maximum likelihood values of the posterior distribution $\alpha = 0.12$ and $\gamma = \beta(\alpha - 1) = 0.15 \times (0.12 - 1) = -0.13$ is shown in Fig. 2 by the solid line. The shaded area around it corresponds to the 95% probability area on the (α, β) plane shown in Fig. 1.

Values of α in each of the 5-yr subperiods in 1975–94 are quite similar (Table 1), but the value of β in the latest period (1990–94) is significantly more negative than in the previous periods. Figure 3 shows maximum

TABLE 1. Analysis results for 5-yr periods in the range of 1975–94. Maximum likelihood estimate (MLE) for the model parameters α and β are shown as well as their 95% confidence intervals. The coefficient α gives the proportion of the air–sea temperature difference appearing as an error in the bucket SST, and β gives the remaining mean offset between the bucket and engine-intake SST once the bucket error has been accounted for. Estimates of the random measurement error variance computed a priori by the variogram method (Part II) are shown ± 1 std dev of their uncertainty.

Period	No. of points	Random error variance estimates			Slope for bucket bias α		Bias in engine intake β ($^{\circ}\text{C}$)	
		SST, ($^{\circ}\text{C}$) ²			MLE	95% confidence interval	MLE	95% confidence interval
		Bucket $c_b \pm s_b$	Engine intake $c_e \pm s_e$	Air temp ($^{\circ}\text{C}$) ² $c_a \pm s_a$				
1975–79	8410	1.56 ± 0.63	2.69 ± 0.56	0.69 ± 0.25	0.16	0.14–0.18	0.09	0.03–0.15
1980–84	11 245	1.25 ± 0.49	2.84 ± 0.64	0.60 ± 0.20	0.12	0.10–0.13	0.15	0.10–0.20
1985–89	11 073	1.25 ± 0.31	2.74 ± 0.53	0.68 ± 0.15	0.12	0.10–0.14	0.18	0.14–0.23
1990–94	5122	1.41 ± 0.43	3.02 ± 0.76	0.65 ± 0.24	0.13	0.10–0.16	–0.13	–0.20––0.06

likelihood estimates of α and β for all four periods, together with their 67%, 95%, and 99% confidence areas. The uncertainties for α and β overlap significantly for the periods from 1975 to 1989, but the period of 1990–94 is well separated from the earlier periods. The latter period is also characterized by larger uncertainties compared to the earlier periods, because it has fewer data points (Table 1).

Figure 4 shows the results of this study applied to ICOADS data in the 1980s for winter and summer periods using averaged values of α and β for the two

periods covering the 1980s in Table 1. Corrections were computed using a monthly mean, $2^{\circ} \times 2^{\circ}$ area data set and were binned separately for reports made using buckets and engine intakes. The corrections ($\Delta T_b = T - T_b$, $\Delta T_e = T - T_e$) required for each dataset were obtained from algebraic rearrangements of Eqs. (1) and (2):

$$\Delta T_b = \frac{\alpha}{\alpha - 1} (T_a - T_b), \tag{9}$$

$$\Delta T_e = -\beta. \tag{10}$$

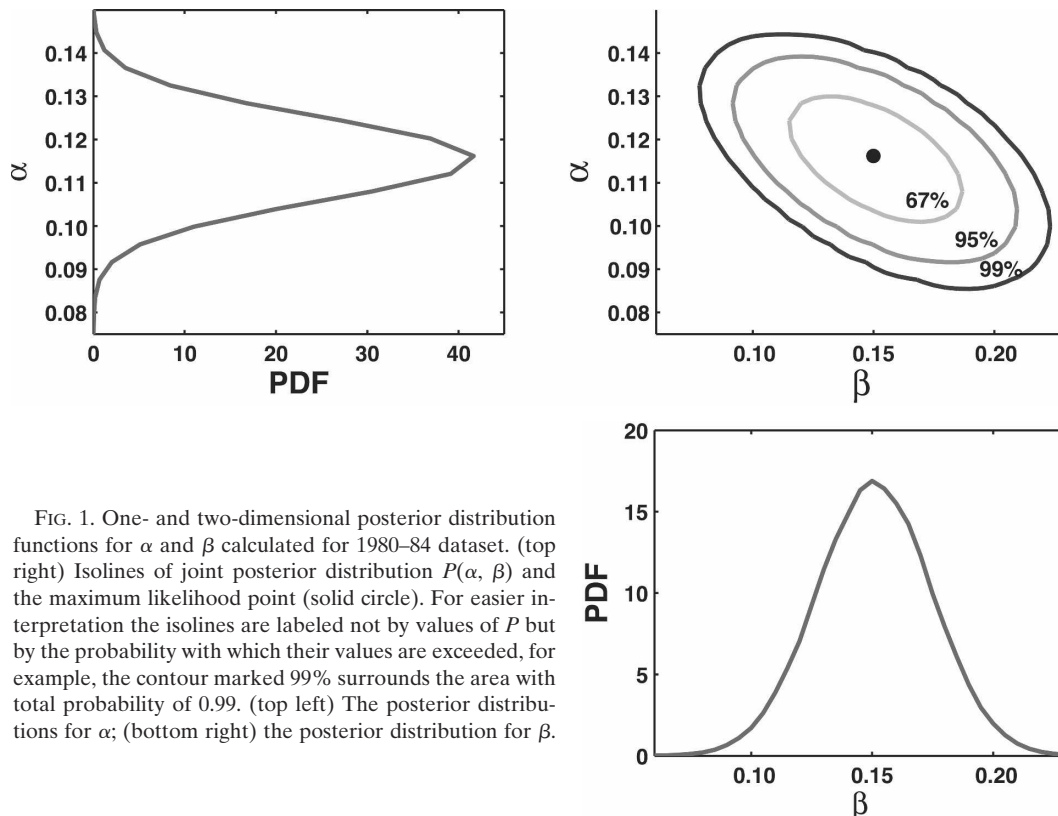


FIG. 1. One- and two-dimensional posterior distribution functions for α and β calculated for 1980–84 dataset. (top right) Isolines of joint posterior distribution $P(\alpha, \beta)$ and the maximum likelihood point (solid circle). For easier interpretation the isolines are labeled not by values of P but by the probability with which their values are exceeded, for example, the contour marked 99% surrounds the area with total probability of 0.99. (top left) The posterior distributions for α ; (bottom right) the posterior distribution for β .

