

ABSTRACT	1
1. Background and Introduction	2
2. Data	4
2.1 Data Sources	4
2.2 Ship Tracking	6
2.3 Quality Control	8
2.4 Metadata Information	9
3. SST Measurement Methods	11
3.1 Bucket Measurements of SST	11
3.2 Engine Intake Measurements of SST	12
3.3 Hull Sensor Measurements of SST	14
3.4 The regional distribution of SST measurement type	14
4. Data Quality Characteristics	15
4.1 Introduction	15
4.2 Data Quality by Country	15
4.3 Data Quality by Ship Parameters	16
4.4 Data Quality by Environmental Parameters	18
5. Analysis Methods	19
5.1 Effect of errors on analysis	19
5.2 Comparison of climatological fields	20
5.3 Comparison with model fields	20
5.4 Observational pairs	20
6. Results	27
6.1 Comparison of fields generated from coincident pairs of reports	27
6.2 Statistical Analysis of Individual Observational Pairs	30
6.3 Referencing Individual Observations to Climatology	31
6.4 Biases in Air Temperature Data	31
6.5 Comparison of SST Results with Previous Studies	33
7. Discussion and Conclusions	38
7.1 Accuracy of Bucket SST Measurements	38
7.2 Accuracy of Engine Intake SST Measurements	40
7.3 Accuracy of Marine Air Temperature Measurements	42
7.4 Future Research	43
Acknowledgements	43
References	44
Tables	50
Figures	58
Appendix	97

ASSESSMENT OF BIASES IN MERCHANT SHIP SURFACE TEMPERATURES

Elizabeth C. Kent, Southampton Oceanography Centre, Southampton, UK.

Abstract

Weather observations from Voluntary Observing Ships (VOS) have been analysed in an attempt to identify biases resulting from different observing practices. The observations were taken from the Comprehensive Ocean Atmosphere Dataset (COADS) for the period 1980 to 1997. A new quality control procedure was developed and implemented based on a comparison of local data rather than on climatological limits and individual ships were tracked to remove data with corrupt positions. Metadata containing information on measurement methods from a World Meteorological Organisation catalogue of VOS was associated with individual ship observations to enhance the information already contained within the COADS.

Observational pairs were constructed from the COADS where reports were co-located within 50 km and 1 hour. This procedure reduced uncertainty about the comparisons of different measurement methods due to spatial and temporal variability. Sea surface temperature (SST) reports made using a bucket and thermometer were shown to be biased cool in conditions where surface heat loss was expected to be large. Bucket SST reports, which sample only the near surface, were relatively warm under conditions where a shallow diurnal warm layer in the ocean might be expected to form. Biases in the nighttime bucket SST reports were quantitatively assessed using a statistical technique that accounted for the error structure of the dataset. This suggested that at moderate wind speeds bucket SST data are biased by an approximately constant fraction of the air - sea temperature difference. SST derived from the engine intake thermometer were much noisier than those measured using buckets but, on average, no significant bias could be detected. This contradicts earlier studies which are comprehensively reviewed.

Biases in air temperature measurements depend on the exposure of the sensor (which depends on recruiting country), on the incident shortwave radiation, the ventilation of the sensor and the past heating of the sensor. Regional variations in the height of the air temperature measurement are shown to be significant as are changes in observation height with time.

1. BACKGROUND AND INTRODUCTION

Recent comparisons between global and tropical temperature trends at the surface and aloft have suggested that either the atmospheric lapse rate has increased, i.e. surface warming has exceeded tropospheric warming, or that biases have arisen in observations made from global observing networks (Christy et al. 2001). This is problematic for the detection and attribution of climatic changes over the ocean. To assess the possibility of biases in the *in situ* data, it is necessary to monitor the observed changes in sea surface temperature (SST) with respect to the different *in situ* platforms and measurement techniques used. In parallel, it is necessary to study effects on marine air temperature (which is also used to monitor global temperature trends) and on air-sea temperature differences which are an important influence on heat fluxes between the atmosphere and ocean. This need is highlighted by recent unexplained cooling of nighttime marine air temperatures, relative to collocated sea surface temperatures, in parts of the Southern Hemisphere (IPCC 2001).

