Ocean Science

Open Access

Comparing historical and modern methods of sea surface temperature measurement – Part 1: Review of methods, field comparisons and dataset adjustments

J. B. R. Matthews

School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada

Correspondence to: J. B. R. Matthews (georobin@uvic.ca)

Abstract. Sea surface temperature (SST) has been obtained from a variety of different platforms, instruments and depths over the past 150 yr. Modern-day platforms include ships, moored and drifting buoys and satellites. Shipboard methods include temperature measurement of seawater sampled by bucket and flowing through engine cooling water intakes.

Here I review SST measurement methods, studies analysing shipboard methods by field or lab experiment and adjustments applied to historical SST datasets to account for variable methods. In general, bucket temperatures have been found to average a few tenths of a °C cooler than simultaneous engine intake temperatures. Field and lab experiments demonstrate that cooling of bucket samples prior to measurement provides a plausible explanation for negative average bucket-intake differences. These can also be credibly attributed to systematic errors in intake temperatures, which have been found to average overly-warm by >0.5 °C on some vessels. However, the precise origin of non-zero average bucket-intake differences reported in field studies is often unclear, given that additional temperatures to those from the buckets and intakes have rarely been obtained. Supplementary accurate in situ temperatures are required to reveal individual errors in bucket and intake temperatures, and the role of near-surface temperature gradients. There is a need for further field experiments of the type reported in Part 2 to address this and other limitations of previous studies.

1 Introduction

Sea surface temperature (SST) is a fundamental geophysical parameter. SST observations are used in climate change detection, as a boundary condition for atmosphere-only models and to diagnose the phase of the El Niño–Southern Oscillation (ENSO). The importance of SST to climate science is reflected in its designation as an Essential Climate Variable of the Global Climate Observing System.

Here I review methods of SST measurement, field and lab analyses of shipboard methods and adjustments applied to historical SST datasets to reduce heterogeneity generated by variable methods. Section 2 describes historical and modern methods and changes in their prevalence over time. Section 3 reviews studies evaluating shipboard methods by field experiment or using wind tunnels. Adjustments developed for bucket and engine cooling water intake temperatures are described in Sect. 4. Error in bucket temperatures can strongly depend on the length of time between sampling and temperature measurement (the so-called exposure time). In Sect. 5 an attempt is made to constrain the range of historical variation in this interval using information in the literature. Synthesis and conclusions are presented in Sect. 6.

2   History of SST measurement

SST measurements are obtained by merchant, navy and scientific vessels. Ships that report SST and other meteorological variables to the World Meteorological Organization (WMO) are known as voluntary observing ships (VOS). Modern VOS include container ships, bulk carriers and tankers. International recommendations for SST measurement were first established at the Brussels Maritime Conference of 1853. The conference report proposed that the temperature of surface seawater be measured using wooden buckets (Woodruff et al., 2008). Folland and Parker (1995; referred to as FP95) describe a 19th century ship's wooden bucket of 12 L capacity. It has been suggested that the buckets used transitioned from predominantly wooden to predominantly canvas between the 1850s and 1920s. As discussed by Jones and Bradley (1992), that such a widespread changeover actually occurred is highly uncertain, with canvas buckets known to have been used since at least the 1840s (Parker, 1993). Regardless, I suggest from practical experience that sampling with general-purpose ships' wooden buckets would have been impractical and dangerous on the steamships that gradually replaced slower sailing vessels of lower freeboard in the late 19th century. Such buckets bounce along the sea surface when suspended from ships travelling in excess of ~ 7 kt (~ 3.5 ms^{-1}) and considerable drag is generated once they dip beneath the surface. Canvas buckets do not bounce along the surface and those used aboard steamships appear to have been of fairly small volume (2–4 L, Brooks, 1928; referred to as B28) and often weighted to help them sink (e.g. with a wood base). A photograph featuring such a bucket is presented in Brooks (1932). Canvas buckets are thought to have remained the dominant bucket type used from the 1920s until their gradual replacement by rubber and other modern "insulated" meteorological buckets in the 1950s and 1960s (Kennedy et al., 2011b). Examples of the latter are described in Kent and Taylor (2006). Retrieval of buckets can be challenging, particularly from the bridge of large modern merchant vessels at 30 m up and underway at speeds of 20 kt (~ 10 ms^{-1}) or more. Hénin and Grelet (1996) note that such hauls "can be an arduous and acrobatic process". Indeed, HMSO (1956) recommends that ships travelling faster than 15 kt should obtain intake readings in preference to bucket temperatures for safety reasons.

Due to supposed differences in their propensities for captured seawater to change temperature following collection, rubber buckets are known as "insulated", wooden as "partially-insulated" and canvas as "uninsulated". The walls of canvas buckets can be permeable to sample seepage, with consequent evaporation from the external bucket surface thought to lead to sample cooling. Evaporation of water absorbed into the walls or adsorbed to their outer surface during sampling can also contribute. For buckets without lids, evaporation can also occur about the exposed upper surface of the sample.

A new method of SST measurement evolved with the advent of steamships. To maintain engine temperatures below critical thresholds, large volumes of subsurface seawater were pumped on board for engine cooling. To monitor the efficiency with which the seawater was removing heat from the engine, ships' engineers began observing seawater temperature in engine cooling water intakes. Meteorologists recognised that intake temperature measured prior to the engine might be representative of seawater temperature at intake depth. Such engine intake temperatures (EIT) are known to have been recorded since at least the 1920s (Brooks, 1926; referred to as B26).

The prevalence of EIT readings in the decades prior to World War II (WWII) is poorly known but assumed small. In the primary compilation of historical SST measurements, the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Woodruff et al., 2011), observations from this period largely originate from British, Dutch and German vessels, for which the bucket method was recommended. EIT are thought to dominate SST measurements from 1942–1945 when there was an increase in the proportion of observations coming from US ships, on which the intake method is thought to have prevailed (Thompson et al., 2008). Furthermore, nighttime bucket deployments were likely avoided during WWII since they would have required use of a light on deck (FP95). Kent et al. (2010) present measurement method attribution plots for ICOADS SST data.

While buckets generally sample the upper few tens of centimetres (note that Parker (1993) describes two weighted buckets designed to sample at 1–2 m depth), depths sampled by intakes can be highly variable. Engine intake inlets are usually close to keel depth to ensure submergence under all sea conditions. Actual sampling depth for intakes on container ships and bulk carriers can vary by several metres, depending on shipload (Beggs et al., 2012). Large ships can have dual seawater intakes, one close to keel depth and another a few metres higher (Ecology and Environment, 2007). The deep intake is used at sea and the upper when in shallow coastal waters or canals. Intake depths reported in the early literature are presented in Table 1. Brooks (1926) reports an intake depth of ~ 7 m on a Canadian Pacific steamship in the 1920s. James and Shank (1964) estimate intake depths of ~ 3–10 m for various US merchant, Navy and Coast Guard observing ships reporting in 1962 and 1963. They defined relations between intake depth and full-load draft for different hull types and categorised observing ships by hull type to estimate their intake depth. More contemporary intake depths averaged by type of VOS ship reporting this between 1995 and 2004 are presented in Table 5 of Kent et al. (2007). Container ships and tankers were found to have intakes at ~ 7–9 m depth while intakes on bulk and livestock carriers were found to often exceed 10 m. Kent and Taylor (2006) report that the average intake depth for VOS reporting this in 1997 was 8.4 ± 4.1 m, with the deepest inlet being at 26 m. Oceanographic research vessels often have dedicated

Table 1. Intake depths reported from observing ships of various type in pre-1980 literature. All are from ships contemporary to the publication year except Collins et al. (1975), which are from vessels operating during 1927 to 1933. CSIRO is the Commonwealth Scientific and Industrial Research Organisation of Australia.