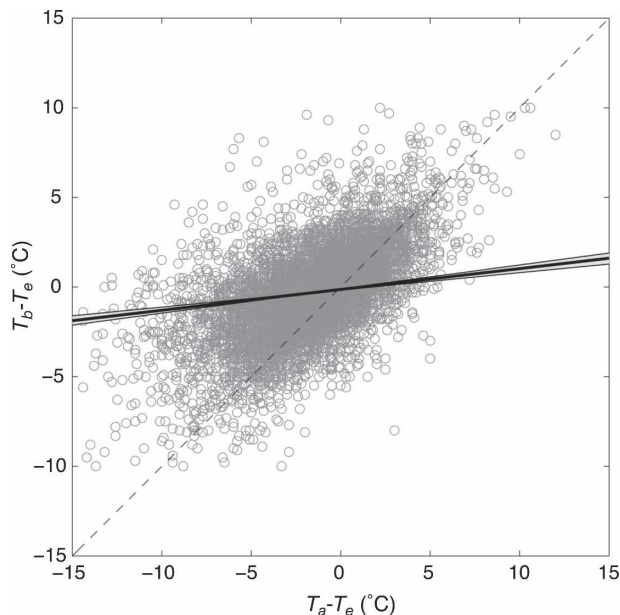


FIG. 2. Difference between collocated SSTs made using different measurement methods (bucket – engine intake; °C) plotted vs air–sea temperature difference (°C) for the period of 1980–84. Collocations are within 100-km separation and at the same reporting hour. The air temperature is taken from the ship in the pair reporting bucket SST. The thick regression line takes into account correlation between errors in two variables and corresponds to the maximum likelihood values for the regression parameters. The 95% confidence intervals for the regression line are shown by shading. The tendency of data to fall along the dashed (1:1) line is because of random errors in T_e and is suppressed by the analysis.

The correction for ICOADS nighttime data is then calculated by Eqs. (9) and (10) for each $2^\circ \times 2^\circ$ area monthly mean grid box and the results are weighted by the proportions of bucket and engine-intake SST in each grid box. Because the measurement method for some of the observations is unknown (Part I), two different correction fields have been calculated. In the calculation for the upper panel (Figs. 4a and 4b) it is assumed that all of the observations with unknown measurement method were made using buckets; in the calculation for the lower panel (Figs. 4c and 4d) it is assumed that all of the observations with unknown measurement method are engine-intake reports. Uncertainties in model parameters have not been accounted for in this simple comparison (but are shown in Fig. 5). Figure 4 shows significant seasonal and regional differences in the calculation. If the SST reports with unknown measurement method are actually bucket reports, then there are significant biases in regions of strong SST gradients and hence air–sea temperature differences. Figure 4a shows SST in the Gulf Stream biased cold by several tenths of a degree Celsius in the

winter months and biased very slightly warm over a more limited region in the summer (Fig. 4b). However, if the reports with unknown method are actually engine-intake reports then most of the North Atlantic region is biased slightly warm in both winter and summer (Figs. 4c and 4d), but with a region of cold bias over the Gulf Stream. This demonstrates the importance of recording methods of observation and the difficulties associated with the interpretation of reports with an unknown method. Figure 5a shows the annual cycle of the correction shown in Fig. 4 averaged over the same North Atlantic region. Our best estimate of the magnitude of the seasonal cycle in the modeled error is 0.08°C if reports of the unknown method are assumed to be from engine intakes and 0.11°C if the reports of the unknown method are from buckets.

The magnitude of the annual cycle in the difference between nighttime gridded $2^\circ \times 2^\circ$ area monthly mean values of engine-intake and bucket SST is shown in Fig. 5b. The solid line shows the annual cycle of the difference between the datasets for the 1980s (similar to Fig. 6b in Part I but for nighttime data only). The dashed line shows the difference between the corrections ($\Delta T_b - \Delta T_e$) for each dataset. The modeled correction differences are slightly larger than the observed values for most of the year. This difference between observed and modeled values falls well within the estimated uncertainty in the modeled correction because of the uncertainty in the model parameters. The values in Fig. 5 are calculated using all nighttime data for which a measurement method is known and therefore include significantly more data than was used in the derivation of the model that required each observation to be matched with a nearby observation made using the alternate method and used only selected data from particular countries.

5. Discussion

The model results confirm that modern insulated buckets lose heat under normal environmental conditions. This tendency has been shown previously. For example, Walden (1966) found an increasing difference between bucket and engine-intake SST with increasing wind speed, which he attributed to evaporation from the bucket or thermometer. Tauber (1969) measured heat loss from the Crawford (1969) bucket. James and Fox (1972) showed that bucket SSTs were colder than engine-intake SSTs, particularly in winter. Differences were greatest at night, at higher wind speeds, and if the bucket observation was made on the windward side of the ship. Parker (1985) and Folland et al. (1993) interpreted maps of SST difference in terms of heating and cooling of the bucket observations, but were hampered

