

Toward Estimating Climatic Trends in SST. Part I: Methods of Measurement

ELIZABETH C. KENT AND PETER K. TAYLOR

National Oceanography Centre, Southampton, United Kingdom

(Manuscript received 24 February 2004, in final form 17 August 2004)

ABSTRACT

To assess climatic changes in sea surface temperature (SST), changes in the measurement method with time and the effect of these changes on the mean SST must be quantified. Observations from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) have been analyzed for the period from 1970 to 1997 using both SST measurement metadata contained within the dataset and a World Meteorological Organization (WMO) catalog of observing ships. The WMO metadata were particularly important in identifying engine-intake SSTs during the 1970s, but increased method identification over the entire period. There are strong regional variations in the preferred SST measurement method, with engine-intake SST most common in the Pacific and bucket SST preferred by countries bordering the Atlantic. The number of engine-intake SSTs increases over time and becomes more numerous than buckets by the early 1980s.

There are significant differences between SST observations made by different methods. The rounding of reports is more common for engine-intake SST than for either bucket or hull sensor SST, which degrades its quality. Significant time-varying biases exist between SST derived from buckets and from engine intakes. The SST difference has a strong seasonal signal with bucket SST being relatively cold in winter, probably resulting from heat loss from the buckets, and warm in summer, probably resulting from solar warming or the sampling of a shallow warm layer. There is also a long-term trend with engine-intake SST being relatively warm in the early period but with a small annual mean difference between the two methods by 1990.

1. Introduction

Observations of in situ SST have an important role to play in the monitoring of climate change. The Intergovernmental Panel on Climate Change (IPCC) Third Assessment (Houghton et al. 2001) shows trends in global surface temperature rising significantly over recent decades with a change of about 0.4°C between 1970 and 2000. Over the ocean, the surface temperature is derived from Met Office SST products [Hadley Centre Global Sea Surface Temperature Version 1 (HadSST1; Jones et al. 2001; Rayner et al. 2003), and Met Office Historical Sea Surface Temperature (MOHSST; Parker et al. 1995)], which use observations of SST from merchant ships participating in the World Meteorological Organization (WMO) Voluntary Observing Ship (VOS) program. Prior to 1942 corrections have been applied to the SST data in an attempt to homog-

enize the observations and reduce time-varying biases resulting from changing observing practices (Folland and Parker 1995). These corrections increased from about 0.1°C in 1850 to 0.4°C in 1940. However, it was then assumed that the needed correction became negligible, initially because of the changes in observation method enforced by wartime conditions, and after about 1950 because of the mix of measurement methods employed. Thus, while it is known that more recent SST observations also contain errors resulting from different observing methods (e.g., James and Fox 1972; Kent et al. 1993; Folland et al. 1993), no corrections have been applied in calculating climatic trends. To quantify any resulting errors it is necessary to identify the methods used by the VOS to measure SST and to quantify biases for each measurement method as a function of environmental conditions and observing practice.

This paper is one of a series of three and will address the determination of measurement methods for VOS data from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Woodruff et al. 1998; Diaz et al. 2002) using metadata from the WMO “List

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

of selected, supplementary and auxiliary ships” (e.g., WMO 1997). Section 2 describes the different measurement methods used by the VOS and reviews studies of their accuracy. Section 3 describes the data and metadata sources and how the measurement methods vary in space and time. Section 4 presents the characteristics of the two main types of observation, buckets, and engine intakes. Section 5 contains a discussion of the results. Subsequent papers will present estimates for random errors calculated for different observing methods (Kent and Challenor 2006, hereafter Part II) and a statistical method utilizing these random errors to estimate biases in the different observing methods (Kent and Kaplan 2006, hereafter Part III).

2. SST measurement methods used by the VOS

a. Introduction

Folland and Parker (1995) have summarized the methods used up to 1941. It was assumed that almost all SST data were derived from bucket samples, with wooden buckets being generally replaced by “uninsulated” canvas buckets as time progressed. During the Second World War it was assumed that readings of the engine room–intake temperature became the dominant method because of the danger of taking bucket measurements, particularly if a light had to be used at night. In postwar years a mix of engine room intake, uninsulated, and more modern insulated buckets was assumed.

There have been many studies over the years comparing the various measurement methods. Brooks (1926) compared nearly simultaneous canvas bucket and engine-intake SST reports both in the wintertime West Indies, and from an ice patrol ship. He found problems with the engine-intake SST resulting from heating by the engines, parallax errors in reading the temperature, and time delays between the reading and the report. However, the buckets were usually biased cold because of heat loss. It was concluded that the problems with the engine intake were more tractable, and that great care was needed to make good quality bucket SST measurements. Roll (1951) also examined simultaneous bucket and engine-intake SST, in the North and Norwegian Seas. He developed a correction for the observed heat loss by the buckets following wind tunnel experiments. Kirk and Gordon (1952) describe comparisons of canvas bucket and engine-intake SST from Netherlands merchant ships and British Ocean Weather Ships. They found evidence for both the evaporative cooling and solar warming of buckets, and for better agreement between observations when the engine intake was shallow and the flow rate higher.



FIG. 1. Examples of SST buckets in current use by the VOS of (left) the United Kingdom, (middle) the Netherlands, and (right) Germany.

Saur (1963) compared simultaneous bucket and engine-intake SST on 12 U.S. Navy ships in the Pacific. A specially designed insulated bucket was used as a standard and large errors in the intake temperature were found. These errors could be attributed to the use of poorly sited mercury thermometers, causing heating and parallax errors, and to time lags in reporting the temperature, which was logged in the engine room, to the bridge. Stevenson (1964) studied SSTs measured on a moving research ship and noted that bucket temperatures were more consistent with bathythermograph measurements at 15–20 ft than with SSTs from a near-surface probe deployed well ahead of the bow. Bucket SST measured from the bridge deck was significantly different from that measured from the work deck. Walden (1966) analyzed simultaneous bucket and engine intake made on German merchant ships in the early 1960s. The type of German bucket with an integral thermometer shown in Fig. 1 was in use before 1960 (V. Wagner 2003, personal communication) and so the buckets were probably insulated. However, evidence for both cooling and solar heating of the buckets as well as for heating of the engine room–intake SST was found. Crawford (1969) describes engine room–intake thermometers with graduations of 5°C and time lags of up to several hours. Tauber (1969) compared bucket and engine-intake SSTs from Russian vessels with higher-quality measurements and showed cooling of insulated buckets and large errors in engine-intake SST, often related to the operating conditions in the engine room. James and Fox (1972) made a comprehensive comparison of a global dataset of deep-water simultaneous bucket and engine-intake SSTs measured on merchant ships from several countries. They found differences that depended on the time of day, ship size, depth of intake, distance inboard of the intake thermometer, wind speed, air–sea temperature difference, precipitation, the type of intake thermometer, the type