Reference	Intake depth(s)	Ship type(s)
Brooks (1926)	22–24 ft ($\sim$ 7 m)	Canadian Pacific steamship (RMS *Empress of Britain*)
	15 ft ($\sim$ 4.5 m)	2 US Coast Guard Ice Patrol vessels (*Tampa* and *Modoc*)
Lumby (1927)	30 ft ($\sim$ 9 m)	Ocean liner (RMS *Majestic*)
Roll (1951a)	4.5 m	German Fishery Patrol Vessel (*Meerkatze*)
Kirk and Gordon (1952)	9 ft ($\sim$ 3 m)	British ocean weather ships
Saur (1963)	10–22 ft ($\sim$ 3–7 m)	3 US Military Sea Transport Service ships, 9 US Navy Radar Picket ships
James and Shank (1964)	10–32 ft ($\sim$ 3–10 m) (average of 21ft)	US merchant, Navy and Coast Guard vessels
James and Fox (1972)	0–9 m	WMO voluntary observing ships
Piip (1974)	2–6 m	Merchant vessels traversing 0–50° S, 135–180° E and reporting to CSIRO
Collins et al. (1975)	5 m	Canadian Pacific steamships
Tabata (1978a)	3 m	Canadian ocean weather ships
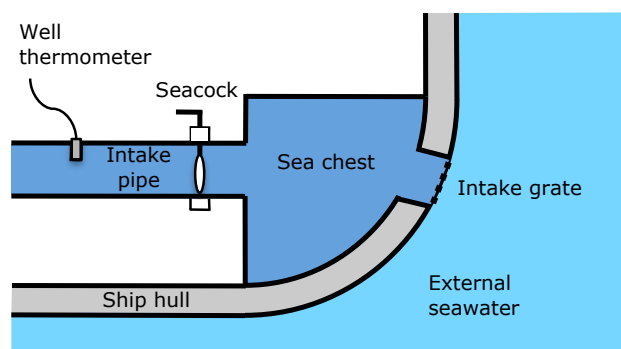
Fig. 1. Schematic of a typical engine cooling water intake system on a modern merchant vessel.

seawater intakes for underwater scientific measurements, typically sampling at $\sim$ 2–4 m depth. These scientific intakes are distinct from engine intakes in that the pipes tend to be of much smaller diameter, a few centimetres (e.g. Tabata, 1978a), as opposed to tens of centimetres (e.g. Kirk and Gordon, 1952).

With EIT readings traditionally being obtained by ships' engineers for engine monitoring purposes, procedures and instruments have varied from ship to ship and remain unstandardised and poorly documented today. As reported in Part 2, intake thermometers are generally mounted within 15 m inboard of the inlet and beyond a seacock (Fig. 1). On modern vessels seawater is often piped aboard through a sea chest, a sealed metal box built into the hull with an external grate. Intake thermometers are sometimes mounted into the sea

chest itself (Tabata, 1978b), reportedly a favoured position for distant-reading thermometers (WMO, 2008). In addition to one or more main engine intake lines (e.g. with multiple engines), ships can have multiple ancillary lines, and temperature be measured on several (B28; Saur, 1963).

Two main types of EIT method can be distinguished: well and faucet. In the well method, a temperature probe or thermometer bulb is mounted inside a well sunk into the intake pipe to around a third its inside diameter (e.g. Kirk and Gordon, 1952; Piip, 1974). Wells may be oil-filled and are sometimes referred to as thermometer pockets or thermowells. Rapid conduction across the well casing allows intake temperature to be measured while at the same time enabling the probe or bulb to be readily removed for maintenance. The sensing element can also be directly inserted into the pipe (e.g. Stupart et al., 1929). Both well and direct insertion temperatures are sometimes referred to as injection temperatures. In the faucet method, seawater is sampled from the intake through an attached pipe fitted with a tap and its temperature measured externally (e.g. B26; HMSO, 1956; Piip, 1974).

Manually-read mercury-in-glass thermometers have been used for both bucket (B28; Tauber, 1969; Collins et al., 1975; Tabata, 1978b) and intake (B28; Saur, 1963; Tauber, 1969) measurements. Intake temperatures have also been taken using mercury-in-steel (HMSO, 1956; Piip, 1974; Collins et al., 1975) and electrical resistance thermometers (B26; Tabata, 1978a), and thermistors (Tabata, 1978c). Hagart-Alexander (2010) provide a review of thermometer types and a description of thermowells. EIT has sometimes been continuously

recorded using a thermograph (e.g. B26; Collins et al., 1975; Piip, 1974). Brooks (1932) presents a photograph of an early version of such a setup. On modern vessels, a distant-reading display may be available in the engine room or on the bridge (WMO, 2008).

In recent decades the number of bucket and engine intake observations has declined, in part due to reduction in the WMO VOS fleet from a peak of over 7500 ships around 1985 to under 4000 today (Kennedy et al., 2011b). Shipboard hull contact sensors, that is temperature sensors mounted to the outside or inside of the hull (e.g. Beggs et al., 2012), have increased in prevalence over this period, providing more SST observations than buckets by the late 1990s (Kent et al., 2007). They presently contribute around a quarter of all VOS SST measurements (Kent et al., 2010). Other dedicated shipboard methods include radiation thermometers, expendable bathythermographs and trailing thermistors.

Since the early 1970s VOS SST measurements have been augmented by temperatures from ocean data acquisition systems (ODAS), primarily moored and drifting buoys. Around 70 % of in situ observations were obtained by buoys in 2006 (Kennedy et al., 2011b). ICOADS contains drifting buoy measurements from 1978 onwards and moored buoy observations from 1971 (Woodruff et al., 2011). Earlier measurements from these platforms may exist but are not included in ICOADS. While drifting buoys are purported to measure sea temperature at a nominal depth of ~ 25 cm (Kennedy et al., 2007), they oscillate within the surface wave field such that actual measurement depth can be anywhere within the upper 2 m (Emery et al., 2001).

Some SST datasets incorporate both satellite and in situ observations. While satellite retrievals of SST have been obtained since the 1960s (Rao et al., 1972), only observations obtained following the advent of the Advanced Very High Resolution Radiometer (which measures in the infrared) are generally utilised today, often from 1981 onwards (e.g. Reynolds et al., 2002). Since 1991, SST retrievals have also been obtained using along-track scanning radiometers, which measure over three channels in the thermal infrared (Merchant et al., 2008). Unlike earlier instruments, these are self-calibrating, providing fairly accurate retrievals without the need for calibration using in situ measurements. SST has also been measured by satellite-borne passive microwave radiometers since 1997 (Wentz et al., 2000). These have an advantage over infrared sensors in that microwaves can penetrate clouds with little attenuation. Satellite instruments observe temperature within the sea surface skin (upper ~ 1 mm) whereas in situ methods measure the so-called bulk temperature beneath. Satellite observations have greatly improved spatial coverage, particularly in the Southern Ocean where in situ sampling remains sparse.