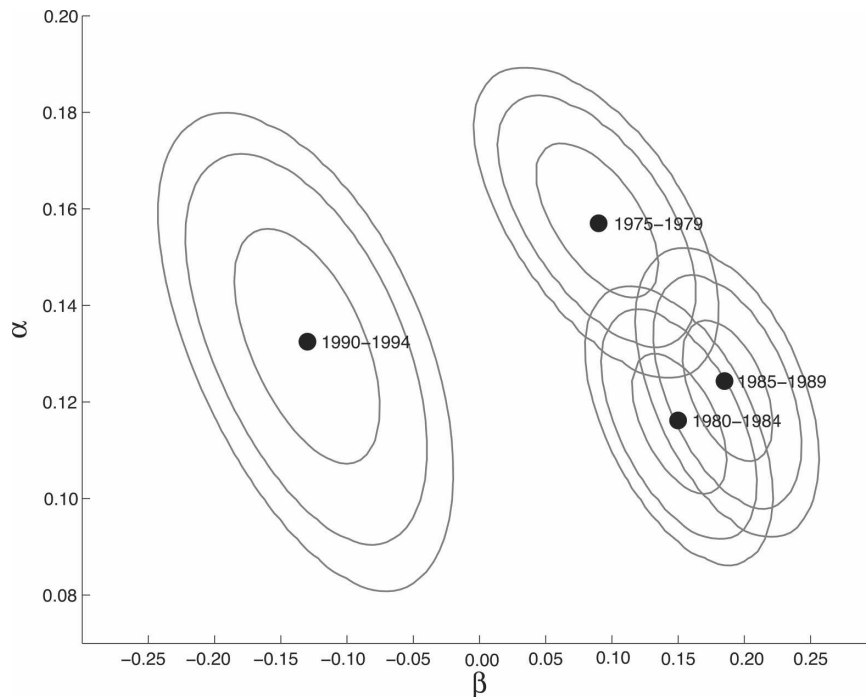


FIG. 3. Maximum likelihood estimates of α and β for four analyzed 5-yr intervals, together with their 67%, 95%, and 99% confidence areas.

by relatively poor knowledge of the measurement method. Kent et al. (1991) showed that bucket SST was cooler than engine-intake SST and became increasingly cool as the sum of the estimated sensible and latent heat fluxes increased. The present study of data collected between 1975 and 1994 has shown that SST measured using buckets of modern design are also subject to heat loss and that this effect can be detected in ICOADS. In addition, a better knowledge of the SST measurement method, resulting from the use of external metadata, has allowed quantification of the effects for a substantial number of VOSSs, rather than for a selected subset of ships. The second result is that engine-intake SST may have been biased slightly warm in the past, but more recent data seem to show a slightly cold bias. Many previous studies have shown a warm bias in engine-intake SST, usually of a larger magnitude than that estimated in the present study (Saur 1963; Walden 1966; Tauber 1969; James and Fox 1972; Kent et al. 1993). James and Fox (1972) showed that the type of thermometer is important, with warm biases of 0.1°C for precision thermometers but significantly larger biases for other types. Saur (1963) found a 0.7°C warm bias in engine-intake SST, but describes conditions of thermometer exposure that would be unlikely on a modern ship. See Part I for a summary of these studies. It seems likely therefore that any warm bias that may

have existed has reduced over time because of better instrumentation and awareness of sources of error in the observation. It is also likely that biases in the engine intake may have been overestimated if heat loss in the buckets used for comparison was not taken into account. A cold bias in the engine-intake SST relative to bucket SST adjusted for heat exchange could be explained if there is a vertical temperature gradient between the surface (sampled by the bucket) and at depth (sampled by the engine intake).

The studies described above all used special comparison datasets. In addition, there are two major studies that attempt to quantify and correct biases in SST data based on climate datasets. FP95 corrected SST observations made prior to 1941. They describe physical models for heat loss from uninsulated (canvas) and partially insulated (wooden) buckets. These models account for direct heat loss, evaporative heat loss, long-wave radiation, solar radiation, the difference between the temperature of the water and the thermometer, and water leakage. They note that many of the details required for the model are variable or unknown including the size of buckets, the mix of bucket types, the length of time of exposure, the speed of the ship, and the instantaneous environmental conditions. A range of assumptions is made for each of the unknowns to make a set of model types. The required correction is then es-

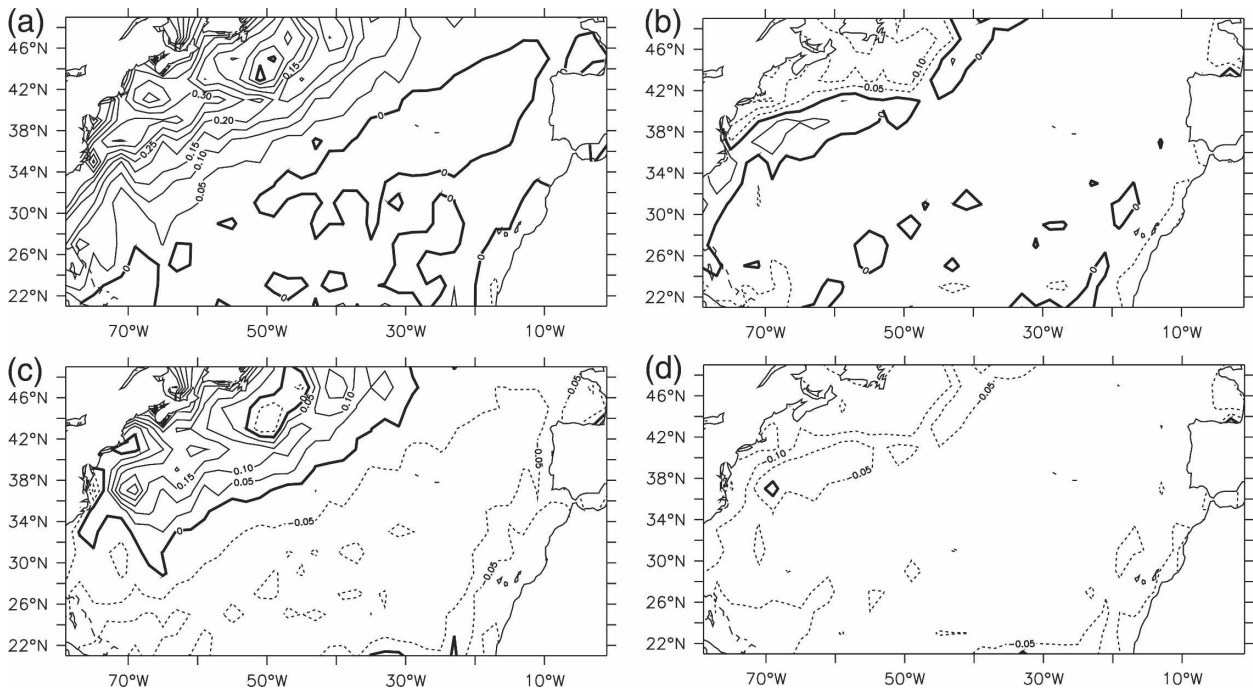


FIG. 4. Maps of North Atlantic region showing the value of the correction based on 1980s average values of α and β and calculated using $2^\circ \times 2^\circ$ area gridded monthly SST and air temperature data from ICOADS for the 1980s. A positive correction indicates that the reported SST is too cold. Contours are in intervals of 0.05°C and a negative correction is indicated by a dotted line. (a) SST correction ($^\circ\text{C}$) for winter [December–February (DJF)] assuming observations of unknown type are made using buckets. (b) SST correction ($^\circ\text{C}$) for summer [June–August (JJA)] assuming observations of unknown type are made using buckets. (c) SST correction ($^\circ\text{C}$) for winter (DJF) assuming observations of unknown type are made using engine intakes. (d) SST correction ($^\circ\text{C}$) for summer (JJA) assuming observations of unknown type are made using engine intakes.