of bucket, and the side of the ship on which the observation was taken. Tabata (1978) compared SST measured on a Canadian research ship using a bucket, engine intake, thermograph, expendable bathythermograph, and oceanographic station data. Using the oceanographic data as a standard, bucket SSTs were found to be biased about 0.1°C warm, engine-intake SST was an order of magnitude more scattered than the other methods and biased 0.3°C warm.

Parker (1985) and Folland et al. (1993) generated seasonal maps of bucket minus nonbucket differences from monthly $5^{\circ} \times 5^{\circ}$ climatologies where there were sufficient data to be confident in the comparison. They found that on average buckets were 0.1°C colder than nonbuckets, and regional and seasonal differences were larger. It was known that the "nonbucket or unknown" category contained a significant number of bucket reports and therefore 0.1°C is likely to be an underestimate of any difference between SSTs from the different measurement methods.

Some information about a subset of the VOS was collected during the Voluntary Observing Ships' Special Observing Project for the North Atlantic (VSOP-NA) project (Kent and Taylor 1991). The ships participating in the project were probably typical of the larger VOSs (although some smaller research vessels participated in the VSOP-NA) and showed that the types of instruments used, observational practice, and environmental conditions could all impact the quality of observations.

Because the results of many of these studies are only published in report form, in the following sections we will summarize the characteristic errors of the different measurement methods.

b. Bucket measurements of SST

To make the measurement the SST bucket is thrown over the side of the ship into the water and a sample of seawater is hauled onto the deck. The temperature of the water sample is then measured, usually with a mercury-in-glass thermometer. The type of bucket used to make SST observations is likely to have changed over the period of 1970–97. In the early 1970s some buckets would have been made of canvas; in the James and Fox (1972) study nearly 10% of the observations came from canvas buckets. Later, most of the buckets are likely to be insulated buckets such as those shown in Fig. 1, which are currently in use by VOSs of Germany, the Netherlands, and the United Kingdom, and are largely made of rubber. The U.K. bucket also has a double-walled construction to hold a sleeve of the seawater around the water sample. The buckets used have very different constructions and volumes (Fig. 1).

The bucket observation of SST is expected to be prone to several sources of error. The sample in the bucket can cool by evaporation from the top surface or from the walls of the bucket in dry or windy conditions (Parker 1985). If the air–sea temperature difference is large the sample may be directly cooled, or less usually, warmed. These effects will be large for canvas buckets (Brooks 1926; James and Fox 1972; Folland and Parker 1995). However, cooling is still detectable for insulated buckets (Roll 1951; Walden 1966; Tauber 1969; James and Fox 1972; Kent et al. 1993). Roll (1951) reported the results of wind tunnel studies using a bucket similar to the modern German bucket, which showed cooling of the bucket SST by 1.5°C in 1 min when the wind speed was greater than 18 m s^{-1} and the air–sea temperature difference was -10°C . The Crawford bucket (Crawford 1969) cooled by 0.2°C in 3 min when the air–sea temperature difference was from -3° to -4°C (Tauber 1969). James and Fox (1972) showed that bucket observations made when the air–sea temperature was greater than 3° or less than -9°C were strongly biased. There was little variation of the difference with increasing wind speed but bucket SSTs were 0.4°C cooler than engine-intake SSTs in the highest wind speed category.

If the temperature of the bucket itself is very different from that of the SST and the bucket is not immersed in the sea long enough to reach equilibrium, the temperature of the water sample will be affected (Parker 1985). This will occur if it has been particularly cold or hot on deck, or when the solar radiation is strong and the bucket has not been stored in the shade. The effect will be larger when conditions do not allow the bucket to be properly immersed in the water (when the ship is moving fast or in strong winds). In a guide for U.K. observers (Meteorological Office 1956) it is suggested that if an initial sample is taken and the canvas bucket and thermometer are allowed to equilibrate before quickly taking another sample then errors should be about $\pm 0.1^{\circ}\text{C}$. However, it is not known how often this technique was used.

The water sample or thermometer may be heated by solar radiation in sunny conditions (Walden 1966) or by direct heating if the air temperature is greater than the SST (Parker 1985). If the thermometer is removed from the sample for reading it will cool by evaporation (Walden 1966). Errors will then occur if the wet-bulb depression is large and may occur more often at lower SSTs because it can be difficult to read the bottom of the scale while the thermometer is immersed in the sample. For buckets with a fixed integral thermometer (e.g., German-issued buckets), on occasion there may not be enough water in the bucket to cover the ther-

momometer bulb. The SST then reported will be that of the air or wet-bulb temperature. Again, this will be more of a problem when deploying the bucket is difficult.

Bucket measurements of SST tend to sample the near-surface waters. This is particularly true when the bucket is deployed from large or fast-moving ships and in high wind speed conditions. Under these conditions it is difficult to immerse the bucket, which can bounce along the sea surface. Differences between bucket and other measurement techniques might therefore occur because of the formation of a diurnal warm layer. The bucket will preferentially sample this near-surface warm layer; the other methods will sample cooler water below. The bucket SST will therefore be relatively warm when low wind speed is combined with strong solar radiation (Kirk and Gordon 1952). However, James and Fox (1972) suggested that other sources of difference between the methods were larger than the ocean vertical gradients.

James and Fox (1972) concluded that precipitation may cool bucket measurements of SST by an amount depending on the water temperature; the amount, type, and temperature of the precipitation; water stability; and mixing, but it is difficult to separate the effects of precipitation itself from the conditions that normally accompany it.

c. Engine-intake measurements of SST

The engine-intake measurement of SST is a report of the temperature of the pumped seawater used to cool the engines and is not a dedicated method of measurement. It is therefore likely to be made near to the ships' engines, whereas for an accurate determination of the SST a measurement close to the seawater inlet is preferable. A guide for U.K. observers (Meteorological Office 1956) describes the preferred method of taking an engine-intake SST with a regularly calibrated thermometer permanently mounted in the intake pipe close to the ship's side. It is not clear how often a described alternative method (holding a thermometer under a tap from the intake) was used.