3 Field and lab evaluations of shipboard methods

3.1 Bucket-intake temperature comparisons

Field evaluations of SST measurement methods have largely focused on average differences between bucket and engine intake temperatures. Brooks (1926) compared tin bucket and engine intake temperatures collected aboard the Canadian Pacific steamship RMS *Empress of Britain* on a cruise between New York and the West Indies in February and March 1924. Faucet and injection temperatures were found to respectively average 0.1° F (~ 0.06 °C) and 0.5° F (~ 0.3 °C) warmer than near-simultaneous tin bucket temperatures. The injection temperatures were obtained from thermometers mounted on the condenser intake pumps, noted as difficult to read to better than 1° F (~ 0.6 °C). Brooks suggests the injection temperatures in error due to parallax in reading and warming of intake seawater about the pumps. A fast-response cylindrical bulb thermometer was used to obtain both the tin bucket and faucet temperatures and appears to have been readable to 0.1° F. This was not the thermometer in standard use for bucket measurements aboard the *Empress of Britain*; rather, a longer-response spherical bulb thermometer read to 0.5 or 1° F was used. Brooks suggests the tin bucket samples cooled slightly pre-measurement, at most by 0.2° F (~ 0.1 °C). Finding the maximum difference between the faucet and tin bucket temperatures to be only 0.25° F (~ 0.15 °C), he concluded that the upper ocean had been well-mixed to at least the intake depth (~ 7 m). He does note, however, that sizeable positive average bucket-intake differences had previously been found in spring and summer in the Grand Banks aboard the *Tampa*. Reported average differences across the upper 5 m were 0.6° F (~ 0.3 °C) in daytime and 0.3° F (~ 0.2 °C) at night for April to July 1925. Note that 0.6° F was added to the intake readings for supposed parallax error so the unadjusted differences were in fact larger. Similar gradients were found for the western North Atlantic in summertime by James and Shank (1964) using bathythermographs. They found the temperature contrast between 10 and 30 ft (~ 3–9 m) exceeded 0.6° F (~ 0.3 °C, ~ 0.05 °C m^{-1}) over 15 % of the time in June, July and August but was ≤ 0.2° F (~ 0.1 °C, ~ 0.02 °C m^{-1}) over 85 % of the time from September to March. Isothermal conditions were observed at least 55 % of the time during the latter period.

Brooks conducted an additional shipboard comparison aboard the ocean liner SS *Finland* on a cruise between San Francisco and New York in May 1928 (B28). Temperatures from the main engine intake were found to average 0.8 °C warmer than those obtained by fast measurement with a rubber-covered tin bucket of small volume (1.7 L). Those from the refrigerator intake in the refrigerator room averaged 0.2 °C warmer. Respectively, the engine intake and refrigerator intake readings were found to average 0.7 and 0.3 °C warmer than those from a specially-fitted intake

thermograph. While details of the engine intake thermometer were not reported, the refrigeration intake thermometer was graduated in intervals of 2° F (~ 1.1 °C). Temperature change of the tin bucket sample pre-measurement was assumed small, although cooling of 0.1 °C was noted in one minute following collection under a wind speed of 9 m s^{-1} and SST-wet bulb temperature contrast of 6 °C.

Roll (1951a) compared bucket and intake temperatures obtained in the North Sea and Norwegian Sea from June to October 1950 by the German Fishery Patrol Vessel *Meerkatze*. A pipe was specially fitted to the engine intake to divert seawater, in a system designed to obtain accurate EIT readings. Bucket temperatures were obtained using a rubber-insulated water scoop of very small volume (0.6 L). An average bucket-intake difference of -0.07 °C was found from 410 comparisons. Small positive average differences of 0.1 °C and below were generally found at low wind speeds (up to Beaufort force 4) and attributed to near-surface temperature gradients. Increasingly negative average differences were found at higher wind speeds and attributed to enhanced cooling of bucket samples, changing from -0.1 °C at Beaufort force 5 to nearly -0.25 °C at force 6.

Kirk and Gordon (1952) compared bucket and intake temperatures obtained aboard Dutch merchant vessels in the eastern North Atlantic south of the British Isles. Intake temperatures tended to be ~ 1° F (~ 0.6 °C) warmer than bucket readings. Considerable scatter was found in the individual bucket-intake differences, with standard deviations ranging from around 0.7 to 0.9° F (~ 0.4–0.5 °C) across the Marsden squares analysed, increasing towards higher latitudes. They also compared bucket (UK Met Office Mk III) and intake thermograph measurements obtained by British ocean weather ships in the eastern subpolar North Atlantic between March and November 1949. The Mk III is a canvas bucket with an internal double-walled copper vessel and spring lid. The average across various cruise-mean intake-bucket differences for three weather ships was 0.4° F (~ 0.2 °C) on-station and 0.2° F (~ 0.1 °C) underway. The larger difference on-station was suggested due to enhanced engine room warming of intake seawater from a reduced volume flow through the intake. The cruise averages varied between -0.6 and $+0.3$ °C for both on-station and underway measurements.

Tauber (1969) evaluated EITs collected by three Soviet research vessels in the Pacific and Indian Oceans and by trawlers in the Black Sea and Sea of Asov between April 1967 and February 1968. Virtually all EITs (98 %) obtained by one research vessel were found to be overly-warm by > 0.5 °C (compared against accurate measurements) while on the other vessels they were 1.2–2.3 °C too warm in 83 % of cases. Tauber thus concluded EIT measurements were unreliable.

Saur (1963) analysed 6826 pairs of bucket and engine intake temperatures obtained aboard 12 US military vessels between May 1959 and January 1962. Three of the vessels were traversing the North Pacific while the remainder

were usually stationed ~ 300 miles off the US west coast. The fleet average intake-bucket difference derived from average differences for the individual vessels was 1.2 ± 0.6° F ($\sim 0.7 \pm 0.3$ °C). There was significant variation in the latter differences, ranging between -0.5 and $+3$° F (around -0.3 and $+1.7$ °C), and between cruise averages for individual vessels, in one case varying between 0.3 and 1.8° F (around 0.2 and 1 °C). Specially-designed buckets and thermometers accurate to at least 0.15° F (~ 0.1 °C) were used for the bucket measurements. Thus the non-zero average differences likely primarily reflect errors in the intake temperatures, although with near-surface temperature gradients playing some role. Intake temperatures were only reported in whole ° F, being read from thermometers graduated in intervals of 2 or sometimes 5° F (around 1.1 and 2.8 °C). Saur notes that a comparison between intake thermometers used aboard five US Coast Guard weather ships and an accurate thermometer had found systematic errors between -2 and $+3.9$° F (around -1.1 and $+2.2$ °C).