Comparisons between recent observations of *in situ* SST from Voluntary Observing Ships (VOS) using buckets and using other techniques have been hampered until recently by a lack of information on the measuring techniques used on individual ships (Folland et al. 1993). However, this metadata has recently become more widely available, making thorough intercomparisons possible. Assessment of *in situ* data by comparison with model output was carried out by Kent et al. (1993a) using data from a small number of ships participating in the VOS Special Observing Project - North Atlantic (VSOP-NA). Here we aim for a wider analysis of biases in SST and air temperature during the past two decades. We assess SST and marine air temperatures from different *in situ* marine instruments between 1980 and 1997 in the Comprehensive Ocean Atmosphere Dataset (COADS) / UK Met Office Marine Data Bank Blend. The eventual aim is to identify any biases associated with particular measurement methods and relate those biases to environmental conditions, for example the air-sea heat fluxes.

The data used contain large random errors (Kent et al. 1999) and data from some ships, buoys and platforms can have large biases. The quality control flags within the dataset are based on differences from climatology and have known problems (Wolter 1997). The approach taken in this study is to quality control the data without using climatology but to use nearby data to assess the quality of each report. As the study will rely heavily on identifying data from, and the instruments carried by, individual ships, efforts have been made to identify ships or buoys that are reporting strongly and consistently biased reports. These biased reports may arise from a

poorly calibrated instrument, incorrect coding of reports or poor observing practice and need to be removed in order that smaller biases resulting from the types of instrument used and the environmental conditions can be studied. Additionally ship tracking has been performed for the first time in a climatological dataset in order to identify and remove mispositioned reports. This procedure is routine in the production of operational datasets (Reynolds and Smith 1994) but a new method is used in this study which better meets the requirements of a climatological dataset.

Although the COADS contains some information about the instruments used to make measurements (for example for wind speed and SST) the use of an external source of metadata to identify instrument types is vital for other variables (for example air temperature and humidity) and to provide additional information for variables with metadata contained within the dataset. The metadata contained within WMO Report No. 47 (e.g. WMO 1994) is merged with the data reports as required during the analysis to avoid the unnecessary generation of a dataset even larger than that already existing. Using metadata from both within the ship reports ("SI", the SST measurement method indicator) and externally (for those reports from VOS listed in WMO Report No. 47 but with no useful value for SI), a measurement type can be associated with most of the VOS SST reports in the period 1980 to 1997. Using this measurement method information, in combination with the environmental variables within the ship weather report we attempt to quantify any differences between SSTs measured by different methods and to determine the environmental conditions under which these differences are significant.

There is a lack of a good comparison standard for this kind of analysis. Engine room intake data are noisy as this is not a dedicated method of measuring SST and is prone to many different types of errors (e.g. James and Fox 1972). Hull sensor SST data, which are thought to be of the best quality (Kent et al. 1993a), are not yet common, although quantities are increasing with time. Satellite and gridded SST products are not suitable for this type of analysis as they may contain errors and biases which cannot be quantified. So we have developed an analysis method using the bucket and engine-intake SSTs. An analysis dataset has been generated containing co-located pairs of SST reports where one ship has used a bucket to measure the SST and the other ship has reported the engine intake cooling temperature. This mixed-type dataset is used for the main part of the analysis but common-type paired datasets (i.e. bucket-bucket and engine intake-engine intake pairs) are also required to allow the necessary assessment of errors. The analysis is complicated by the fact that the random errors in both SST measurement methods are larger than the possible cooling effect we are trying to detect. Further complicating the calculation is the need to use the measured SST in the calculation of air-sea temperature difference making the errors in these two regression variables correlated.

A parallel study has investigated regional changes in the height of the observing platform through the period 1980 to 1997. There are large regional variations in the size of VOS and hence differing corrections will need to be made to the air temperature for the lapse rate. In addition, over much of the ocean the rate of increase of the height of the observing platform is greater than has been previously thought (Parker and Rayner 2002).