There will have been changes in the method of engine-intake SST measurement over the period of this study. Early engine-intake temperatures may have been taken with a mercury thermometer in a well in the intake pipe or from a dial with temperature intervals of several degrees. Engine-intake temperatures are now more likely to be taken with an electrical thermometer. On the U.S. Navy ships studied by Saur (1963), the engine-intake thermometer was a mercury thermometer protected by a metal sleeve inserted into a well in the engine-intake pipe. There were examples of incrus-

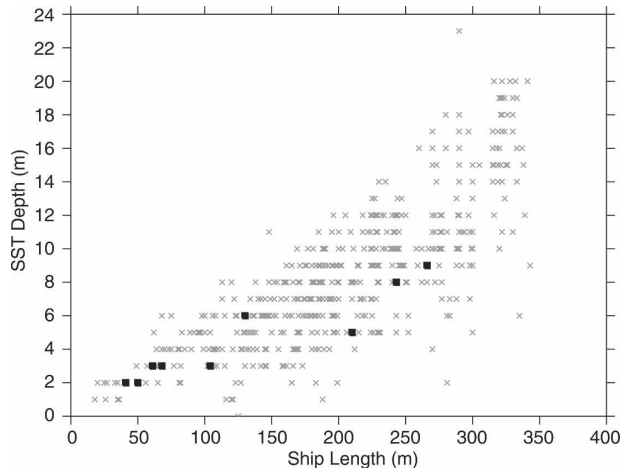


FIG. 2. Depth of the SST measurement against ship length for 1997 from WMO metadata, with engine-intake (crosses) and hull sensor (squares) measurement methods.

tation and fouling, poor exposure of the thermometer well to the flow in the pipe, air pockets inside the thermometer well, and heat conduction along metal supports to the thermometer bulb. All of these could give rise to errors in engine-intake measurements. The type of thermometer used for the intake measurement is significant (James and Fox 1972). For precision thermometers and thermistor probes the bucket – engine-intake SST ΔT was -0.1°C , compared to -0.3°C for mercury thermometers and -0.6°C for “other” types. We have no information on how often engine-intake thermometers are calibrated. Some engine-intake thermometer scales are hard to read with an accuracy of better than a degree. Graduations can be up to 5°C and parallax errors can be large (Brooks 1926; Crawford 1969).

It is expected that the water sampled by the engine intake will be at a greater depth (and therefore possibly cooler) than that sampled by the bucket (Kirk and Gordon 1952; Walden 1966; Kent and Taylor 1991). Deep measurement is usually required because the inlet must be below the waterline, whatever the ships' loading. The ships using engine intakes in the VSOP-NA project had inlets at a depth between 1 and 9 m below the waterline (Kent and Taylor 1991). Figure 2 shows the depth of the SST sensor derived from WMO (1997) as a function of ship length. Only a few countries have reported SST depth because this field was introduced in 1995. Most of the depth information comes from ships from Australia, France, Hong Kong, and Japan. This shows that some of the intakes are much deeper than those of the VSOP-NA ships using engine intakes, which were mainly from the United States. In 1997 the depths range from 1 to 26 m with an average of 8.4 m

and a standard deviation of 4.1 m. Figure 2 shows that the maximum possible depth can be estimated from the ship length but for each ship length there is a wide range of depths. James and Fox (1972) show that intakes at 7-m depth or less showed a bias of 0.2°C and deeper inlets had biases of 0.6°C.

James and Fox (1972) stress the importance of positioning intake thermometers close to the hull. The engine-intake SST could be biased warm if it is made a significant distance inboard or close to the ships' engines (Brooks 1926; Saur 1963; Walden 1966). If the intake thermometer was 3 m or less from the intake the average bias with respect to the bucket data was 0.2°C, increasing to 0.8°C for thermometers further from the intake. The error may be larger for larger ships (James and Fox 1972; Kent et al. 1993), although the water flow rate and depth of sampling may increase on larger ships, which could reduce any heating effect. Errors in engine-intake SST may depend strongly on the operating conditions in the engine room (Tauber 1969), and the circulation of the water may stop if the ship is stationary leading to large errors. James and Fox (1972) confirmed that the size of the ship and the distance inboard are more important than the vertical thermal gradient in the ocean.

In the earlier period, the procedure for making an engine-intake SST report was found to be variable. Unknown time lags were present between the logging of the temperature in the engine room log and its later use in the meteorological report (Brooks 1926). These time lags could be detected in the dataset in regions of strong SST gradients (Saur 1963). On some ships there was more than one engine room temperature log maintained and there was no standard procedure for which temperature was given to the bridge for weather observations. It is expected that with the increased availability of remote readouts of the SST on the bridge that such errors will have reduced over time.

d. Hull sensor measurements of SST

Hull sensors are dedicated SST sensors that measure the SST through the hull or through a hole in the ship's side. The VSOP-NA project showed the hull sensors to be a reliable measurement method and recommended the extension of their use. They are still relatively uncommon because installation costs can be prohibitive unless the sensor and cabling are fitted when the ship is built. Some progress has been made using acoustic modems to relay the SST measurement to a remote readout, which will reduce the cost of installing hull sensor systems in the future (M. J. Yelland and R. W. Pascal 2001, personal communication). Hull sensor SSTs are expected to be of a higher quality than the other meth-

ods of measurement because this is a dedicated sensor. Hull sensor measurements will be taken at greater depths than the bucket measurements, comparable to depths for engine-intake inlets (Kent and Taylor 1991). Figure 2 shows the few reported depths for the ships using hull sensors as squares.

3. Spatial and temporal distribution of measurement methods

a. Data sources

This study uses data from the ICOADS (Woodruff et al. 1998; Diaz et al. 2002). Owing to the availability of metadata, only reports for the period of 1970–97 have been used. The dataset contains reports from VOS, moored and drifting buoys, platforms, and oceanographic measurements, but only VOS SST is considered in this paper. The VOS reports come from a wide range of different ships, ranging from small fishing and research vessels to large container ships and tankers. Several different types of instrumentation are used by these ships and observational practice is variable. ICOADS contains a limited amount of metadata within the ship reports that gives information about the source of the data, recruiting country, platform type (ship, buoy, etc.), the method of SST measurement, and source units and precision.