timated using a statistically fitted combination of these model types, which optimize the consistency of the climatological record using the period of 1951–80 as a standard.

Smith and Reynolds (2002, hereafter SR02) took a different approach to the same problem. They assumed that large-scale air–sea interaction was the same for earlier periods as for the period of 1968–97. Coefficients are then derived to correct the SST to match the nighttime air–sea temperature difference of historical periods to the modern data. The coefficients are smoothed to allow small-scale variation in the air–sea temperature differences. SR02 compare their corrections to those of FP95 and conclude that the approximate size and spatial distribution of the corrections are similar. Because the correction models are based on very different assumptions, the difference between the two corrections may give an idea of the uncertainty in the correction. They further conclude that SST bias corrections should be recalculated whenever substantial additions are made to the historical data.

The bias correction of SR02 has the same form as that developed in the present study. From their Fig. 4b a value of α can be deduced to be about 0.2 in 1856,

rising to nearly 0.4 in 1941 when a dramatic change in observation practice occurred (FP95). The rise in their computed value of α between 1856 and 1941 results from a shift from insulated wooden buckets to uninsulated canvas buckets for practical reasons. We might expect that the modern buckets would perform better than the early buckets, so our smaller values of α ranging from 0.12 ± 0.02 to 0.16 ± 0.02 are therefore consistent with the results of the SR02 study obtained for much earlier observations.

Figure 6 compares the nighttime bucket-only correction derived in the present study for the 1980s with the all-hours correction of FP95 averaged over the period of 1856–75. This period has been chosen for comparison because it was dominated by wooden buckets. The wooden buckets are thought to be better insulated than the canvas buckets that gradually replaced them because the wooden buckets were difficult to deploy and it was easy to damage the bucket or to cause damage to the ship. The magnitude of our correction (Fig. 6a) is similar to that of FP95 (Fig. 6b) in this region for these very different periods. Over much of the North Atlantic the correction is $0.1^\circ\text{--}0.15^\circ\text{C}$ rising to a maximum of over 0.35°C in the Gulf Stream region, which is

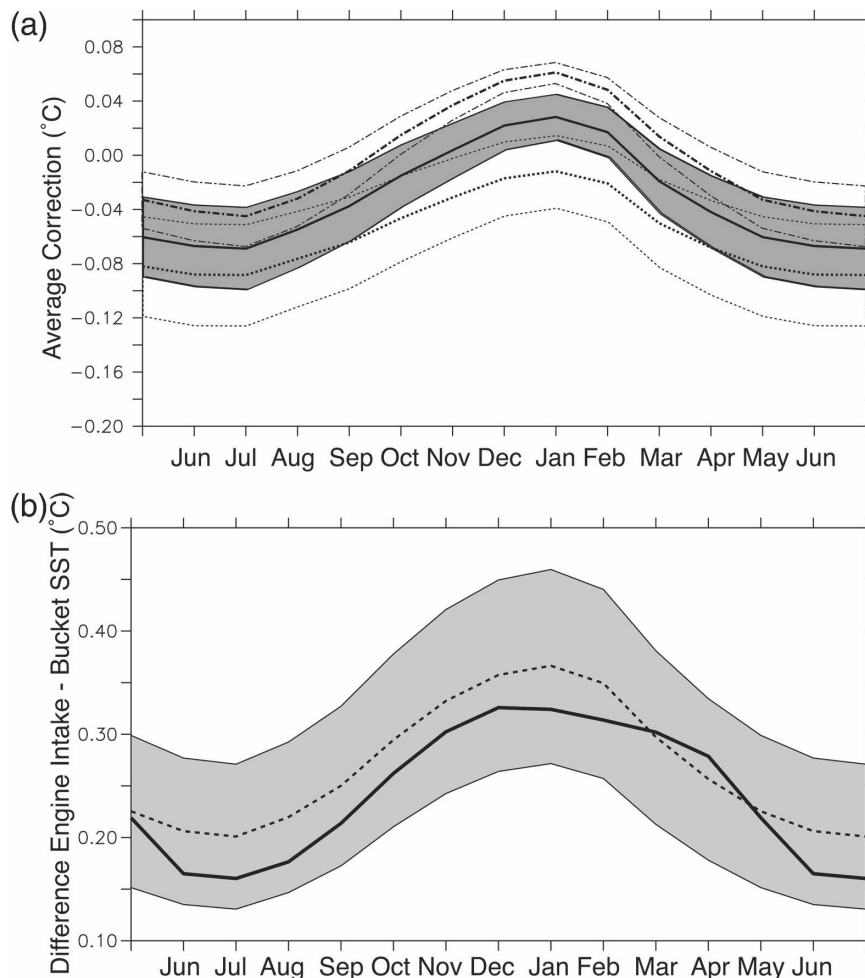


FIG. 5. (a) Average value of correction for the North Atlantic between 20° and 50°N for the 1980s calculated from separate monthly 2° × 2° area gridded datasets of bucket and engine-intake SST and air temperature. The solid line is calculated with the assumption that the reports of unknown method are distributed in the same proportion as those for which the measurement method is known. Gray shading represents 95% uncertainty in the model parameters (Table 1). The three dot-dash lines (MLE and 95% range) are calculated with the assumption that observations of unknown type are actually made using buckets, the dashed lines with the assumption that observations of unknown type are actually made using engine intakes. (b) Estimated difference between engine intake and bucket for the North Atlantic between 20° and 50°N for the 1980s calculated from 2° × 2° area monthly mean gridded datasets as in (a). Actual difference (solid line) and difference calculated using the model parameters from Table 1 (dashed line). Gray shading represents 95% uncertainty in model parameters taken from Table 1.

slightly larger than that of the FP95 correction. The pattern of the corrections predicted by SR02 and FP95 are slightly different—that of SR02 tends to follow contours of sensible heat flux, while that of FP95 tends to follow contours of latent heat flux (SR02). The correction from the present study depends on the air–sea temperature difference and is therefore more similar to the correction of SR02.