2. DATA

2.1 Data Sources

2.1.1 COADS / UK Met Office Marine Data Bank Blend

This study uses data from the Comprehensive Ocean Atmosphere Dataset (COADS) which has been merged with the UK Met Office Marine Data Bank (MOMDB) and will be referred to here as the COADS (Woodruff et al. 1999). Only reports for the period 1980-1997 have been used, corresponding to the COADS Release 1a (Woodruff et al. 1993). The dataset contains reports from Voluntary Observing Ships (VOS), moored and drifting buoys, platforms and oceanographic measurements. We have only utilised the VOS reports but there should be useful information to be gained from looking at the other types of measurements, particularly in the light of the Emery et al. (2001) comparison of drifting and moored buoy and VOS SSTs. The VOS reports contain only the basic meteorological variables (wind speed and direction, air and sea temperatures, humidity and pressure) and the heat and momentum fluxes need to be calculated from these basic variables using bulk formulae when required. The formulae chosen follow Josey et al. (1999) and are as follows: sensible and latent heat fluxes from Smith (1988), momentum flux from Smith (1980), longwave flux from Clark et al. (1974) and shortwave flux from Dobson and Smith (1988). Additionally the visual estimates of wind speed are converted to the Lindau (1995) visual equivalent scale.

The VOS reports come from a wide range of different ships ranging from small fishing and research vessels to large container ships and tankers (Moat 2002). A wide range of different types of instrumentation are used by these ships and observational practice and sensor locations and exposures are also highly variable. Some information about a subset of the VOS was collected during the VSOP-NA project (Kent and Taylor 1991). The ships participating in the project were probably typical of the larger VOS ships (although some smaller research vessels participated in the VSOP-NA) and showed that not only the type of instruments used were important but also their positioning and how they were used. COADS contains a limited amount

of metadata within the ship reports: the deck number and source identifier (which give information about the source of the data, GTS or logbook, country of origin etc.); platform type (ship, buoy etc.); the method of SST measurement; the type of wind observation (visual or anemometer) and information about the source units and precision. This metadata, whilst useful, is not sufficient to characterise the data and in this study we appeal to an external source of metadata, the World Meteorological Organisation International List of Selected, Supplementary and Auxiliary Ships (WMO Report No. 47) described in the following section.

2.1.2 WMO Report No. 47 Metadata

Metadata from WMO Report No. 47 (e.g. WMO 1994) has been used to provide information about the instrumentation carried by the VOS that contribute reports to the COADS. The metadata is available in digital form for the period 1973 to 1999 and in paper form from 1955. It is hoped that more recent metadata will become available from the WMO in the near future. The metadata is organised by year and by ship callsign which can be matched to individual reports within the COADS. Information is available about the methods of measurement of surface pressure, air and sea temperature and humidity. Also contained is the height of the observing platform and the height of the anemometer (if carried). There is also a list of any other instruments that are carried by the ship, for example a rain gauge, radiosonde equipment or meteorological radar. The metadata from WMO Report No. 47 that have been used in this study are SST and air temperature measurement methods and observing platform heights. Recent changes 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 period analysed here (1980 to 1997) this information was too scarce to be used.

2.1.3 Additional Sources of Data

Other data sources that have been utilised are the Met Office SST climatologies, HadISST (Rayner et al. 2002) and GISST2.2 (Rayner et al. 1996), sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data (Cavalieri et al. 1997), and an ocean mask generated using data from the ETOPO5 5-minute gridded elevation data (NOAA 1988).

2.2 Ship Tracking

2.2.1 Linked records within COADS netCDF Files

In order to efficiently identify data from individual ships within COADS for ship tracking and the association of metadata with each report a 'linked list' was generated. Each COADS report has two extra variables appended, a file and record number identifier for the next report from a ship with the same call sign. Information on the file name and record number of the first report for each ship is kept in a separate ASCII file. Thus data for an individual ship can be identified efficiently and extracted from multiple netCDF files without searching through each monthly file.