Since about 1950 the VOS data in ICOADS is predominantly derived from two sources. First, there are data transmitted by radio from the VOS and distributed over the Global Telecommunication System (GTS). Second, there are data extracted from ships' meteorological logbooks and distributed in the electronic International Marine Meteorological Tape (IMMT) format. Where data are available from both sources the IMMT data, which are more complete and better quality assured, take precedence. Until the inception of new meteorological codes in 1982 the GTS data contained no information as to SST measurement method, while the IMMT data contained a flag indicating either "bucket" or "nonbucket" methods. Since 1982 both GTS and IMMT data have a flag indicating the SST measurement method.

Within ICOADS the method of SST measurement, if known, is contained in the SST measurement method indicator flag ("SI"). However, many reports have no information about how the SST was measured. Additional information is available for many of the VOS in WMO's (1997) "List of selected, supplementary and auxiliary ships." The metadata are available annually in digital format for the period from 1973 to the present and on paper for most years from 1955. The metadata prior to 1973 has been imaged by the National Climatic

Data Center Climate Database Modernization Program and will shortly be digitized (S. Woodruff 2005, personal communication). The metadata are indexed by year and ship call sign and can be matched to individual reports within the ICOADS. This study uses only the SST measurement method. Changes in 1995 to the metadata report format allow the reporting of extra information that should prove useful, such as the vessel dimensions and the depth of the SST sensor. However, within the 1970–97 period analyzed here this information was too scarce to be used.

Using the SI indicator and metadata from the WMO, where there is no useful value for SI, a measurement type can be associated with most of the VOS SST reports in the period of 1970–97. In just over 20% of the reports for which a method is available from both sources the metadata disagree. In that case the ICOADS SI flag is given the higher priority.

b. Change in measurement method with time

Figure 3 shows how the composition of the ICOADS ship-measured SST data changes over the period of 1970–97. The number of SST reports per month is from about 120 000 to 140 000 from 1970 until 1979 when it rises to about 180 000 with the start of the First Global Atmospheric Research Program Global Experiment (FGGE) year. This level of reports is maintained until the late 1980s when the number of SST reports starts to decline to about 100 000 month⁻¹ by 1997. In 1970 information is only available for about 30% of the SST reports but by 1997 about 80% of the reports have a known method. Figure 3a shows the information on observation method available using the SI flag only. In 1970 about 20% of the reports are identified as being made using buckets from the SI flag. This drops to less than 5% by 1973, rises to nearly 30% in 1980, and then declines slowly. Engine intakes are only identified in significant numbers by the SI flag from 1982 following a code change to allow the reporting of the method with the GTS report. From 1982 to 1994 about 15% of reports are identified as being engine-intake SSTs, and there is a large peak in 1995–97 with nearly 50% engine-intake data that are a result of SST method information being included in the National Centers for Environmental Prediction (NCEP) data. The number of hull sensor reports identified is a few percent.

Figure 3b shows the improvement in SST method identification gained by the use of the WMO metadata. Because WMO metadata have not yet been digitized prior to 1973 the information on ships instrumentation from 1970 to 1972 uses the metadata for 1973 and will be less reliable than for the later period. From 5% to

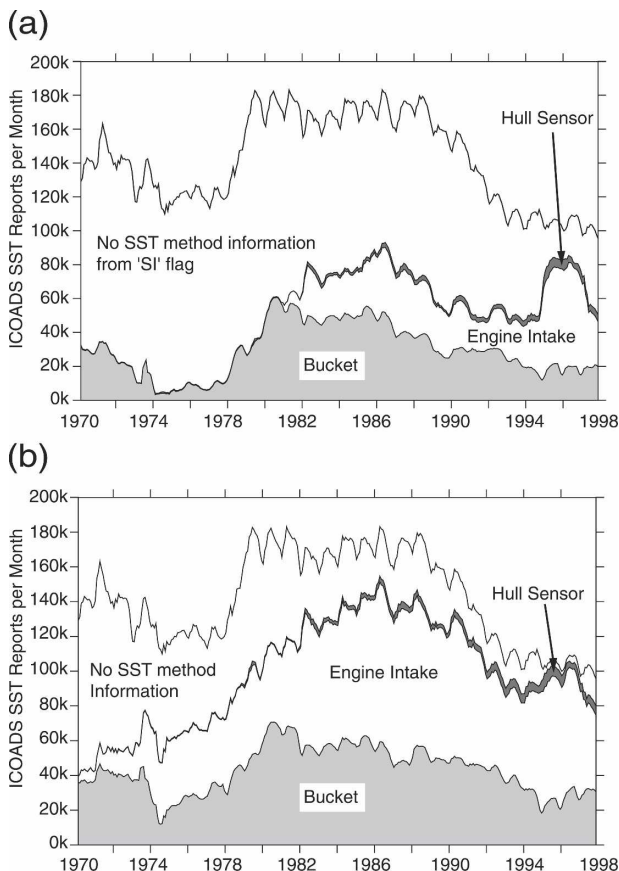


FIG. 3. (a) Breakdown of the average number of ship SST reports by measurement method using information from the ICOADS SI flag for the period of 1970–97. Values have been smoothed with a 3-month running mean filter. (b) As in (a), but using also metadata information from the WMO.

10% more reports are identified as being made using buckets and by 1973 30% of the reports are identified as being made using engine intakes. The 30% figure remains fairly constant until large numbers of engine intakes are identified with the SI flag in 1995. A small number of extra hull sensor reports are identified.

There are a number of reports with SI indicator that is unknown or nonbucket. In the period from 1976 to 1979, 11% of these reports were identified using WMO metadata as being made using a bucket, and 7% as using an engine intake. After 1979 2% are identified as bucket reports and 25% as engine intakes. Prior to 1976 the numbers identified from WMO metadata are negligible. This suggests that studies such as those of Parker (1985) and Folland et al. (1993) who analyzed SST from 1975 to 1981 with metadata containing only bucket and unknown or nonbucket categories may have underestimated the significance of the differences found between the two categories.

