

## ***Interactive comment on “Comparing historical and modern methods of Sea Surface Temperature measurement – Part 1: Review of methods, field comparisons and dataset adjustments” by J. B. R. Matthews***

**R. Matthews**

georobin@uvic.ca

Received and published: 19 December 2012

I thank AR1 for their extensive comments. Evidently there has been some confusion as to exactly what is and what is not being suggested in the paper. I will endeavour to clarify this here and improve the expression of my ideas in a revised manuscript.

General comments >"This paper has the potential to be a useful contribution to the literature but will require revision before publication. Some references that are not usually quoted have been identified which is helpful. The review is however presently

C1346

selective in picking papers, or parts of papers, that support the thesis that bucket measurements are accurate to  $\pm 0.1^\circ\text{C}$  and engine intake measurements are so noisy as to be useless."

The main impetus for conducting the review was to critically assess the evidence for cooling of bucket samples and warming of intake seawater. The intention was to be fairly comprehensive in literature coverage, although with focus given to the most frequently cited papers (e.g. Brooks, 1926; James and Fox, 1972; Saur, 1963) and a few papers that have been little considered in the modern literature (e.g. Hénin and Grelet, 1996; James and Shank, 1964; Piip, 1974). Recently I have been provided with several additional hard-to-obtain papers that were unavailable to me at the time of writing. These include Amot (1954), Brooks (1928), Kirk and Gordon (1952), Lumby (1927), Roll (1951) and Tauber (1969). These will be covered in the revised manuscript.

The thesis of the paper is not that bucket measurements are accurate to  $0.1^\circ\text{C}$  or that engine intake temperatures (EIT) are necessarily useless. Bucket temperatures clearly have not been consistently accurate to  $0.1^\circ\text{C}$  since, as noted in the paper, they have often only been measured to a half or whole  $^\circ\text{C}$  or  $^\circ\text{F}$ . It is also not my hypothesis that all EIT are useless. Indeed, in Part 2 it is noted that EIT can be useful 'where accurate and of known sampling depth'. Unfortunately few historical EIT meet these criteria, with those that do not being of little utility. However, given that engine intake measurements are unlikely to be affected by heating from warm engine room air (as demonstrated in Part 2), accurate subsurface temperatures can be obtained from thermosalinographs on voluntary observing ships. I must admit I have been surprised by the general poor quality of modern EIT measurements given the wide range of scientific applications for which they are used.

>"This leads to a conclusion (in Part 2) that bucket measurements require no adjustment and all engine intake observations should be discarded, which is not supported by the full literature. For example, Ashford (1948) and Roll (1951) made observations of temperature change in wind tunnels for several types of bucket and Brooks (1926)

C1347

describes "errors closely related to the depression of wet bulb or air temperature below the water temperature".

Part 2 is actually proposing a paradigm shift in the way we think about historical SST observations. It suggests that adjustment of low precision, inaccurate data is inappropriate and that this data should rather be discarded. AR1 does not provide any evidence to suggest historical EIT are of reasonable accuracy and thus should be retained. Critically, the extent to which subsurface temperatures at several meters depth can be assumed equivalent to those in the upper few tens of centimetres is unclear and has been little investigated on a large-scale. We can and should be more selective about the data we use to construct SST datasets, using only that which is of reasonable quality (i.e. measured to at least 0.1°C). Further, the paper is not arguing that bucket samples cannot cool post-sampling and pre-measurement, but is rather questioning the magnitude of this cooling, particularly the substantial cooling simulated by Folland and Parker (1995; hereafter FP95).

To be clear, the conclusions of the paper are as follows:

- 1) Reported average bucket-intake temperature differences are mostly negative, although the individual differences show considerable spread between positive and negative values. Given this spread and the variable quality of the actual observations, non-zero average differences cannot be attributed to physical causes with much certainty.
- 2) Canvas bucket samples can cool between collection and temperature measurement, with the magnitude being dependent on the duration of this period (the exposure time) and the cooling rate. Estimates of hauling and thermometer equilibration periods from the literature suggest exposure time varied between tens of seconds and a few minutes.
- 3) We should not rely on the historical literature to assess whether rapid bucket cooling or engine room warming of intake seawater are physically plausible. Several of the field experiments that have attempted to assess the accuracy of bucket temperatures suffer

C1348

from substantial methodological limitations (e.g. poor observing, historically-unrealistic methods). There is thus a need for new field experiments, with testing of buckets of variable type and volume.

Note that while the last conclusion was not directly stated in the paper, it was very much implied.