2.2.2 Ship tracking method used in this study

Figure 1 shows a flow chart for the ship tracking as used in this study. The program simply extracts all the data for a particular ship into memory and checks each report position with the previous one for consistency. A generous speed of 100 km hr^{-1} is allowed. The maximum speed chosen turns out to be relatively unimportant as most erroneous reports appear at least twice in the dataset, once with a correct position and one or more times with incorrect positions. As the reports are simultaneous the speed calculated is infinite. If the two reports are consistent within these limits, and less than a week apart, both reports are accepted and the next report is examined. A 'last good data' flag is set to show the last report that was accepted for a particular ship. If problems are encountered with consecutive reports then checks are made against the last flagged good data. Special attention has to be paid to the first data in a block of reports for a given ship. A new block is started whenever there is a gap of a week or more in the ships' observations. The method used examines the first ten reports (or all reports if less than 10) in the block and finds the first report that gives the maximum number of good comparisons. This report is set as the 'last good report', the reports before it in the block set to bad data and the tracking continues. Typically 1-2% of the ship data are removed by the ship tracking process. It should be noted that only reports with a valid callsign could be tracked in this way. Any reports without a callsign, or with generic callsigns such as 'SHIP', 'BUOY' or 'PLAT', are left unchecked.

The ship tracking method used in this study was compared with the method used in the generation of the NCEP SST fields (Reynolds and Smith 1994). The tracking method used here

retained more good data than the NCEP method without leaving extra bad data and is therefore preferred.

2.2.3 Results of Ship Tracking

Figure 2 shows the results of ship tracking for the Karl Libknekht (a Russian 'selected ship', callsign EWAE) which reported regularly during the analysis period. Reports plotted in yellow (or light grey if in black and white) have passed the track checking, those in red (or dark grey) have failed. In this example the track checking has removed most reports away from the track of the ship and therefore is working as expected. Reports on or near the ship track (on this global scale) also proved erroneous on examination, but no examples were found of good data being removed by this procedure.

Figure 3 shows qualitatively the effect of the ship tracking on SST in the North Atlantic. Overlaid on a smoothed SST field for January 1990 are the 1° averages of the ship reports that have been removed by the tracking in that month (Figure 3a). Figure 3b shows the same information for July 1990. Both of these plots show that the ship tracking removes data that is often significantly different from the surrounding values. In addition many of the reports removed are close to, or actually on, land (COADS uses a coarse 2° mask to identify landlocked points). Many of the erroneously warm SSTs in the Arctic region are removed.

We decided to attempt to remove data from ships that consistently report poor quality data. Figure 4 shows data from two ships which reported SST outside 4.5 standard deviations of the 10° area monthly mean about 30% of the time in January 1990. The data from these ships are plotted over the smoothed monthly mean SST field for the same month. Both ships are reporting strongly biased SST although most of the time the reports are within the coarse quality control limits set at this time. More than 30% of the data from each ship are also accepted within the COADS 4.5 σ trimming limits. This 'blacklisting' type of quality control allows data from individual ships that report consistently biased data to be removed from the dataset. This is done on a month-by-month basis.

An example of the effect of ship tracking on the SST field for January 1980 is shown in Figure 5. The SST reports have been averaged on a 1° grid using the quality control flags described in Section 2.3. One gridded file uses the information from the ship tracking flags, the other does not. Note that the quality control flags are also calculated for the points failing the track check so if the SST's are outside the quality control limits set they are removed from the

non-track checked data as normal. The difference in the mean SST between the fields caused by the ship tracking is shown in Figure 5a. The ratio of the standard deviations is plotted in Figure 5b. 14% of the monthly mean standard deviation values were increased by the ship tracking due to the removal of mis-placed data that was not very different from the surrounding data. For these 1° averages, the standard deviation was increased by an average of 3%. 14% of the monthly mean standard deviation values were decreased by the ship tracking when the misplaced data was significantly different from the surrounding data, for these 1° averages the standard deviation was decreased by 16%. Overall the track checking reduces the standard deviation of all the 1° values by 2%, the reduction for those 28% of 1° boxes where data were removed was 6%. The track checking therefore increases the quality of the dataset by removing data that is usually different from nearby data.

Figure 6 shows the locations of data removed by track checking but not by the quality control in January 1980. Figure 6a shows the 1° squares where all the data were found to be mispositioned. Figure 6b shows the proportion of data retained after the track checking. 1° squares where no data were removed are not plotted.