TABLE 1. Breakdown of the contribution to VOS SST reports by instrument type and country for each of the six countries contributing the greatest number of reports, and for all data. The monthly mean number of reports is given by decade, along with the percentage contribution from each measurement method [bucket, engine room intake (ERI), hull sensor (hull), other, and unknown]. Also shown is an indication of the "reporting preference": the percentage of reports from each source made in whole degrees. For uniformly distributed data this would be 10%.

	Method	1970s monthly mean	Percent whole degrees	1980s monthly mean	Percent whole degrees	1990s monthly mean	Percent whole degrees
All data	Number	164 493	43%	170 660	48%	128 757	47%
	Bucket	31%	38%	32%	39%	31%	37%
	ERI	18%	50%	39%	56%	44%	57%
	Hull	0.4%	22%	2%	22%	4%	22%
	Other	<0.1%	57%	0.5%	23%	0.7%	15%
	Unknown	51%	46%	28%	48%	21%	45%
United States	Number	35 865	30%	28 816	43%	26 712	42%
	Bucket	10%	55%	4%	33%	5%	39%
	ERI	44%	27%	54%	38%	57%	38%
	Hull	0.2%	34%	0.3%	50%	0.5%	40%
	Other	<0.1%	65%	0.2%	57%	0.1%	36%
	Unknown	46%	28%	42%	51%	38%	50%
Russia	Number	30 367	57%	37 419	58%	25 956	62%
	Bucket	22%	57%	14%	52%	16%	37%
	ERI	8%	59%	58%	62%	73%	69%
	Hull	<0.1%	25%	<0.1%	20%	1%	21%
	Other	0%	...	0%	...	0.1%	27%
	Unknown	70%	54%	27%	54%	10%	60%
United Kingdom	Number	20 648	30%	20 352	30%	16 951	30%
	Bucket	69%	27%	70%	27%	67%	30%
	ERI	5%	51%	9%	62%	7%	61%
	Hull	1%	17%	5%	19%	13%	20%
	Other	0%	...	0%	...	0%	...
	Unknown	25%	34%	17%	29%	13%	21%
Germany	Number	9465	31%	15 456	38%	16 757	35%
	Bucket	69%	29%	84%	36%	83%	35%
	ERI	3%	53%	13%	55%	16%	36%
	Hull	0%	...	0%	...	0%	...
	Other	0%	...	0%	...	0%	...
	Unknown	28%	38%	3%	56%	0.3%	38%
Japan	Number	12 015	68%	15 755	67%	8450	70%
	Bucket	32%	68%	24%	59%	6%	45%
	ERI	32%	71%	68%	69%	80%	71%
	Hull	0%	...	0.5%	31%	0.3%	30%
	Other	0%	...	0.1%	17%	0.2%	30%
	Unknown	36%	63%	7%	61%	13%	68%
France	Number	4617	55%	6271	40%	4601	38%
	Bucket	17%	45%	22%	38%	17%	41%
	ERI	51%	58%	61%	39%	80%	36%
	Hull	0%	...	0%	...	0%	...
	Other	0%	...	0%	...	0%	...
	Unknown	32%	54%	16%	56%	3%	60%

c. The regional distribution of SST measurement type

Because different measurement types are preferred by different recruiting countries there are regional variations in the number of reports for each method. Table 1 shows the number of reports for some of the most frequently reporting countries by decade. Table 1 shows that the ships recruited by the United States and France are most likely to report engine-intake SST. Both Japan and Russia have moved from bucket re-

ports to engine-intake reports over time. The ships from the United Kingdom and Germany predominantly make bucket reports of SST in the period of 1970–97. The difference in preferred national shipping routes of different countries leads to an inhomogeneous global distribution of SST measurement method. Figure 4 shows the annual average number of reports per 5° area for each of the three main SST measurement methods for each decade. The preference of many European countries for bucket SST and that of Japan for