FP95 and SR02 both assumed that modern data were

unbiased, FP95 used the period from 1951 to 1980 as a standard and SR02 used the period from 1968 to 1997. Although these periods are not entirely covered by the present study, this work and Part I suggests that an additional correction may be needed for modern data. The correction of the modern data used by FP95 and SR02 as a comparison standard will cause a corresponding increase in the correction required to the early data.

The 1980s is probably the period for which the small-

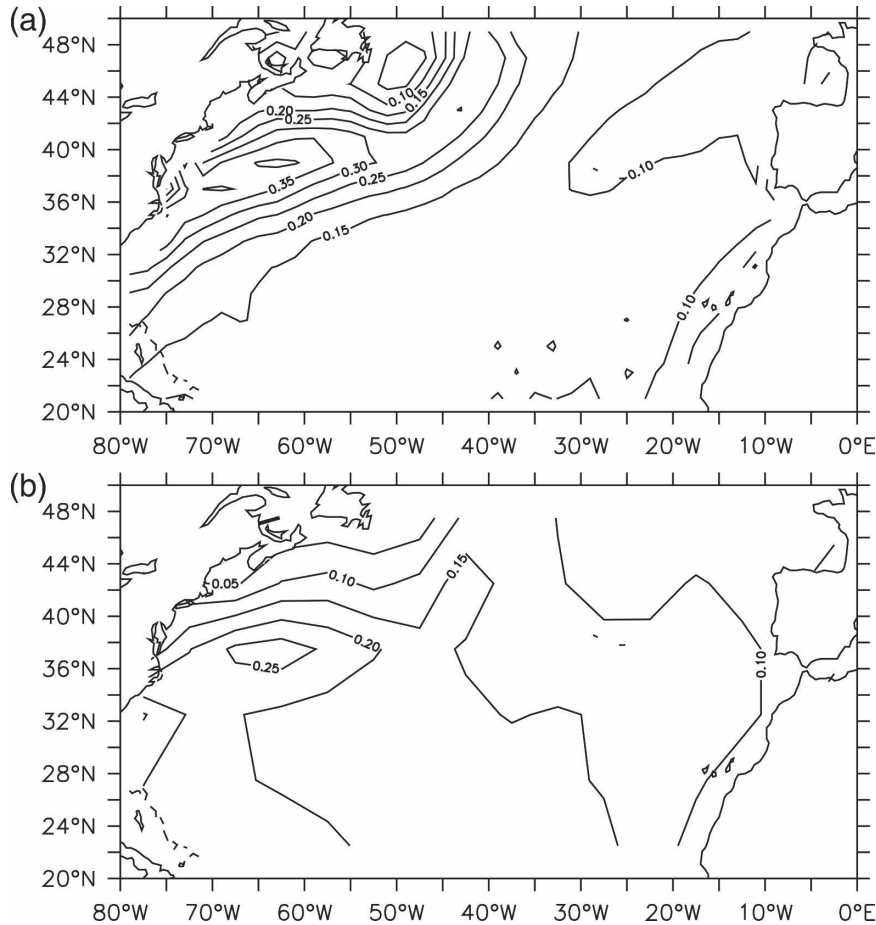


FIG. 6. Comparison of results of FP95 with those from the present study. (a) Correction to insulated buckets derived in present study and applied to data from the 1980s using the same data as in Fig. 5. Smoothing has been applied to the $2^\circ \times 2^\circ$ data to aid the comparison with (b). (b) Correction of FP95 on a $5^\circ \times 5^\circ$ grid, averaged for the period dominated by insulated wooden buckets (1856–75). Contours are in intervals of 0.05°C .

est corrections are required. In the 1970s the biases in the individual data sources are similar to those in the 1980s, but the balance of observational types is likely different (Part I), with a higher percentage of bucket reports, although for many observations the methods used are unknown. In the 1980s the typically cold bias in the bucket reports is offset by the warm bias in the engine-intake reports, leaving small overall biases but with some strong regional signals. In the 1970s it is expected that a higher proportion of reports was made using buckets and therefore the SST measurements were biased colder than in the 1980s. In the 1990s the proportion of engine-intake SST observations increased, but their quality improved. The results of this study suggest that in the 1990s SST observations may be biased cold overall relative to the 1980s. Part I shows the mean difference between SST derived from buckets and that derived from engine intakes in the North At-

lantic (their Fig. 6a). It clearly shows that data from the two sources agree best in the period after 1990. If we believe that the bucket SST is typically biased cold over the entire period shown, then this supports the conclusion drawn from this study that engine-intake SST may, at least in the period of 1990–97, be biased slightly cold.

Figures 4–6 apply the results of the model (1)–(2) to separate bucket and engine-intake $2^\circ \times 2^\circ$ area binned monthly mean datasets derived from nighttime data in ICOADS. However, these datasets contain a significant amount of data not represented in our analysis, which selected only data from certain recruiting countries and restricted the wind speed range. Further, to be included in the samples used to determine the model coefficients (α and β) SST reports were required to have recorded measurement method and to be paired with a nearby report using the alternate method of measurement. We cannot therefore be sure that the data we have used in

our analysis is representative of that in ICOADS. However, the comparison of calculated and modeled differences between bucket- and engine-intake-gridded datasets shows encouraging similarities (Fig. 5b). We must always remember that corrections are not expected to necessarily improve an individual observation, but should, on average, improve the accuracy and consistency of the dataset.