2.3 Quality Control

The purpose of this study is to identify biases in the dataset, so we did not attempt rigorous initial quality control. This is because we want to retain observations with biases that result from the use of different types of instruments in different environmental conditions, and retain the distribution of the random uncertainty associated with these biased data. The data that need to be excluded are those that result from misreporting, miscoding, transmission or keying errors, have large biases due to poorly calibrated instruments or any other gross errors. The initial quality control is thus limited. Although COADS contains a 'trimming flag' which indicates whether each report is within 2.8, 3.5 or 4.5 climatological standard deviations of a climatological monthly mean, it was decided not to use this flag as it is known to remove good data in climatologically extreme months (Wolter 1997). Kent et al. (1999) showed in addition that very poor quality data can remain unflagged when ambient conditions deviate from the climatological norm.

Quality control was performed globally on all the reports passing the track check in the following way. Firstly, all data outside 4.5 standard deviations of a 30° zonal mean were removed. This was necessary to remove mispositioned reports in the Arctic regions that were not removed by the track checking (possibly because the report does not contain a callsign). A

wide band was required as erroneous data can be in the majority in these data sparse regions. Drifting buoy and oceanographic data were excluded from the calculation of the mean and standard deviation as sometimes these data were strongly biased and present in large quantities which could result in biased data being retained and good data removed. Once the mean and standard deviation were calculated, the quality control procedure was applied to these data sources. The process was repeated, recalculating the standard deviation without the extreme points until no data were removed. The next pass iteratively removes all data outside 4.5 standard deviations of a 10° zonal mean, the third pass removes all data outside 4.5 standard deviations of a 10° area mean. The blacklisting procedure was then carried out and the removal of data for the 3 different sized areas repeated.

Figure 7 shows the North Atlantic gridded 1° monthly mean and standard deviation SST for January 1985 before (a) and after (b) the quality control procedure outlined above. The quality control smoothes the mean SST removing many of the anomalous values. The standard deviation is reduced by 3%, for the 0.2% of the 1° averages for which there is an increase in standard deviation the mean increase is 9%, for the 5.9% of the 1° averages for which there is a decrease in standard deviation the mean decrease is 55%. The effect of the quality control is therefore significant but localised.

Figure 8 shows the SST reports that have been removed by the blacklisting procedure in January 1985 plotted over the smoothed monthly mean SST field for the same month. Tracks of individual ships showing consistent biases of both signs are clear. The blacklisting of ships should therefore improve the quality of the dataset for analysis.

2.4 Metadata Information

The 'SST Indicator' flag (SI) within COADS gives the method used for measuring the SST for each report. In addition, we have information by ship callsign on the type of instrument used to take measurements within WMO Report No. 47 (see Section 2.1.2). We therefore have two sources of SST sensor information to which we can appeal. Table 1 shows the sensor information for April 1985. The left hand side of the table compares the number of observations identified as coming from particular sensors for each of the main three methods of measurement; bucket, engine room intake (ERI) and hull sensor (combining hull-contact and through-hull sensors) from the two different sources of information. Most of the data show the same method: of those reports for which two types of valid information were available 78% showed agreement, 15% indicated a bucket from the data and ERI from the metadata and 6% the reverse. The

information given the highest priority will be the information from COADS. However only 36% of the ship reports in April 1985 contained the SI flag giving useful information. If we look at this information [column 4] it appears that most of the reports come from buckets. If however we add in the information from WMO Report No 47 [column 5] then the number of reports from buckets increases by only 5% but the number of ERI reports by 27% [column 6]. Using metadata both from within COADS and from the external source we have information about more than two-thirds of the measurement methods rather than just over a third from using the COADS-metadata alone.

Table 2 shows how this information changes during each April for the period under analysis and includes metadata from WMO Report No. 47 for the method of air temperature measurement for which there is no metadata contained within COADS. Figure 9 shows how the availability of metadata for the VOS reports varies throughout the analysis period. Whilst at the beginning of the period less than 30% of the ship reports have associated air temperature metadata (the ship callsign was only included in the ship data transmission code in 1982) this rises to nearly 70% by the early 1990s. The proportion of reports with metadata then reduces to about 60% at the end of the period. The peaks in the amount of metadata may reflect the impact of various projects showing the importance of metadata. For example the amount of SST metadata peaks just after the results of the VSOP-NA (Kent et al. 1993a) experiment were publicised, but the effect seems to have been shortlived. However, the proportion of data with an identifiable measurement method is large enough for analysis to take place.