>"It has long been known that bucket measurements can be accurate, especially if the bucket is well-designed, the exposure time is short and the response time of the thermometers used is short (e.g. Brooks 1926, Saur 1963). However the important question for the adjustment of real data is what the biases are likely to be in the observations actually collected. Folland and Parker (1995) were careful to give more weight to those observing instructions that were likely to have been relevant to a large number of the observations made. In the present review there is more focus on studies where particular care was taken (e.g. Brooks took quick samples with a large bucket, the Saur study used a bucket specially designed by Scripps and Tabata considers observations made at a weather station). It is not clear therefore that the conclusions of this review can be directly applied to the observations in ICOADS or the SST datasets, like HadSST3, derived from them."

I am not suggesting that the minimal errors found for bucket temperatures taken aboard weather ships and oceanographic research vessels are directly applicable to those from merchant ships. Further, I do not think the review gives unreasonable credence to field experiments in which observation was careful. These studies provide the most reliable results for assessing whether bucket and intake seawater samples change temperature prior to measurement, this being one of the main foci of the paper.

Also, unlike FP95, I do not think we can take observing instructions to be representative of actual observing practices aboard merchant vessels. For instance, I think it unreasonable to assume the on-deck period for bucket measurements generally lasted several minutes based on recommended waiting periods for thermometer equilibration.

C1349

Critically though, I think inaccurate data should be discarded rather than adjusted.

>"However the paper does provide some useful focus on the expected differences between accurate measurements at the surface and at depth. This has not been entirely ignored in the literature, most studies make it very clear that differences are expected between buckets and engine intakes because of the depth (e.g. Brooks, Saur, James and Fox 1972, Folland and Parker, Kennedy et al. 2011b). However the large ship-to-ship differences evident in such studies have precluded the analysis of variations with environmental parameters expected to be important for relating temperature at depth to the surface. It should be noted that most of the evidence for particularly serious problems with engine intakes comes from the 1960s or earlier (although Kent et al. 1993 show some fairly large differences for engine intake measurements from 1988-1990)."

Relating environmental parameters such as wind speed to bucket-intake temperature differences is a very indirect and unreliable method of looking at near-surface temperature gradients. What we need are accurate in situ measurements of the vertical temperature structure alongside the bucket and intake temperatures. With regards changes in the accuracy of intake measurements over time, an important consideration is the degree of automation in their observation and recording. This has not been thoroughly assessed.

>"To become suitable for publication two main improvements are required: 1) A more balanced and complete review of the available literature"

This point has already been addressed. The manuscript will be improved by more clearly elucidating what is and what is not being suggested. The revised version will also include discussion of the aforementioned literature recently made available to the author.

>"2) A more careful consideration of how the different types of studies relate to the historical temperature record as measured"

C1350

That the reasonable accuracy of bucket temperatures found in some studies may not be characteristic of those in ICOADS as a whole will be made clear. Addressing this point thoroughly, however, would be a very large task that is well beyond the scope of this review. It would involve analysing the available metadata to extract details such as the types of ships reporting and the precision to which measurements were made. This task might best be undertaken as part of the new international project 'Historical Ocean Surface Temperatures: Accuracy, Characterisation and Evaluation (HOSTACE)'.

Specific comments >"Abstract, line 3: post-industrial period?"

Thanks for spotting this. The term was intended to refer to the period from the start of the Industrial Revolution to the present, but is evidently ambiguous and so will be changed.

>"Abstract, lines 3-4: Most SST observations have come from the platforms listed at all times."

Thanks for spotting this - it will be reworded.

>"Abstract, lines 8-end: As outlined above I do not believe that the summary presented in the abstract is a fair representation of the literature."

Note that I am not claiming that average bucket-intake temperature differences reported in the literature are statistically insignificant. Rather, I am saying that the considerable heterogeneity in the data precludes attribution of physical causes for non-zero averages.

>"Page 2954, lines 1-3: Kennedy et al. (2011b) describe the difficulty of measuring from large and fast ships. The information reported there is that UK ships were advised to use ERI when the speed is > 15 kt."

This is an interesting insight. I would certainly not want to be doing bucket deployments from a ship travelling at such speeds.

C1351

>"Page 2955, lines 1-3: Reference needed for information about dual intakes (Saur?)"

I do not think this needs a reference as it is common knowledge within the marine industry.

>"Page 2955, lines 16-17: Reference needed for diameter of intake pipes"

This is the range of intake pipe diameters for seawater partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>) systems aboard oceanographic research vessels. For ease of reference, I shall quote the diameter of the scientific seawater intake pipe on the Seamans in the revised paper. The critical consideration with regards intake warming is the volumetric flow rate. Note that flow velocity is fairly independent of pipe diameter, typically being around 1-1.5 m/s.

>"Page 2956, line 23: usually called Ocean Data Acquisition Systems (ODAS)"

The phrase 'Automated Ocean Acquisition Systems' will be replaced with this term.