One of the most observation-rich bucket-intake comparisons ever conducted was that of James and Fox (1972). They analysed 13 876 pairs of near-simultaneous bucket and intake temperatures obtained aboard VOS ships between 1968 and 1970. Although of global distribution, reports were mainly from the North Atlantic and North Pacific shipping lanes. From a compilation of all observations, intake temperatures averaged 0.3 °C warmer than bucket readings. Considerable spread was found in the individual differences with 68 % falling within ± 0.9 °C and the largest differences exceeding ± 2.5 °C. This noise is not surprising given the temporal and spatial coverage of the collated observations and the heterogeneity of the bucket and intake methods (e.g. variable thermometer quality and observer care). They found that intake temperatures from mercury thermometers yielded a larger average intake-bucket difference (0.3 °C) than those from precision thermometers or thermistors (both 0.09 °C).

On the whole, these studies suggest a tendency for intakes to read warmer than buckets, in opposition to what we would expect from typical near-surface temperature gradients (cooler with depth). The precise cause of reported average bucket-intake differences is not always clear, potentially being due to both bucket and intake errors where neither has been shown to be accurate. Confusing matters, buckets and intakes cannot be assumed to sample seawater of the same temperature in the presence of near-surface temperature gradients. This leads us to a discussion of terminology. The term "bias" is sometimes applied to average bucket-intake differences (e.g. Kennedy et al., 2011b) yet seems inappropriate given that both bucket and intake temperatures may show average deviations from the actual SST. By the latter I mean the actual temperature in the upper few centimetres. Similarly, use of the term "correction" to describe adjustment of bucket temperatures to be more consistent with EIT and vice versa is also unsuitable.

Identification of individual errors in bucket and intake temperatures in field comparisons requires supplementary accurate in situ temperature measurements. Studies by Susumu Tabata published in the late 1970s are amongst the most comprehensive in this regard. Tabata (1978a) analyses upper ocean temperatures collected over 1956–1976 by Canadian weather ships at Station P and traversing Line P in the northeast Pacific, a ~ 1425 km-long transect extending from the coastal waters of southwestern Vancouver Island, British Columbia, to Station P in the mid-Gulf of Alaska (Crawford et al., 2007). The mean difference between temperatures from a specially-designed meteorological bucket and an accurate reversing thermometer in the upper 1 m was 0.04 ± 0.13 °C over 1969–1976, with bucket temperatures thus concluded accurate to ± 0.1 °C. Like Saur (1963), average bucket-intake differences were found to vary widely between ships and between cruises on the same ship, although the individual cruise standard deviations were generally smaller and more consistent at around 0.05–0.25 °C (compared to around 0.3-0.8 °C for Saur). The latter likely reflects reduced noise in the intake temperatures from the weather ships due to better observing practices (they were collected by meteorological observers) and use of higher precision instruments (precision of ± 0.2 °C). Mean cruise intake-bucket differences were -0.02 and $+0.18$ on two weather ships over 1962–1967 (*St. Catherines* and *Stonetown*), and -0.05 and -0.02 °C for two other weather ships over 1967–1976 (*Vancouver* and *Quadra*). Except for the *St. Catherines* there was considerable variation in the average differences for individual cruises on these ships, for example, mostly varying within ± 0.3 °C for the *Quadra*.

Tabata (1978b, d) conducted a similar analysis using measurements collected by a Canadian oceanographic research vessel in the northeast Pacific in August and September 1975. Only observations coincident with wind speeds exceeding ~ 6 m s^{-1} were analysed, conditions under which the upper 10 m was considered isothermal. EIT (inlet at 4 m) averaged 0.3 ± 1.2 °C warmer than accurate temperatures from a salinity-temperature-depth (STD) meter. Tabata attributed the large standard deviation to reading error of the intake thermometer by the engine room crew, with the largest differences exceeding ± 2 °C.

More recently, Hénin and Grelet (1996) compared meteorological bucket temperatures to conductivity-temperature-depth (CTD) temperatures at 1–2 m depth obtained by research vessels in the western equatorial Pacific. Bucket temperatures were found to average 0.13 ± 0.34 °C and 0.16 ± 0.22 °C warmer than CTD temperatures on two cruises and 0.60 ± 0.48 °C cooler on another cruise. The warm average differences may have been attributable to temperature gradients over the upper few metres. The cause of the cool average difference is unclear but apparently due to the bucket measurements since the corresponding average CTD-thermosalinograph difference was similar to those for the other cruises.

3.2 Canvas bucket experiments by the Sea Education Association

The accuracy of canvas bucket temperatures was tested by field experiments in the early 1990s aboard the Sea Education Association (SEA) sailing vessel *Corwith Cramer*. The *Cramer* is the Atlantic sister ship of the *Robert C. Seamans* used for Part 2 of this study, the *Seamans* operating in the Pacific. The experiments, undertaken for the late Reginald Newell of the Massachusetts Institute of Technology, were conducted over several cruises across the western North Atlantic and Caribbean. They are described in a series of student project reports in the SEA archives in Falmouth, MA, USA. FP95 compared observations from one cruise to results from their canvas bucket model (described in Sect. 4). Underway at around 15 to 25 locations on each cruise, a replica Mk II canvas bucket was filled with surface seawater and hung on deck for 10 min in a wind-exposed, sun-shaded location. During this 10 min period, the sample temperature was measured each minute and the bucket agitated every half minute to mix the sample. The Mk II was in use aboard British ships (likely mostly motor vessels) from the 1930s until at least the 1950s (FP95; HMSO, 1956). Cooling over 5 or 10 min equating to average rates of ~ 0.05–0.10 °C min^{-1} was generally reported. Cooling rates in the first minute (mostly unreported) were likely faster due to non-linearity, with cooling of 0.2–0.3 °C or more found in one minute in some cases.

One peculiarity in the experimental method is that the replica canvas bucket itself appears often not to have been used for seawater collection, apparently due to concerns this valuable bucket would be damaged. Instead, a plastic bucket was used for sampling and the Mk II then filled with seawater from this. In one report it is noted that the the canvas bucket was dipped into the plastic bucket for filling so that its walls were made wet, although the extent to which this was the case for other cruises is unclear. Results for both wet and dry walls are reported for some cruises. Regardless, the experiments suggest that samples in small-volume canvas buckets can cool rapidly.

3.3 Field comparisons of different bucket types

Few shipboard comparisons between different bucket types have ever been conducted. James and Fox (1972) report average bucket-intake differences for various bucket types but no direct differences between bucket types.

B26 compared temperatures from canvas and tin buckets (4 L and, 2 or 4 L, respectively) obtained aboard the *Empress of Britain*. When dropped from the bridge, the canvas bucket measured an average of 0.5° F (~ 0.3 °C) cooler than a tin bucket launched from a lower deck in 10 comparisons, increasing to 1° F (~ 0.6 °C) when the quartermasters took the canvas bucket measurements rather than Brooks himself ($n = 79$). The bulk of the latter comparisons ($n = 65$) were conducted south of 35° N (and above 9° N), for which the

average difference was smaller at 0.3 °C. The extent to which the larger difference found for the quartermasters' measurements is due to additional sample cooling rather than thermometric error is unclear. The quartermasters were using the ship's slow-response (and perhaps poorly-calibrated) thermometer while Brooks was using a calibrated fast-response thermometer. Recall also that the quartermasters were only measuring to a half or whole ° F. Brooks attributes the difference to several sources, including cooling by or of the thermometer.