Parker (1985) and Folland et al. (1993) noted there was a mismatch between the large number of SST reports identified as using buckets from the report flag (SI in COADS) and the much smaller number of ships using buckets according to WMO Report No. 47. This was based on more limited metadata than is available in the study with SST divided only into 'bucket' and 'non-bucket' categories. They noted from examination of the WMO metadata that much of the data for which a method of measurement was unknown must come from engine intakes, but as they did not associate metadata with individual reports were unable to use this information in their analyses. We therefore confirm their expectation that matching metadata from WMO Report No. 47 is particularly valuable in identifying reports from engine intakes. The extra engine intake reports make the numbers of bucket and engine intake reports comparable. Since, as noted by Parker (1985), any error in the identification of the method of measurement will tend to reduce any difference calculated, we can be confident that any biases found will not be overestimated due to errors in the SST flags and WMO Report No. 47 metadata.

3. SST MEASUREMENT METHODS

3.1 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 bucket is usually insulated (largely made of rubber) and may have a double walled construction to hold a sleeve of the seawater around the actual water sample (e.g. UK issue buckets). Examples of some of the types of SST bucket in current use are shown in Figure 10a (reproduced from Kent and Taylor 1991). In addition, Figure 10b shows examples of three types of bucket, from the Meteorological Services of the UK, Holland and Germany. It is hoped to experimentally determine the heat loss from these types of buckets in different environmental conditions. The buckets have very different construction and volumes. 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. Errors can therefore be caused by:

- B1 Cooling of the sample by evaporation from the top surface of the sample or from the walls of the bucket in dry and/or windy conditions (Parker 1985).
- B2 Cooling or warming of the sample by direct heating when the air-sea temperature difference is large (James and Fox 1972).
- B3 Cooling or warming of the sample due to the bucket not having enough time to equilibrate with the SST (Parker 1985). This will occur if it has been particularly cold or hot on deck and the bucket is not kept in the water for a sufficient time. This will be a problem when the solar radiation is strong (if the bucket has not been in the shade), or in very cold conditions. The effect will be larger when conditions do not allow the bucket to be properly immersed in the water (from a large fast moving ship or in strong winds).
- B4 Direct solar heating of the sample or thermometer in sunny conditions (Walden 1966).
- B5 Cooling of the thermometer by evaporation if the thermometer is removed from the sample to read (Walden 1966). This will cause an error if the wet bulb depression is large. It may

occur more often at lower SSTs as it can be difficult to read the bottom of the scale while the thermometer is immersed in the sample.

- B6 For buckets with a fixed integral thermometer (e.g. German issue buckets) there may not be enough water in the bucket to cover the thermometer bulb. The SST then reported will be that of the air or wet bulb temperature. Again, this will be more of a problem on large fast moving ship or in strong winds when deploying the bucket is difficult.
- B7 Precipitation may affect bucket measurements of SST but the extent to which this occurs will depend on the water temperature, the amount, type and temperature of the precipitation, water stability and mixing (James and Fox 1972). It is difficult to separate the effects of precipitation itself from the conditions (higher winds etc.) that normally accompany it.
- B8 Whilst not actually an error, differences between bucket and other measurement techniques might occur due to 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.
- B9 It is assumed that any errors due to miscalibrated thermometers will contribute to the random error in the measurement.

3.2 Engine Intake Measurements of SST

The engine intake measurement of SST is a report of the temperature of the seawater used to cool the engines. It is therefore likely to be made near to the ships' engines. For a good determination of the SST, a measurement close to the seawater inlet is preferable. Seawater is pumped into the ship at a depth between 1 and 9 metres below the waterline (Kent and Taylor 1991) and a remote readout of the temperature on the bridge or in the engine room is recorded. Deep measurement is required as the inlet must be below the waterline whatever the ships' loading. It is not a dedicated measurement of the SST. Engine room intake SST reports show reporting preferences of multiples of whole or half degrees.