>"Page 2956, line 22-end of paragraph: The information presented here refers to the observations available in ICOADS. There were probably earlier measurements made by these different platforms which are not in ICOADS. The text should make it clear that the start dates are for ICOADS observations."

Thanks for alerting me to this distinction. This will be made clear in the revised version.

>"Page 2958, lines 3-9: SST measurements from AVHRR are adjusted for biases using in situ observations, as described in the Reynolds reference quoted and in many other papers by Reynolds and others. AVHRR cannot be described as accurate, e.g. from Reynolds conclusions "However, because of large potential biases in satellite retrievals, accurate bias corrections are needed particularly for climate studies." The SST from ATSR is more suitable for applications requiring higher accuracy (e.g. Merchant et al. 2012, <http://www.agu.org/pubs/crossref/2012/JC008400.shtml>)"

Thanks for spotting this – it will be amended. I was actually referring to a quote from

C1352

Reynolds (1999) which states: "In late 1981, accurate SST retrievals became available from the Advanced Very High Resolution Radiometer (AVHRR) instrument...". In hindsight I agree that describing SST measurements derived from AVHRR retrievals as accurate is inappropriate.

>"Page 2957, Section 3.1: As noted in major points above, this review is not adequate to support the conclusions drawn."

The review provides by far the most detailed analysis of the literature ever conducted, with discussion of the full range of findings and conclusions reported by both oft-cited and overlooked papers. Important methodological details of these studies are highlighted (e.g. the precision to which thermometers were read) to enable discussion of the reliability and wider relevance of their conclusions. Again, I think AR1's point comes from some confusion over exactly what the conclusions of the paper are.

>"Page 2960, Section 3.2: Some of the information presented here is a valuable addition to the literature. It is a shame however that none of the data are presented. I realise that the somewhat unusual method reduces the utility of the observations, but the data may still be useful, especially if there is additional environmental information available. The Mk II bucket was issued to UK observers for many years."

I think it would be better to conduct original field experiments rather than use these data given the unrealistic and poorly-documented method used. The raw data may have been retained by the Sea Education Association.

>"Page 2962, Section 3.3: Ashford (1948) compared several different types of bucket and there are other studies too."

Thank you for alerting me to the study by Ashford. It will be covered in the revised manuscript.

>"Also the comments about likely observing practice are rather sweeping. Yes there is anecdotal evidence of poor observing practice, but equally Meteorological Agencies

C1353

have invested much effort (since at least 1853) in trying to make mariners understand the importance of the observations and how to take them properly. Some observers will have done as instructed, some will not."

There are actually (at least) two separate issues here. First, the care with which the thermometers were read and the values recorded, and second the time allowed for thermometer equilibration. I am actually assuming that thermometer reading and temperature recording was generally careful for bucket measurements, while assuming the time allowed for thermometer equilibration was short (one minute or less). We don't actually need to invoke impatience amongst marine observers to claim on-deck periods were short. FP95 cite evidence to suggest slow-response historical thermometers could indeed equilibrate within one minute.

As for manual EIT readings, we cannot assume these were carefully observed and recorded given that this was not a requirement in their traditional role as an engine monitoring tool. For instance, Brooks (1928) reports that an engineer told him intake temperatures read to within 5°F would be sufficient.

The general point I am making is that we should not expect data obtained by untrained, unsupervised non-scientist observers at sea to be of high quality. We can make this assumption for meteorological observations obtained by amateur scientists on land since they collect them out of a passion for doing so. It would be useful to consult retired and active mariners for insight on the care with which bucket and intake temperatures were obtained aboard observing ships.

>"Page 2966, lines 21-23: There are several studies which have tried to separate the contribution of variations with depth from bucket cooling (e.g. James and Fox 1972)."

Yes, some studies have attempted to relate bucket-intake temperature differences to variations in intake depth. I will clarify this in the revised version.

>"Page 2967, lines 6-7: If observations can be considered independent or only

C1354

partly dependent then the standard deviation is not the relevant measure. See e.g. <http://www.metoffice.gov.uk/hadobs/hadsst3/faq.html>"

James and Fox report that 68% of their intake-bucket temperature differences fall within  $\pm 0.9^{\circ}\text{C}$  of the mean, which I referred to as the standard deviation. This 'percentage of observations' measure is the only type of dispersion indicator they report. It would not be possible to calculate another measure of spread without their raw data. Note that their histogram (Table III) appears to be slightly in error, with some observations unaccounted for. I think it important to report some measure of spread for bucket-intake temperature differences, and the standard deviation remains a good indicator of this even if better ones do exist. Note that Kennedy et al. (2011b) do not provide any measure of spread in their table of ERI biases reported in the literature (Table 3).

---

Interactive comment on Ocean Sci. Discuss., 9, 2951, 2012.

C1355