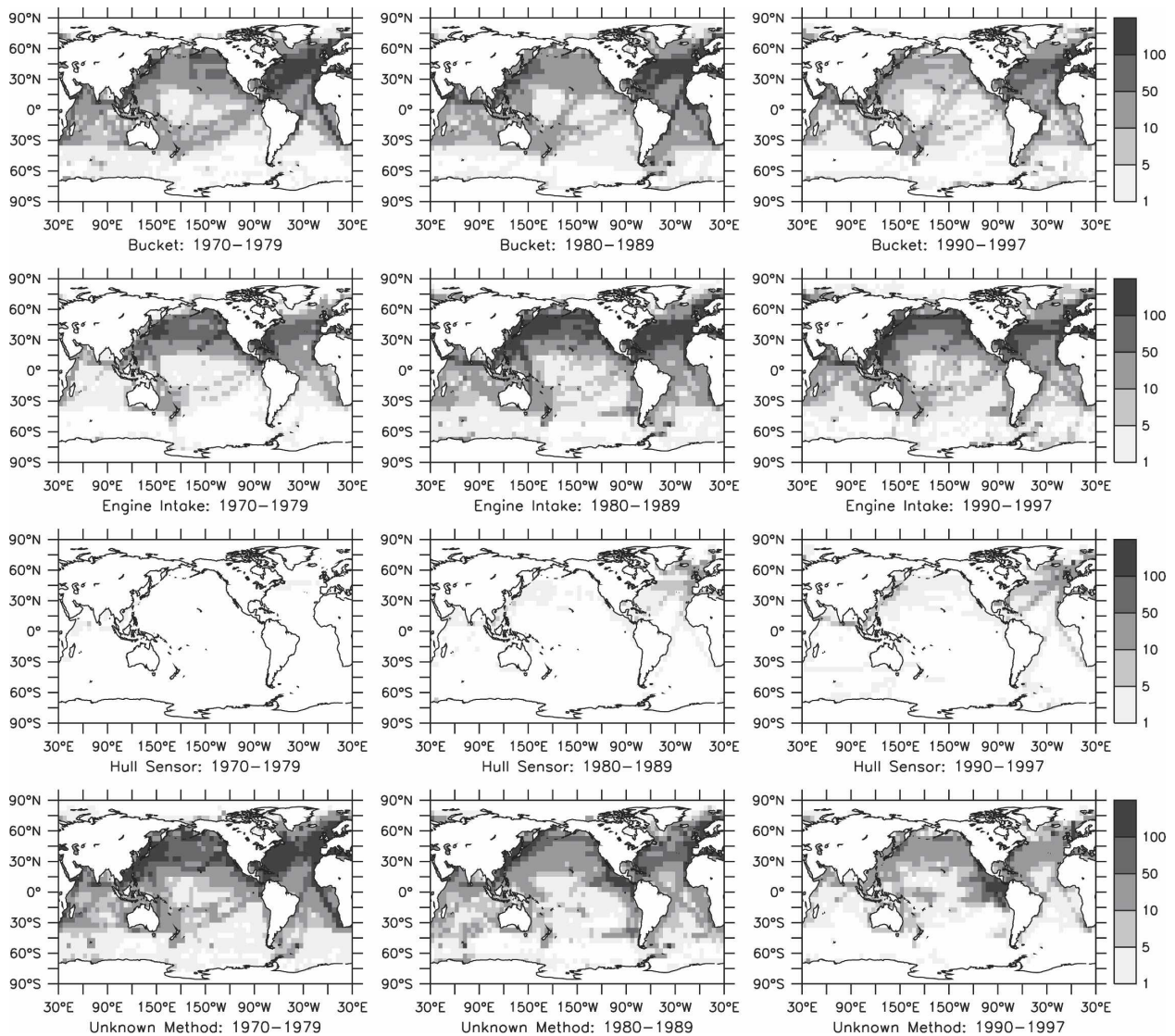


FIG. 4. Regional distribution of SST measurement methods by decade. Plotted are the monthly number of reports per 5° area averaged over the periods of (left) 1970–79, (middle) 1980–89, and (right) 1990–97 for (top) the reports identified as being made by buckets, (second row) engine intakes, (third row) hull sensors, and (bottom row) those for which the measurement method was unknown.

engine-intake SST leads to strong regional variations in the observation type. The proportion of SST reports from engine intakes and hull sensors is increasing with time; the proportion from buckets and of unknown methods is decreasing with time. Bucket and hull sensor SSTs are most common in the Atlantic, whereas the engine-intake reports are more widely distributed.

4. Impact of measurement methods on data quality

a. Reporting preferences

The term “reporting preference” is used to describe the effect of the human observer on the distribution of

measurements. When there is manual intervention in making a report there is a natural tendency to round the report, perhaps to a whole number or to half a degree. The VOS SST reports contained within ICOADS show reporting preferences that vary between countries and instrument types. Excepting half degrees, even decimal places are favored over odd. Over the whole period of 1970–97 45% of the VOS SST reports are in whole degrees and 12% are half degrees. Of the remainder of the reports, even-numbered decimal places are most numerous with temperatures with a last digit of 2 or 8 contributing 7% each and those with 6 or 4 slightly over 5%. SSTs with an odd-numbered last

digit, excepting 5, are most poorly represented, contributing less than 5% each. Although each country shows some preference for whole and half degrees and even last digits there is a strong variation between countries. Table 1 summarizes the prevalence of preferential reporting for the recruiting countries that contribute the most identifiable reports to ICOADS. The table gives the percentage of SST reports from each country that is reported in whole degrees. Typically, bucket reports show less effect of reporting preference than engine-intake reports, particularly in the 1970s. For Japanese engine-intake SST about 70% of reports are made in whole degrees; of reports from the former Union of Soviet Socialist Republics, typically 60%–70% are in whole degrees. For all countries over half the engine-intake reports have zero as their final digit. Bucket reports for all countries typically have less than 40% of the reports made in whole degrees, for hull sensors the figure is 22%. For some countries, such as France and Germany, the amount of reporting preference decreases with time for engine-intake reports. This might be because of the increased use of digital readouts by these countries, but there is no information available on this. Figure 5 illustrates the effect of reporting preference on SST distributions showing a frequency histogram of SST reports near the coast of Japan in July 1995. In this region the peaks at whole numbers of degrees are very obvious.

b. Differences in SST derived from bucket and engine room reports

Having identified the measurement method for most of the SST reports we can generate gridded monthly

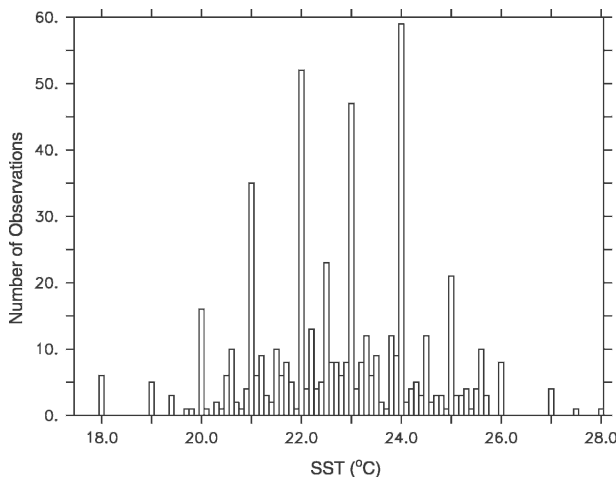


FIG. 5. Illustration of preferential reporting. Histogram of the frequency of occurrence of SST values for July 1995 in the region of 30°–40°N, 140°–150°E. Intervals are in 0.1°C and clearly show that most reports are made in whole degrees.

mean 10° latitude × 10° longitude SST fields for the period of 1970–97 using simple averages of data for each method separately. The difference between the methods shows variability on both interannual and longer time scales. Figure 6a shows the mean difference between bucket and engine-intake SST calculated from these gridded fields (using values only for 10° areas and months where there were at least 10 observations from each method) and averaged over the North Atlantic. The decadal time-scale variability shows the engine intake being relatively warm compared to the bucket SST until the late 1980s when the difference between the two methods decreases. Variability on the shorter multiyear time scale may be because of changes in environ-