Brooks conducted a similar comparison aboard the *Finland* in May 1928 (B28). Canvas bucket temperatures (4 L bucket, double-walled) obtained by the crew averaged 0.4 °C lower than tin bucket temperatures obtained by Brooks with a calibrated thermometer, both buckets being deployed from a similar low deck level. Although attributed to cooling of the canvas bucket samples, the main thermometer used by the crew (a galley thermometer) exhibited variable error between −0.5 and +0.75° F (−0.3 and +0.4 °C) dependent upon temperature. However, the cool error of the canvas bucket temperatures was found to increase with larger depressions of the wet bulb temperature below the SST, as would be expected for sample cooling. It was also found to be substantially larger at nighttime than daytime, averaging 1.1 and 0.4 °C, respectively. The larger nighttime error was attributed to the observers removing the reservoir thermometer from the bucket to hold under a light for reading.

3.4 Wind tunnel experiments

Ashford (1949) measured the temperature change of water samples in several types of bucket when suspended in a wind tunnel. The buckets were first dipped in a water bath, the temperature of which was varied to yield a range of air–water temperature contrasts. Wind speed was held fixed at 20 mph ($\sim 9\,\mathrm{m\,s^{-1}}$) while air temperature and relative humidity varied from 15.6–18.3 °C and 50–60 %, respectively. Note that the latter is fairly low compared to typical open ocean values. The rate of sample temperature change was found to increase with greater contrast between the wet bulb and water temperature, with warming observed for positive differences and cooling for negative. Measured cooling rates with an Mk II bucket were intermediate between those for a rubber-walled German scoop thermometer and a German rubber pail. With a 3 °C water-wet bulb contrast, the scoop thermometer cooled at $\sim 0.2\,°\mathrm{C\,min^{-1}}$ and the Mk II by $\sim 0.1\,°\mathrm{C\,min^{-1}}$, while the sample in the rubber pail did not change temperature perceptively. These contrasting cooling rates may partly reflect the different volumetric capacities of the buckets (0.6 L for the scoop thermometer, 4 L for the Mk II and unknown volume for the rubber pail). Cooling rates were found to be independent of whether the external surface of the bucket was wet or dry.

Roll (1951b) conducted wind tunnel experiments with the same model of scoop thermometer as used by Ashford (1949). This was again immersed in a tank of water at a desired temperature and then suspended. Wind speed was varied between 2 and $19\,\mathrm{m\,s^{-1}}$ and air–water temperature contrast was varied between +5 and −10 °C. With a −2.5 °C air–water temperature contrast, the sample cooled, respectively, by 0.1 and 0.25 °C in the first minute at wind speeds of 8 and $10\,\mathrm{m\,s^{-1}}$, with cooling not detected at lower wind speeds. No cooling was detected in the first minute for wind speeds of $6\,\mathrm{m\,s^{-1}}$ and below. The rate of temperature change was found to decline over the 10 min measurement period as the temperature contrast was diminished by heat exchange. Roll (1951a) stresses the difficulty of using results from wind tunnels to correct bucket temperatures given that the wind conditions experienced by buckets during the exposure period aboard ships cannot be reliably estimated.

4 Bucket and engine intake temperature adjustments

FP95 developed physical models for temperature change of seawater samples in wood and canvas buckets. Modelled temperature change is dependent on air–sea temperature difference, relative humidity and apparent wind speed. Different versions of the models were developed by altering parameters such as ship speed and bucket exposure to solar radiation. Two canvas buckets of different dimension were modelled, one the size of the Mk II and the other half its diameter at 8 cm. Adjustments were derived for both "fast" and "slow" ships to represent motor and sailing vessels, with ship speed set to 7 and $4\,\mathrm{m\,s^{-1}}$, respectively.

The FP95 bucket models are particularly sensitive to the choice of exposure time. For canvas bucket adjustments in non-equatorial regions with appreciable seasonal SST cycles and sufficient data, exposure time was determined using the finding that seasonal cycle amplitudes were generally larger in pre-1942 years (Folland, 2005). FP95 assumed the larger amplitudes were due to environmental cooling of wood and canvas buckets, the strength of which varies seasonally in their adjustments. Exposure time was altered in 10° latitude bands to find adjustments that would minimise the variance of three pre-1942 30-year average seasonal cycles relative to the total variance of their complete record. The longest exposure times so derived exceeded 5 min and the shortest were under 2 min. An "optimum integration time" (not reported) was calculated for each model version by averaging over derived times for all 30-year averages across all latitude bands. The exposure time for wooden bucket adjustments was set to 4 min everywhere, partitioned into a 1 min hauling period and a 3 min on-deck phase.

To generate final pre-1942 "corrections", the adjustments from different model versions were combined to fit a time-variant ratio of the number of wood to canvas bucket observations and an assumed linear increase in ship speed from 4 to $7\,\mathrm{m\,s^{-1}}$ between 1870 and 1940. The former was set so that the resulting adjustments would minimise the difference

between night marine air temperature (NMAT) and SST anomalies in the tropical Pacific and southern tropical Indian Ocean between 1856 and 1920. FP95 found pre-1942 annual-mean northern- and southern-hemispheric NMAT anomalies were up to 0.5 °C larger than the corresponding SST anomalies and attributed this to bucket cooling. It is commonly assumed that NMAT and SST anomalies should be similar on seasonal and longer timescales.

The FP95 adjustments have been applied with some modifications to pre-1942 bucket temperatures in the UK Met Office Hadley Centre Sea Ice and SST dataset, HadISST (Rayner et al., 2003), and the second and third versions of the Hadley Centre SST dataset, HadSST2 (Rayner et al., 2006) and HadSST3 (Kennedy et al., 2011a, b). Independent bucket adjustments have been applied to the US National Oceanic and Atmospheric Administration's Extended Reconstruction SST version 3, ERSSTv3 (Smith et al., 2008), derived by Smith and Reynolds (2002) using the assumption of similarity between NMAT and SST anomalies. Kent et al. (2010) compare bucket adjustments applied to HadSST2 and ERSSTv3. Both generally increase on a global annual-average from the mid-19th century to around 1920 and then plateau to the late 1930s. In HadSST2, the global-mean of the adjustments increases from ∼ 0.2 °C in 1880 to ∼ 0.4 °C in 1920. This is due to the specification of increases in the proportion of canvas to wooden bucket measurements and "fast" to "slow" ships over this period.

As of 2008, in situ observations in historical SST datasets had not been adjusted post-1941. Thompson et al. (2008) suggested a need to apply adjustments to more recent observations, arguing an abrupt 1945 drop of ∼ 0.3 °C in global-mean SST from HadSST2 was the result of uncorrected method changeover. In HadSST3, adjustments have been applied to measurements from buckets, buoys and engine intakes over the duration of the record (1850–2006). The FP95 "fast ship" adjustments are used post-1941, with their wooden bucket adjustments applied to temperatures from modern "insulated" buckets. A linear switchover from canvas buckets to the latter is specified over the 1950s and 1960s. As for HadSST2, different realisations of the FP95 adjustments were derived by varying bucket model parameters within their supposed uncertainty ranges.

Multiple realisations of EIT adjustments were also developed for HadSST3. For measurements obtained in the North Atlantic between 1970 and 1994, adjustments were generated from the EIT errors of Kent and Kaplan (2006). Adjustments for other regions and years were derived by taking the best estimate for the average EIT error from the literature to be 0.2 °C too warm. Note that "strictly speaking" adjustments are intended to be relative to the mix of observations in the respective dataset reference period (in this case 1961–1990) rather than corrections back towards "true" values.