6. Conclusions

There are differences in VOS SST that can be attributed to differences in measurement method. Using a combination of metadata within ICOADS and WMO (1997), comparisons of different measurement methods can be extended back to the 1970s.

It is important to properly account for errors, both random and systematic, when comparing observations. This paper describes a method that accounts for these errors, but uses very simplistic physics. Therefore, to fit the model, only a subset of data is used, comprising nighttime reports excluding very low and very high wind speeds and only from ships recruited by selected countries. However, the results are consistent with expectations of the error derived from the literature, but also indicate possible improvements in the engine-intake SST over time.

The model results suggest that bucket reports of SST made in the North Atlantic at nighttime and moderate wind speeds may be biased by the order of 10% of the air–sea temperature difference in the period of 1975–94. Under similar conditions, the engine-intake SST may be biased warm by 0.1° – 0.2° C for the period between 1975 and 1989. After 1990 the model indicates that engine-intake SST may be biased cold by a similar amount. The timing of this possible transition from a mean warm bias to a mean cold bias is not clear from our analysis.

The need for corrections to the modern data implies an increase in correction to those required for the earlier data. The changing mix of bucket and engine-intake reports in both time and space has important implications for the homogenization of SST. It would therefore be desirable to correct modern data for biases using metadata where available, and then repeat the FP95 and SR02 procedures to reference the early data to an improved modern climatology.

The results of this study are tentative and more work is required to confirm the results and extend conclusions to cases excluded from the present analysis. We have, however, gained some insight into the characteristics of the individual observations. It is desirable that a range of techniques is used to analyze the data, uti-

lizing both the individual observations and climatological analyses. This will be an important future step toward making corrections that can be applied to gridded representations of observational datasets. This work challenges the assumption that engine-intake SST is, on average, biased warm. The presence of biases in observations measured using different measurement techniques has important implications for the detection of climate change as the mix of observational methods continues to evolve.

Acknowledgments. This work was supported by funding from the U.K. Government Meteorological Research Programme. AK was supported by NOAA Grants NA06GP0567 and NA03NES4400012. The authors thank Steven Worley of the National Center for Atmospheric Research Data Support Section for providing the ICOADS and Scott Woodruff of the NOAA–CIRES Climate Diagnostics Center for help and advice on ICOADS. Comments by Simon Tett of the Hadley Centre for Climate Prediction and Research and Michael Evans of the Laboratory of Tree-Ring Research, University of Arizona, resulted in significant improvement of the manuscript.

APPENDIX

Fitting a Straight Line When Errors in Two Variables Are Intercorrelated and Have Uncertainty in Their Covariance Parameters

a. Problem formulation

Our goal is to solve the problem (4)–(8), which is formalized in the following way. There is a “true” linear model,

$$y = \alpha x + \beta(\alpha - 1), \quad (\text{A1})$$

which describes the population whose n samples (x_i, y_i) are observed as (x_i^o, y_i^o) , with $i = 1, \dots, n$. Each observed point (x_i^o, y_i^o) corresponds to a point of true values (x_i, y_i) , lying precisely on the line (A1). The observations differ from the true values by random measurement errors ε_i^x and ε_i^y :

$$x_i^o = x_i + \varepsilon_i^x, \quad y_i^o = y_i + \varepsilon_i^y, \quad i = 1, \dots, n. \quad (\text{A2})$$

Model parameters α and β are to be estimated.

Errors for different points are independent from each other, but each individual point errors in the x and y directions are not independent. More precisely, the vector of errors $\xi_i = (\varepsilon_i^x, \varepsilon_i^y)^T$ (superscript T denotes transposition hereafter) for the i th observed point (x_i, y_i) is assumed to have the bivariate Gaussian distribution

$$\xi_i \sim \mathcal{N}(0, \mathbf{C}_i), \quad (\text{A3})$$

with $(0, 0)^T$ mean and covariance matrix

$$\mathbf{C}_i = \begin{pmatrix} c_i^a + c_i^e & c_i^e \\ c_i^e & c_i^b + c_i^e \end{pmatrix}$$

[cf. Eq. (7)]. Error parameters c_i^b , c_i^e , c_i^a have the Gaussian prior distributions

$$c_i^b \sim \mathcal{N}(c_0^b, s^b), \quad c_i^e \sim \mathcal{N}(c_0^e, s^e), \quad c_i^a \sim \mathcal{N}(c_0^a, s^a).$$

b. Maximum likelihood approach

Based on (A1)–(A3), the likelihood function for the observations (their conditional probability given all other parameters) can be written as

$$L = \prod_{i=1}^n p(x_i^o, y_i^o | x_i, \alpha, \beta, \mathbf{C}_i),$$

where

$$p(x_i^o, y_i^o | \alpha, \beta, x_i, \mathbf{C}_i) = \frac{1}{2\pi\sqrt{\det\mathbf{C}_i}} \exp\left(-\frac{1}{2} \delta_i^T \mathbf{C}_i^{-1} \delta_i\right),$$

with

$$\delta_i = [x_i - x_i^o, \alpha x_i + \beta(\alpha - 1) - y_i^o]^T.$$

Prior probability for the parameters α , β , x_i , \mathbf{C}_i , $i = 1, \dots, n$ is based on the given above Gaussian distributions for c_i^b , c_i^e , c_i^a and noninformative priors for α , β , x_i :

$$P_{\text{pr}} = \prod_{i=1}^n \frac{1}{(2\pi)^{3/2} s^b s^e s^a} \exp\left\{-\frac{1}{2} \left[\left(\frac{c_i^b - c_0^b}{s^b} \right)^2 + \left(\frac{c_i^e - c_0^e}{s^e} \right)^2 + \left(\frac{c_i^a - c_0^a}{s^a} \right)^2 \right]\right\}.$$