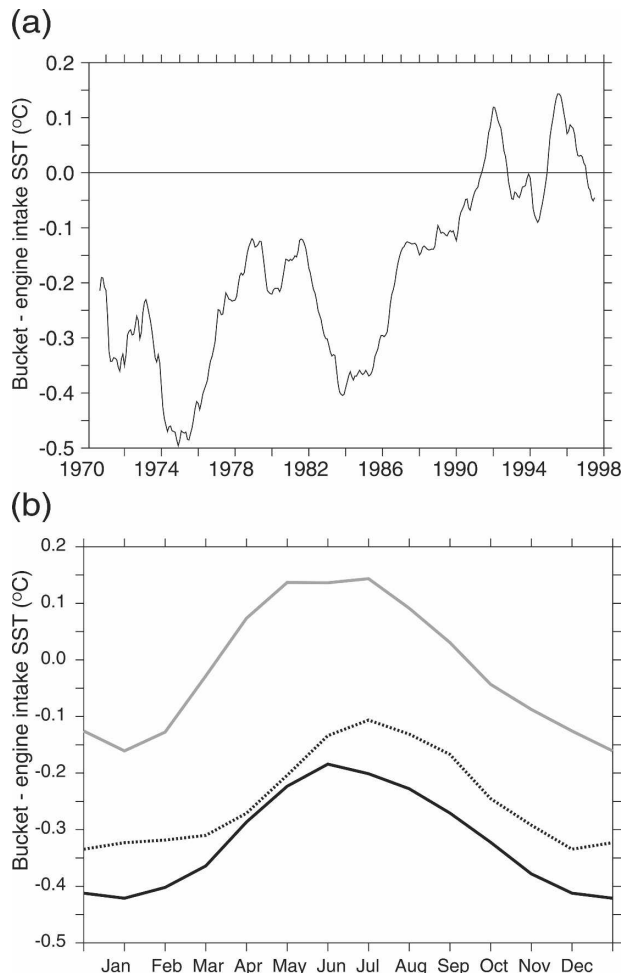


FIG. 6. (a) Mean difference between bucket and engine-intake SST (°C) averaged over the North Atlantic between 20°–50°N and 80°W–0°. Averages have been calculated for all 10° areas and months where there were at least 10 observations made using each method and smoothed with a 12-month running mean filter. (b) Annual cycle of difference between bucket and engine-intake SST (°C) for the same data as in (a): average annual cycle for the 1970s (dark line), 1980s (dotted line), and 1990–97 (light line).

mental conditions (Part III). Figure 6b shows the annual cycle of these gridded differences between bucket and engine-intake SST separately for data from the 1970s, 1980s, and from 1990–97. This interannual variability follows the annual cycle of air–sea temperature difference suggesting that the buckets are cooled by surface turbulent fluxes, particularly in the winter. In the summer we may be seeing the effects of reduced cooling along with stratification of the near-surface waters or heating of the bucket by solar radiation. The size of the annual signal is similar in all three decades, although the mean difference varies from -0.31°C in the 1970s to -0.23°C in the 1980s and is 0 for the 1990s. If the negative bias in the annual difference (Fig. 6a) resulted from the use of uninsulated buckets in the earlier period we would expect the size of the annual cycle to increase, but it remains approximately constant at about 0.3°C over the whole period (Fig. 6b). The likely reason for the negative difference may therefore be a warm bias in the engine-intake SST that is present up to about 1986 and then declines by about 1990. There is evidence from the literature that engine-intake SST may have been of poor quality in the past with warm biases resulting from engine room heating and inadequate thermometers (Saur 1963; Tauber 1969; James and Fox 1972). It is now possible that these problems may have been solved by the early 1990s and that recent engine-intake SST is, on average, unbiased when compared to bucket SST. It should be remembered however that the difference in measurement depth (surface for the bucket compared with an average of 10 m for the engine intake) means that the engine-intake SST should, if anything, be colder than the bucket SST.

5. Discussion and summary

There is a large amount of literature that suggests that there are significant differences between SST reports made by different methods. It is therefore desirable to identify the measurement method for as many SST reports as is possible. We do this for ICOADS using external metadata from the WMO to supplement the SST indicator flag contained within ICOADS. The WMO metadata is particularly important for identification of engine-intake reports in the 1970s. Use of the WMO metadata does however mean that some reports will have an incorrect method ascribed to them. The metadata has space for up to three methods of SST measurement, and we have assumed that the first of these methods is used. There is evidence from WMO metadata that some ships may report methods of measurement interchangeably. Some ships will change method depending on conditions (e.g., normally report-

ing with a bucket SST but using engine-intake SST when the weather is bad or navigational duties are pressing). For those reports for which information was available from both metadata sources, 22% gave differing methods. This suggests that a significant minority of reports have an incorrect method ascribed to them. The problem is likely to lie mainly with the WMO metadata, which is a more indirect method. The WMO files are updated annually (quarterly since 1998), which will lead to some misidentification; ships using multiple methods are a problem and there are known errors with some entries in the metadatabase. Figures 3a and 3b show that WMO metadata are used most heavily in the 1970s, so it is likely that most misidentification of reports occurs in this period. In addition, the use of metadata from 1973, with data between 1970 and 1972, will have increased this problem in the early 1970s. We therefore conclude that any differences found between different methods of measurement are likely to be reduced by misidentification of reports and this underestimate of difference will become more severe going back in time.

Having maximized the number of SST observations with a known measurement method we are able to determine the distribution of the different measurement methods in both space and time. There is a shift over the period of 1970–97 from bucket reports to engine-intake and hull sensor reports. The full picture is however unclear because many reports in the early period do not have a known measurement method. There are differences in the effectiveness of method identification, which may mean that the reports with known methods are not representative of reports as a whole. For example, in the 1970s the ICOADS SI indicator is only useful for identifying bucket reports, almost all of the engine-intake reports in this period are identified from the WMO metadata. The availability of metadata depends both on the country of recruitment of the VOS (for the WMO metadata) and on the data acquisition route [the ICOADS source identifier flag (SID)]. It is therefore not possible to reliably attribute the proportion of unknown SST reports made using each method. However, for the reports of known methods, the bucket SSTs are concentrated in the North Atlantic with engine-intake SST being more widely spread but most common in the North Pacific.