HadSST3 has been combined with the fourth version of the Climatic Research Unit (CRU) near-surface land air temperature dataset, CRUTEM4 (Jones et al., 2012), to produce a new global instrumental surface temperature record, HadCRUT4 (Morice et al., 2012).

5　Exposure time

The magnitude of the temperature change simulated by the FP95 bucket models critically depends on the specified exposure time. This can be partitioned into a hauling phase and an on-deck period. Here an attempt is made to constrain the historical durations of these periods using information in the literature.

The length of the hauling period depends on the height of the observer above the waterline and the quickness of the haul. Lumby (1928) notes that buckets could be drawn upward a distance of 30–60 ft (∼ 9–18 m) or more. Brooks (1926) reported that quick hauls with a 4 L tin bucket from 10 to 20 ft (∼ 3–6 m) up on the leeward stern of the *Empress of Britain* took him 20 to 30 s, equating to hauling speeds of ∼ 0.2 m s^{-1}. The hauling period would undoubtedly have been longer for the canvas bucket measurements conducted by the crew from the bridge, but no estimates are given. Lumby (1928) estimates the exposure time for these measurements to have been no longer than 2 min based on his own experimental sample cooling rates (0.11–0.12 °C min^{-1}) and the portion of the canvas bucket error attributed to sample cooling by Brooks (∼ 0.25 °C), both of which are rather uncertain. On the *Finland*, the typical hauling period for the mariners' canvas bucket deployments from a low deck (likely ∼ 9 m up) was apparently 2 min, equating to a very slow hauling speed of 0.08 m s^{-1}. Brooks suggested this could have been reduced to 1 min by faster handwork. For comparison, the hauling period for Brooks' tin bucket measurements from a similar deck level (9 m up) appears to have been ∼ 20–30 s (his exposure time was ∼ 30–40 s and his fast-response thermometer stabilised in ∼ 10 s), equating to hauling speeds ∼ 0.3–0.45 m s^{-1}. Note that all these values are for bucket deployments conducted on large ocean liners. It is unclear whether deployments were generally conducted from the bridge or from a lower deck on such vessels. More generally, the extent to which deployments are and were conducted from heights exceeding 10 m is unknown, with deployments becoming increasingly difficult at greater heights and vessel speeds.

With regards to the on-deck period, this can generally be assumed to be largely comprised of the waiting period for thermometer stabilisation following insertion, at least for those buckets without built-in thermometers. Bucket temperature readings conducted aboard the *Finland* by the ships' crew took ∼ 45–60 s, suggesting that the thermometers used stabilised within a similar period. This would be consistent with Lumby (1928) who notes that a thermometer will indicate the water temperature in one minute with reasonably active stirring. On-deck periods for pre-WWII bucket measurements are generally assumed to have been much longer than

this based on recommended waiting periods for thermometer equilibration of 2–3 min or more (FP95). However, written instructions do not necessarily equate to the actual practices of mariners. Schott (1893) suggests that periods of 3–4 min noted in some books are much too long for most instruments in use given the potential for bucket cooling. He reports waiting an average of 1 min before obtaining a reading. A post-WWII source, HMSO (1956), states that thermometers attain a steady reading after about 30 s with vigourous stirring, with Stubbs (1965) noting that this was the waiting period respected on a British ocean weather ship. The response time of liquid-in-glass thermometers is almost entirely dependent on bulb diameter (Nicholas and White, 2001), being longer for larger diameter bulbs. Placing greater weight on actual reports of the duration of thermometer stabilisation periods over recommendations from observing instructions, I suggest that the on-deck period would generally have been around 1 min.

6 Synthesis and conclusions

Various techniques have been used to measure sea surface temperature since the mid-19th century. Methods differ in terms of platform, measurement depth and extent of automation (e.g. manual observation and recording). Shipboard methods include temperature measurement of bucket samples and of engine cooling water intakes. Methodological details are generally poorly documented for both methods, but particularly so for intakes. The latter not being a dedicated scientific method, instruments and procedures have likely varied widely between ships. Many details of shipboard methods show general changes over time. Indeed ships themselves have clearly altered dramatically since the 1850s, with a general increase in average speed, freeboard and the deepest drafts. Intake depths on modern voluntary observing ships appear typically around 7–10 m, although can exceed 15 m.

Accurate temperatures can be obtained with either the bucket or intake method. However, measurements cannot be expected to be of high accuracy or precision when obtained by untrained sailors using poorly-calibrated, low-resolution thermometers. This is not of major concern with regards the accuracy of large-scale area-average SST records since random and systematic errors associated with individual observations and instruments tend to cancel out across large numbers of observations. The literature suggests a tendency for the lowest resolution liquid-in-glass thermometers in use to have generally been poorer for intake readings than for bucket measurements. There are reports of intake thermometers graduated in intervals of only 2 or 5° F (B28; Saur, 1963), consistent with the idea that EIT readings would only have been needed to accuracy of 1–2 °C in their traditional engine monitoring role. However, whether intake thermometers were generally of poorer accuracy and precision than

those used for bucket measurements is unclear. Saur (1963) describes a study in which several were found to read in systematic error by between −1 and +2 °C, while B28 notes that a galley thermometer used for bucket measurements aboard the *Finland* read in variable error between −0.3 and +0.4 °C.

Bucket temperatures have generally been found to average a few tenths of °C cooler than simultaneous intake temperatures in field studies, although with considerable scatter amongst the individual bucket-intake differences (e.g. James and Fox, 1972). Such variability is likely, at least in part, due to poor observation and recording with thermometers of variable accuracy and resolution. Such noise does not necessarily negate the accuracy of average differences, however. Average bucket-intake differences are found to vary widely, both between ships and between cruises on the same ship (Saur, 1963; Tabata, 1978a). Crucially though, individual errors in bucket and intake temperatures cannot be directly distinguished from relative bucket-intake differences. To do so requires supplementary accurate in situ temperatures and these have rarely been obtained in field comparisons. In their absence it is difficult to distinguish, for instance, between contributions from thermometric errors, temperature change of bucket samples and near-surface temperature gradients in non-zero average differences found for individual ships and cruises.

The magnitude of bucket cooling depends on the cooling rate and the time elapsed between sampling and thermometer reading (the exposure time). Field and lab experiments suggest samples in small-volume canvas buckets can cool at rates of 0.05–0.10 °C min^{-1} or more (e.g. 0.2 °C min^{-1}). Wind tunnel experiments (Ashford, 1949; Roll, 1951b) have shown cooling to be faster under larger sea–air temperature contrasts and at higher wind speeds. From physical principles we would expect cooling rates to vary with bucket type and construction (e.g. material, presence of a lid) and sample volume. Different buckets can have quite different volumes, so the influence of each of these factors is often unclear in field and lab experiments. Canvas buckets of volumetric capacity between 2 and 12 L are known to have been used (B28; Schott, 1893; Uwai and Komura, 1992).