Multiplying the likelihood function by the prior probability, we obtain the joint probability distribution function of all parameters and observations:

$$P = LP_{\text{pr}} \propto \prod_{i=1}^n P_i = \prod_{i=1}^n \frac{1}{\sqrt{\det\mathbf{C}_i}} \exp\left\{-\frac{1}{2} \left[\delta_i^T \mathbf{C}_i^{-1} \delta_i + \left(\frac{c_i^b - c_0^b}{s^b} \right)^2 + \left(\frac{c_i^e - c_0^e}{s^e} \right)^2 + \left(\frac{c_i^a - c_0^a}{s^a} \right)^2 \right]\right\}. \quad (\text{A4})$$

The maximum likelihood estimation of α and β requires maximizing P over all the “nuisance” parameters x_i , c_i^b , c_i^e , c_i^a , $i = 1, \dots, n$. The resulting function of α and

β approximates their posterior distribution in the vicinity of the probability maximum (e.g., Sivia 1996).

c. Solution

Maximization of P over nuisance parameters can be done independently for each factor P_i in Eq. (A4). Note that each factor

$$P_i = \frac{1}{\sqrt{\det\mathbf{C}_i}} \exp\left\{-\frac{1}{2} \left[\delta_i^T \mathbf{C}_i^{-1} \delta_i + \left(\frac{c_i^b - c_0^b}{s^b} \right)^2 + \left(\frac{c_i^e - c_0^e}{s^e} \right)^2 + \left(\frac{c_i^a - c_0^a}{s^a} \right)^2 \right]\right\} \quad (\text{A5})$$

only depends on the parameters corresponding to the i th observational point.

Because only the term $\delta_i^T \mathbf{C}_i^{-1} \delta_i$ in P_i depends on x_i , the minimization of P_i over x_i can be easily performed analytically. Introducing

$$x_i' = x_i - x_i^o, \quad r_i = y_i^o - \alpha x_i^o - \beta(\alpha - 1),$$

$$\mathbf{h}_1 = \begin{pmatrix} 1 \\ \alpha \end{pmatrix}, \quad \mathbf{h}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

so that

$$\delta_i = x_i' \mathbf{h}_1 - r_i \mathbf{h}_2,$$

we rewrite

$$\delta_i^T \mathbf{C}_i^{-1} \delta_i = (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1) [x_i' - r_i (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2) / (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1)]^2 + r_i^2 [(\mathbf{h}_2^T \mathbf{C}_i^{-1} \mathbf{h}_2) - (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2)^2 / (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1)].$$

Obviously, value $x_i' = r_i (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2) / (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1)$ corresponding to the maximum likelihood estimate of x_i

$$\hat{x}_i = x_i^o + r_i (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2) / (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1)$$

minimizes $\delta_i^T \mathbf{C}_i^{-1} \delta_i$. Note that

$$\begin{aligned} & (\mathbf{h}_2^T \mathbf{C}_i^{-1} \mathbf{h}_2) - (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2)^2 / (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1) \\ &= \frac{(\mathbf{h}_2^T \mathbf{C}_i^{-1} \mathbf{h}_2)(\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1) - (\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_2)^2}{\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1} \\ &= \frac{\det(\mathbf{H}^T \mathbf{C}_i^{-1} \mathbf{H})}{\mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1} = \frac{1}{\det\mathbf{C}_i \mathbf{h}_1^T \mathbf{C}_i^{-1} \mathbf{h}_1} \\ &= \frac{1}{(c_i^a + c_i^e)\alpha^2 - 2c_i^e\alpha + c_i^b + c_i^e}, \end{aligned}$$

where we introduced $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2]$, and made use of $\det\mathbf{H} = 1$, as well as

$$\mathbf{C}_i^{-1} \det\mathbf{C}_i = \begin{pmatrix} c_i^b + c_i^e & -c_i^e \\ -c_i^e & c_i^a + c_i^e \end{pmatrix}.$$

Therefore,

$$\min_{x_i} \delta_i^T \mathbf{C}_i^{-1} \delta_i = \frac{r_i^2}{(c_i^a + c_i^e)\alpha^2 - 2c_i^e\alpha + c_i^b + c_i^e}.$$

Because the minimum of $\delta_i^T \mathbf{C}_i^{-1} \delta_i$ in (A5) corresponds to the maximum of P_i

$$\max_{x_i} P_i = \frac{1}{\sqrt{\det \mathbf{C}_i}} \exp \left\{ -\frac{1}{2} \left[\frac{r_i^2}{(c_i^a + c_i^e)\alpha^2 - 2c_i^e\alpha + c_i^b + c_i^e} + \left(\frac{c_i^b - c_o^b}{s^b} \right)^2 + \left(\frac{c_i^e - c_o^e}{s^e} \right)^2 + \left(\frac{c_i^a - c_o^a}{s^a} \right)^2 \right] \right\}.$$

Maximizing the latter over c_i^b , c_i^e , and c_i^a appears analytically intractable, thus it is done numerically.

Rewrite the last equation as

$$\max_{x_i} P_i = \exp(-S_i/2),$$

where

$$S_i = \frac{r_i^2}{(c_i^a + c_i^e)\alpha^2 - 2c_i^e\alpha + c_i^b + c_i^e} + \left(\frac{c_i^b - c_o^b}{s^b} \right)^2 + \left(\frac{c_i^e - c_o^e}{s^e} \right)^2 + \left(\frac{c_i^a - c_o^a}{s^a} \right)^2 + \ln(c_i^a c_i^b + c_i^b c_i^e + c_i^e c_i^a).$$

The maximum of P_i is achieved at the minimum of S_i . The latter can be easily minimized over c_i^b , c_i^e , and c_i^a using the gradient descent method. To facilitate the computation of values of P_i maximized over these parameters, we observe that the minimum value of S_i only depends on α and r_i^2 , and that this functional dependence is the same for all i . Therefore, we tabulate the function

$$F(\alpha, R) = \min_{c^b, c^e, c^a} \left[\frac{R}{(c^b + c^e)\alpha^2 - 2c^e\alpha + c^a + c^e} + \left(\frac{c^b - c_o^b}{s^b} \right)^2 + \left(\frac{c^e - c_o^e}{s^e} \right)^2 + \left(\frac{c^a - c_o^a}{s^a} \right)^2 + \ln(c^a c^b + c^b c^e + c^e c^a) \right].$$