Any bias between the methods will therefore lead to regional- and basin-scale biases in the SST field. These biases will also vary with time as measurement methods change and the methods themselves improve: canvas buckets are phased out and remote-reading precision engine-intake thermometers become more common. There are significant differences between SST derived from the different methods (Fig. 6a). The analysis sup-

ports the results of earlier studies in many respects but also shows that it is not always appropriate to directly relate the results of older work to the constantly changing VOS reports. The strong seasonal cycle in the difference between the bucket and engine-intake SST (Fig. 6b) is probably mainly because of the bucket SST being cooled by the surface fluxes (Roll 1951; Walden 1966; Tauber 1969; James and Fox 1972; Parker 1985; Folland and Parker 1995). There will also be contributions from solar heating of the buckets increasing the daytime SST, and from the bucket sampling any shallow diurnal warm layer. There is no evidence for the amplitude of the seasonal cycle varying significantly, suggesting that the majority of buckets used in the period of 1970–97 were insulated. The longer-term trend (Fig. 6a) suggests that the quality of engine-intake SST may have improved over time. Any improvement in the bucket SST would be indicated by a decrease in the seasonal signal so that the improvement can probably be attributed to the engine-intake SST. Prior to about 1986 the typical annual mean difference between the methods ranges from about -0.5 to about -0.1°C , with -0.3°C being a typical value. This size offset between the methods has been observed before (Saur 1963; Walden 1966; James and Fox 1972; Kent et al. 1993) and attributed to warming of the seawater between the inlet and the thermometer (Brooks 1926; Saur 1963; Tauber 1969) or to the poor quality or siting of the engine-intake thermometer (Saur 1963; James and Fox 1972). Since about 1990 however the annual mean difference between the methods is small with the engine intakes being relatively cool in summer and warm in winter. It is therefore likely that the warm bias in engine-intake SST has reduced with the use of more sophisticated thermometers in better positions. The increased use of remote readouts should also help to reduce any random errors resulting from time lags in reporting the engine-intake SST. It is possible however that the engine intakes are biased relatively cold in summer because they would measure below any shallow thermocline.

These ship reports of SST are the basis for several gridded climate datasets (e.g., Kaplan et al. 1998; Rayner et al. 2003; Smith and Reynolds 2003, 2004). In these datasets there has been no discrimination between observations made using different methods, and no bias adjustments have been made in these datasets for the period after 1970 analyzed in the present study. Any bias in the final field will vary with environmental conditions and with the proportion of reports using each method. The impact of the difference between SSTs measured by different methods on gridded datasets is as yet unclear. At best, if the true SST lies

between the SST estimates from the two single-method fields, the mean would be biased by a smaller amount than either of the single-method fields. However, it is possible that the true SST is not bounded by the single-method estimates.

Preferential reporting is examined by calculating the proportion of reports made in whole degrees. This is most common for engine-intake SST, with over half of the reports being made in whole degrees. This is consistent with the reported difficulty of reading intake thermometers (Saur 1963; Crawford 1969). There is evidence from some countries (e.g., France and Germany) that engine-intake SST-reporting preference may be decreasing over time, as would be expected from the introduction of remote digital readouts. However, for all ships using engine intakes the reporting preference is increasing with time. Bucket reports tend to have less reporting preference, overall less than 40%. This however varies from country to country with less than 30% for the Netherlands, the United Kingdom, and Germany, and over 60% for Japan. Hull sensors show the smallest amount of preferential reporting. Preferential reporting presents challenges for the statistical interpretation of SST data.

We therefore conclude that the use of metadata is important to the understanding of errors in the mean SST in the period of 1970–97. It is also important to take account of the changing characteristics of individual SST measurement methods as well as the changing proportions of each instrument type.

Acknowledgments. This work was supported by funding from the U.K. Government Meteorological Research Programme. The authors thank the reviewers for their helpful comments. We 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. WMO metadata ASCII files were provided by Joe Elms at the National Climatic Data Center (Asheville, North Carolina) and the WMO. Martin Bridger took the photograph of the SST buckets. The Ferret program (information available online at <http://www.ferret.noaa.gov/>), a product of NOAA's Pacific Marine Environmental Laboratory, was used for some of the analysis and graphics in this paper.

REFERENCES

- Brooks, C. F., 1926: Observing water-surface temperatures at sea. *Mon. Wea. Rev.*, **54**, 241–253.
- Crawford, A. B., 1969: Sea-surface temperatures: Some instru-

- ments, methods and comparisons. World Meteorological Organization Tech. Note 103, 151 pp.
- Diaz, H. F., C. K. Folland, T. Manabe, D. E. Parker, R. W. Reynolds, and S. D. Woodruff, 2002: Workshop on advances in the use of historical marine climate data (Boulder, Co., USA, 29th Jan–1st Feb 2002). *WMO Bull.*, **51**, 377–380.
- 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.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaosu, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 892 pp.
- James, R. W., and P. T. Fox, 1972: Comparative sea surface temperature measurements. World Meteorological Organization Reports on Marine Science Affairs, Rep. 5, WMO 336, 27 pp.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner, 2001: Adjusting for sampling density in grid-box land and ocean surface temperature time series. *J. Geophys. Res.*, **106**, 3371–3380.
- Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.*, **103**, 18 567–18 589.
- 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 P. G. Challenor, 2006: Toward estimating climatic trends in SST. Part II: Random errors. *J. Atmos. Oceanic Technol.*, **23**, 476–486.
- , and A. Kaplan, 2006: Toward estimating climatic trends in SST. Part III: Systematic biases. *J. Atmos. Oceanic Technol.*, **23**, 487–500.
- , P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993: The accuracy of voluntary observing ships' meteorological observations—Results of the VSOP-NA. *J. Atmos. Oceanic Technol.*, **10**, 591–608.
- Kirk, T. H., and A. H. Gordon, 1952: Comparison of intake and bucket methods for measuring sea temperature. *Mar. Observer*, **22**, 33–39.
- Meteorological Office, 1956: *Handbook of Meteorological Instruments, Part 1: Instruments for Surface Observation*. HMSO, 458 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.
- , C. K. Folland, and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements. *Climatic Change*, **31**, 559–600.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- 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.
- Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate*, **16**, 1495–1510.
- , and —, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate*, **17**, 2466–2477.
- Stevenson, R. E., 1964: The influence of a ship on the surrounding air and water temperatures. *J. Appl. Meteor.*, **3**, 115–118.
- Tabata, S., 1978: On the accuracy of sea-surface temperatures and salinities observed in the Northeast Pacific Ocean. *Atmos.–Ocean*, **16**, 237–247.
- Tauber, G. M., 1969: The comparative measurements of sea surface temperature in the U.S.S.R. World Meteorological Organization Tech. Note 103, 151 pp.
- 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.