Systematic warm error in intake temperatures is also a plausible explanation for negative average bucket-intake differences. For instance, Tabata (1978b, d) found EIT to average 0.3 ± 1.2 °C warmer than accurate in situ temperatures on a research vessel, while Brooks (1928) found EIT to be overly-warm by 0.7 °C on average on an ocean liner. Given the large magnitude of these errors, it is possible that the principal cause of the 0.3 °C average intake-bucket difference found by James and Fox (1972) is EIT error rather than bucket cooling. Note that the general origin of systematic warm errors in EIT is poorly known, with it being demonstrated in Part 2 that warming of intake seawater by hot engine room air is an unlikely explanation.

Bucket adjustments have been applied to historical SST datasets in an attempt to reduce supposed bucket cooling

error. In the case of the Hadley Centre SST datasets (e.g. HadSST3), these were derived using variants of the FP95 bucket models. These models are particularly sensitive to the choice of exposure time, an interval comprised of a hauling period and an on-deck phase. Based on the literature, there is scope for both of these periods to have ranged between tens of seconds and a few minutes. For their wooden bucket adjustments, FP95 assume a 1 min hauling phase and a 3 min on-deck period, giving a total exposure time of 4 min. They support their use of on-deck periods of several minutes by citing instructions recommending waiting periods for thermometer equilibration of 2–3 min or more. However, the few reports we have detailing actual durations of thermometer reading periods suggest they could typically have been only around a minute in duration (Schott, 1893; B28). Since this is uncertain, I suggest that the range of possible average exposure times used to derive bucket adjustments be widened to allow for periods of 1–2 min.

Bucket-intake field comparisons are of variable relevance to the bulk of the historical SST data in ICOADS. Studies vary greatly in terms of the type(s) of vessel used (e.g. scientific or merchant; sail or motor), the methods assessed, and the spatial and temporal coverage of measurements (e.g. region(s), season(s) and number of observations). Further, a minor variant of a historical method may have been tested (e.g. a particular type of bucket) that was not in widespread use. This in itself is difficult to assess given the lack of metadata accompanying historical SST measurements. In terms of deducing the extent to which bucket and intake errors are due to actual change in sample temperature, the utility of several field studies is reduced by poor measurement quality (e.g. the use of an inaccurate ships' thermometer in the B28 study). Accurate measurement of sample temperature change requires use of well-calibrated, high precision, fast-response thermometers. Such limitations of previous studies can be addressed through new field experiments of the type presented in Part 2.

References

Ashford, O. M.: A new bucket for measurement of sea surface temperature, Mar. Observ., 19, 107–112, 1949.

Beggs, H. M., Verein, R., Paltoglou, G., Kippo, H., and Underwood, M.: Enhancing ship of opportunity sea surface temperature observations in the Australian region, J. Operat. Oceanogr., 5, 59–73, 2012.

Brooks, C. F.: Observing Water-Surface Temperatures at Sea, Mon. Wea. Rev., 54, 241–253, doi:10.1175/1520-0493(1926)54<241:OWTAS>2.0.CO;2, 1926.

Brooks, C. F.: Reliability of different methods of taking sea-surface temperatures, J. Washington Acad. Sci., 18, 525–545, 1928.

Brooks, C. F.: Chapter 14 – Oceanograpy and Meteorology, in: Physics of the Earth – V: Oceanography, Bulletin of the National Research Council, 85, National Academy of Sciences, Washington, DC, USA, 457–519, 1932.

Collins, C. A., Giovando, L. F., and Abbott-Smith, K. B.: Comparison of Canadian and Japanese Merchant-Ship Observations of Sea-Surface Temperature in the Vicinity of Present Ocean Weather Station "P", 1927–1933, J. Fish. Res. Bd. Can., 32, 253–258, doi:10.1139/f75-023, 1975.

Crawford, W., Galbraith, J., and Bolingbroke, N.: Line P ocean temperature and salinity, 1956–2005, Prog. Oceanogr., 75, 161–178, doi:10.1016/j.pocean.2007.08.017, 2007.

Ecology and Environment, Inc.: Final Environmental Impact Statement/Environmental Impact Report for the Cabrillo Port Liquified Natural Gas Deepwater Port License Application: Volume III: Appendix D5, California State Lands Commission, Sacramento, CA, USA, 2007.

Emery, W. J., Castro, S., Wick, G. A., Schluessel, P., and Donlon, C.: Estimating Sea Surface Temperature from Infrared Satellite and In Situ Temperature Data, Bull. Amer. Meteor. Soc., 82, 2773–2785, doi:10.1175/1520-0477(2001)082<2773:ESSTFI>2.3.CO;2, 2001.

Fedorov, K. N. and Ginsburg, A. I.: The Near-surface Layer of the Ocean, VSP, Utrecht, The Netherlands, 1992.

Folland, C. K.: Assessing Bias Corrections in Historical Sea Surface Temperature using a Climate Model, Int. J. Climatol., 25, 895–911, doi:10.1002/joc.1171, 2005.

Folland, C. K. and Parker, D. E.: Correction of instrumental biases in historical sea surface temperature data, Q. J. R. Meteorol. Soc., 121, 319–367, doi:10.1002/qj.49712152206, 1995.

Hagart-Alexander, C.: Chapter 21 - Temperature Measurement, in: Instrumentation Reference Book (4th Edition), edited by: Boyes, W., Butterworth-Heinemann, Burlington, MA, USA, 269–326, 2010.

Hénin, C. and Grelet, J.: A merchant ship thermo-salinograph network in the Pacific Ocean, Deep-Sea Res. Pt. I, 43, 1833–1855, doi:http://dx.doi.org/10.1016/S0967-0637(96)00084-2, 1996.

HMSO: Handbook of meteorological instruments part 1, 404–407, HMSO, 1956.

James, R. W. and Fox, P. T.: Comparative Sea-Surface Temperature Measurements, WMO, Geneva, Switzerland, 1972.

James, R. W. and Shank, M. K.: Effect of Variation of Intake Depths on Water Injection Temperatures, Marine Sciences Department, US Naval Oceanographic Office, Washington, DC, USA, 1964.

Jones, P. D. and Bradley, R. S.: Climatic variations in the longest instrumental records, in: Climate since A.D. 1500, edited by: Bradley, R. S. and Jones, P. D., Routledge, London, UK, 1992.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res., 117, D05127, doi:10.1029/2011JD017139, 2012.

Kennedy, J. J., Brohan, P., and Tett, S. F. B.: A global climatology of the diurnal variations in sea-surface temperature and implications for MSU Temperature Trends, Geophys. Res. Lett., 34, L05712, doi:10.1029/2006GL028920, 2007.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. measurement and sampling uncertainties, J. Geophys. Res., 116, D14103, doi:10.1029/2010JD015218, 2011a.

Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, J. Geophys. Res., 116, D14104, doi:10.1029/2010JD015220, 2011b.

Kent, E. C. and Kaplan, A.: Toward Estimating Climatic Trends in SST. Part III: Systematic Biases, J. Atmos. Oceanic Technol., 23, 487–500, doi:10.1175/JTECH1845.1, 2006.

Kent, E. C. and Taylor, P. K.: Toward Estimating Climatic Trends in SST. Part I: Methods of Measurement, J. Atmos. Oceanic Technol., 23, 464–475, doi:10.1175/JTECH1843.1, 2006.