The gradient descent is used to find the minimum for each given pair of α and R . For the applications presented in this work and summarized in Table 1, no more than a few hundred iterations were necessary to determine the minimizing values of c^b , c^e , and c^a with the accuracy of $0.001(^{\circ}\text{C})^2$. When F is available, the minimum value of S_i can be found as

$$\min_{c_i^b, c_i^e, c_i^a} S_i = F(\alpha, r_i^2).$$

Finally, the maximum likelihood approximation of the joint posterior distribution for α and β , obtained as a maximum of P over all nuisance parameters, is computed by

$$P(\alpha, \beta) \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n F[\alpha, (y_i^o - \alpha x_i^o - \beta(\alpha - 1))^2] \right\}. \tag{A6}$$

Maximum likelihood approximations of posterior distributions for α and β can be found by further maximizations:

$$P(\alpha) \propto \max_{\beta} P(\alpha, \beta), \tag{A7}$$

$$P(\beta) \propto \max_{\alpha} P(\alpha, \beta). \tag{A8}$$

Modes of these distributions give maximum likelihood estimates for α and β . The distributions (A7) and (A8) are also used to obtain 95% confidence intervals for these parameters reported in Table 1, along with their maximum likelihood estimates.

REFERENCES

- 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.
- , —, and P. K. Taylor, 2004: An analytical model of heating errors in marine air temperatures from ships. *J. Atmos. Oceanic Technol.*, **21**, 1198–1215.
- Boggs, P. T., and J. E. Rogers, 1990: Orthogonal distance regression. *Statistical Analysis of Measurement Error Models and Their Applications*, P. J. Brown and W. A. Fuller, Eds., Vol. 112, *Contemporary Mathematics*, American Mathematical Society, 183–194.
- Crawford, A. B., 1969: Sea-surface temperatures: Some instruments, methods and comparisons. World Meteorological Organization Tech. Note 103, WMO 247.TP.135, 151 pp.
- Deming, W. E., 1964: *Statistical Adjustment of Data*. Dover, 261 pp.
- Diaz, H., C. Folland, T. Manabe, D. Parker, R. Reynolds, and S. Woodruff, 2002: Workshop on advances in the use of historical marine climate data. *WMO Bull.*, **51**, 377–380.
- Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis*. John Wiley and Sons, 603 pp.
- Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319–367.
- , R. W. Reynolds, M. Gordon, and D. E. Parker, 1993: A study of six operational sea surface temperature analyses. *J. Climate*, **6**, 96–113.
- James, R. W., and P. T. Fox, 1972: Comparative sea surface temperature measurements. Reports on Marine Science Affairs No. 5, WMO336, 27 pp.
- Kent, E. C., and P. K. Taylor, 1991: Ships observing marine climate: A catalogue of the voluntary observing ships participating in the VSOP-NA. Marine Meteorology and Related Oceanographic Activities Rep. 25, WMO Tech. Doc. 456, 123 pp.

- , and —, 1997: Choice of a Beaufort equivalent scale. *J. Atmos. Oceanic Technol.*, **14**, 228–242.
- , and D. I. Berry, 2005: Quantifying random measurement errors in voluntary observing ships meteorological observations. *Int. J. Climatol.*, **25**, 843–856.
- , and P. G. Challenor, 2006: Toward estimating climatic trends in SST. Part II: Random errors. *J. Atmos. Oceanic Technol.*, **23**, 476–486.
- , and P. K. Taylor, 2006: Toward estimating climatic trends in SST. Part I: Methods of measurement. *J. Atmos. Oceanic Technol.*, **23**, 464–475.
- , P. G. Challenor, and P. K. Taylor, 1999: A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *J. Atmos. Oceanic Technol.*, **16**, 905–914.
- , B. S. Truscott, J. S. Hopkins, and P. K. Taylor, 1991: The accuracy of ship's meteorological observations: Results of the VSOP-NA. Marine Meteorology and Related Oceanographic Activities Rep. 26, WMO Tech. Doc. 455, 86 pp.
- , P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993: The accuracy of voluntary observing ship's meteorological observations: Results of the VSOP-NA. *J. Atmos. Oceanic Technol.*, **10**, 591–608.
- , —, and P. G. Challenor, 1998: A comparison of ship and scatterometer-derived wind speed data in open ocean and coastal areas. *Int. J. Remote Sens.*, **19**, 3361–3381.
- Lindau, R., 1995: A new Beaufort equivalent scale. *Proc. Int. COADS Winds Workshop*, Kiel, Germany, Institut für Meereskunde and NOAA, 232–252.
- National Geophysical Data Center, 1993: Digital relief of the surface of the earth. NOAA Data Announcement 93-MGG-01, 1 pp.
- Parker, D. E., 1985: A comparison of bucket and non-bucket measurements of sea surface temperature. Meteorological Office, Met O 13 Branch Memo., 16 pp.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in C: The Art of Scientific Computing*. 2d ed. Cambridge University Press, 1020 pp.
- Roll, H. U., 1951: Water temperature measurements on deck and in the engine room. *Ann. Meteor.*, **4**, 439–443.
- Saur, J. F. T., 1963: A study of the quality of sea water temperatures reported in the logs of ships' weather observations. *J. Appl. Meteor.*, **2**, 417–425.
- Sivia, D. S., 1996: *Data Analysis: A Bayesian Tutorial*. Clarendon Press, 189 pp.
- Smith, T. M., and R. W. Reynolds, 2002: Bias corrections for historical sea surface temperatures based on marine air temperatures. *J. Climate*, **15**, 73–87.
- Tauber, G. M., 1969: The comparative measurements of sea surface temperature in the U.S.S.R. World Meteorological Organization Tech. Note 103, WMO 247.TP.135, 151 pp.
- Thomas, B. R., E. C. Kent, and V. R. Swail, 2005: Methods to homogenize wind speeds from ships and buoys. *Int. J. Climatol.*, **25**, 979–995.
- Walden, H., 1966: On water temperature measurements aboard merchant vessels (in German). *Dtsch. Hydrogr. Z.*, **19**, 21–28.
- WMO, 1997: International list of selected, supplementary and auxiliary ships. WMO Rep. 47.
- Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth*, **23**, 517–527.