Kent, E. C., Woodruff, S. D., and Berry, D. I.: Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS, J. Atmos. Oceanic Technol., 24, 214–234, doi:10.1175/JTECH1949.1, 2007.

Kent, E. C., Kennedy, J. J., Berry, D. I., and Smith, R. O.: Effects of instrumentation changes on sea surface temperature measured in situ, Wiley Interdisciplinary Reviews: Climate Change, 1, 718–728, doi:10.1002/wcc.55, 2010.

Kirk, T. H. and Gordon, A. H.: Comparison of intake and bucket methods for measuring sea temperature, Mar. Observ., 22, 33–39, 1952.

Lumby, J. R.: The surface sampler, an apparatus for the collection of samples from the sea surface from ships in motion, with a note on surface temperature observations, J. Cons. Perm. Int. Explor. Mer., 2, 332–342, 1927.

Lumby, J. R.: Modification of the Surface Sampler with a view to the Improvement of Temperature Observation, J. Cons. Perm. Int. Explor. Mer., 3, 340–350, 1928.

Matthews, J. B. R. and Matthews, J. B.: Comparing historical and modern methods of sea surface temperature measurement – Part 2: Field Comparison in the central tropical pacific, Ocean Sci., 9, 695–711, doi:10.5194/os-9-695-2013, 2013.

Merchant, C. J., Llewellyn-Jones, D., Saunders, R. W., Rayner, N. A., Kent, E. C., Old, C. P., Berry, D., Birks, A. R., Blackmore, T., Corlett, G. K., Embury, O., Jay, V. L., Kennedy, J., Mutlow, C. T., Nightingale, T. J., Carroll, A. G., Pritchard, M. J., Remedios, J. J., and Tett, S.: Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers, Adv. Space Res., 41, 1–11, 2008.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, J. Geophys. Res., 117, D08101, doi:10.1029/2011JD017187, 2012.

Nicholas, J. V. and White, D. R.: Traceable Temperatures: An Introduction to Temperature Measurement and Calibration (2 Edn), Wiley, Chichester, UK, 2001.

Parker, D. E.: Hadley Centre Internal Note 34: Summary of information on sampling apparatus for sea surface temperature measurement, Meteorological Office, Bracknell, UK, 1993.

Piip, A. T.: A Critical Description of the CSIRO Sea Surface Temperature and Salinity Sampling Program from Merchant Ships, Marine Laboratory, Cronulla, NSW, Australia, 1974.

Rao, P. K., Smith, W. L., and Koffler, R.: Global sea-surface temperature distribution determined from an environmental satellite, Mon. Wea. Rev., 100, 10–14, doi:10.1175/1520-0493(1972)100<0010:GSTDDF>2.3.CO;2, 1972.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670, 2003.

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T. J., and Tett, S. F. B.: Improved Analyses of Changes and Uncertainties in Sea Surface Temperature Measured In Situ since the Mid-Nineteenth Century: The HadSST2 Dataset, J. Climate, 19, 446–469, doi:10.1175/JCLI3637.1, 2006.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An Improved In Situ and Satellite SST Analysis for Climate, J. Climate, 15, 1609–1625, 2002.

Roll, H. U.: Water temperatures measurements on deck and in the engine room. Ann. Meteor., 4, 439–443, 1951a.

Roll, H. U.: The accuracy of measuring water temperature with the water scoop thermometer (Marinepütz – German scoop thermometer), Ann. Meteor., 10–12, 480–482, 1951b.

Saur, J. F. T.: A Study of the Quality of Sea Water Temperatures Reported in Logs of Ships' Weather Observations, J. Appl. Meteor., 2, 417–425, doi:10.1175/1520-0450(1963)002<0417:ASOTQO>2.0.CO;2, 1963.

Schott, G.: Scientific results of a research voyage at sea carried out in the years 1891 and 1892, Justus Perthes, Gotha, Germany, 1893.

Smith, T. M. and Reynolds, R. W.: Bias Corrections for Historical Sea Surface Temperatures Based on Marine Air Temperatures, J. Climate, 15, 73–87, doi:10.1175/1520-0442(2002)015<0073:BCFHSS>2.0.CO;2, 2002.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21, 2283–2296, doi:10.1175/2007JCLI2100.1, 2008.

Stubbs, M. W.: The standard error of a sea surface temperature as measured using a canvas bucket, Meteorological Magazine, 94, 66–69, 1965.

Stupart, F., Patterson, J., and Smith, H. G.: Ocean surface-water temperatures – methods of measuring and preliminary results, Bulletin of the National Research Council, 68, 1929.

Tabata, S.: An Evaluation of the Quality of Sea Surface Temperatures and Salinities Measured at Station P and Line P in the Northeast Pacific Ocean, J. Phys. Oceanogr., 8, 970–986, doi:10.1175/1520-0485(1978)008<0970:AEOTQO>2.0.CO;2, 1978a.

Tabata, S.: An examination of the quality of sea-surface temperatures and salinities observed recently in the northeast Pacific Ocean, Pacific Marine Science Report 78-3, Institute of Ocean Sciences, Sidney, BC, Canada, 1978b.

Tabata, S.: Comparison of Observations of Sea Surface Temperatures at Ocean Station P and NOAA Buoy Stations and Those Made by Merchant Ships Traveling in Their Vicinities, in the Northeast Pacific Ocean, J. Appl. Meteor., 17, 374–385, doi:10.1175/1520-0450(1978)017<0374:COOOSS>2.0.CO;2, 1978c.

Tabata, S.: On The Accuracy of Sea-Surface Temperatures and Salinities Observed in the Northeast Pacific Ocean, Atmos. Ocean, 16, 237–247, doi:10.1080/07055900.1978.9649032, 1978d.

Tauber, G. M.: The comparative measurements of sea-surface temperature in the USSR, Tech. Rep. 103, WMO, 1969.

Thompson, D. W. J., Kennedy, J. J., Wallace, J. M. and Jones, P. D.: A large discontinuity in the mid-twentieth century in observed global-mean surface temperature, Nature, 453, 646–649, doi:10.1038/nature06982, 2008.

Uwai, T. and Komura, K.: The Collection of Historical Ships' Data in Kobe Marine Observatory, in: Proceedings of the International COADS Workshop, Boulder, CO, USA, 13–15 January 1992, 1992.

Wentz, F. J., Gentemann, C., Smith, D., and Chelton, D.: Satellite Measurements of Sea Surface Temperature Through Clouds, Science, 288, 847–850, doi:10.1126/science.288.5467.847, 2000.

WMO: Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8): Part II – Observing Systems: Chapter 4 – Marine Observations, WMO, Geneva, Switzerland, 2008.

Woodruff, S. D., Diaz, H. F., Kent, E. C., Reynolds, R. W., and Worley, S. J.: The Evolving SST Record from ICOADS, in: Climate Variability and Extremes during the Past 100 Years, edited by: Brönnimann, S., Luterbacher, J., Ewen, T., Diaz, H. F., Stolarski, R. S., and Neu, U., Advances in Global Change Research Vol. 33, Springer, Dordrecht, The Netherlands, 2008.

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Freeman, J. E., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., and Wilkinson, C.: ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive, Int. J. Climatol., 31, 951–967, doi:10.1002/joc.2103, 2011.