- E1 The engine intake SST could be warmer than the bucket sampled SST if it is made a significant distance inboard or close to the ships' engines (Saur 1963, Walden 1966, Kent

et al. 1993a). The error may be larger for larger ships (James and Fox 1972, Kent et al. 1993a) although the water flow rate and depth of sampling may increase on larger ships, which would reduce any heating effect.

- E2 It is expected that the water sampled by the engine intake will be at greater depth than that sampled by the bucket (Walden 1966, Kent and Taylor 1991). This could lead the engine intake SST to be cooler than the bucket SST if vertical temperature gradients occur in the near-surface waters. This real oceanographic effect would only be significant in low wind speed and high solar radiation conditions.
- E3 It is usually hard to read the engine room intake thermometer remote-readout to better than a degree. Graduations can be up to 5°C and parallax errors can be large (Crawford 1969).
- E4 The type of thermometer used for the intake measurement can be significant. James and Fox (1972) found that the relatively warm bias of 0.3°C in the engine intake reports reduced to 0.1°C when only observations using thermistors or precision thermometers were used.
- E5 It is assumed that errors due to miscalibrated sensors will contribute to the random error in the measurement.

Saur (1963) noted that, in addition to any increase in engine intake temperature close to the engine room, there might be other problems with engine intake temperatures. In this study of US Navy ships the engine intake thermometer was a mercury thermometer protected by a metal sleeve inserted into a well in the engine intake pipe. It is not clear to what extent this system was employed on VOS ships and whether mercury thermometers were used in the period 1980 to 1997 studied here. Saur found examples of incrustation 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 but it is not clear how relevant these problems would be to more recent engine intake SSTs. An additional problem in the earlier period studied by Saur (1963) was that, due to a lack of remote readouts, there could often be a time lag in the recorded SST and a resulting mispositioning of the SST report. This could be detected in the dataset in regions of strong SST gradient. We have no information on whether the engine intake measurements we have analysed use remote readouts or whether they use mercury thermometers in the intake pipe.

Compared with SST measurements from Ocean Weather Station (OWS) P, Tabata (1978) found a warm bias in VOS data overall. He ascribed this, based on the literature, to a warm bias in engine intakes, but no identification of measurement method was attempted. Emery et al. (2001) also ascribe a warm bias in VOS SST (relative to drifting buoy measurements) to engine room intake warming again without examining the method of measurement used by the ships.

3.3 Hull Sensor Measurements of SST

Hull sensors are a dedicated SST sensor and measure the SST through the hull or through a hole in the ships 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 as installation costs can be prohibitive unless the sensor and cabling is 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 cut down on the cost of cabling for hull sensor systems in the future (Yelland and Pascal 2001).

- H1 Hull sensor SSTs are expected to be of higher quality than the other methods of measurement as this is a dedicated sensor (Kent et al. 1993a).
- H2 Hull sensor measurements will be at greater depth than the bucket measurements, comparable to those for engine intake inlets (Kent and Taylor 1991).
- H3 It is assumed that errors due to miscalibrated thermometers will contribute to the random error in the measurement. However for any particular ship this will be a systematic bias which may be important as the platinum resistance thermometers commonly used in this type of sensor are prone to drift with time (Peter Taylor, pers. comm.).

3.4 The regional distribution of SST measurement type

As different measurement types are preferred by different recruiting countries there are regional variations in the number of reports for each method. Figure 11 shows the annual average number of reports per 1° area for each of the three main SST measurement methods for the period January 1986 to December 1995. Only regions where there are three or more reports over the whole period are shaded. Figure 11 shows that bucket measurements are the most common, but that engine intake measurements are also reasonably common. Hull sensor SST

reports are fairly uncommon and make up about 5% of the total number of SST reports for which a method can be identified. The North Atlantic is the best sampled region by all the methods will therefore be used for comparisons between the methods.

4. DATA QUALITY CHARACTERISTICS

4.1 Introduction

The quality of both the SST and air temperature data from merchant ships is variable. In order to analyse the data we need to know their error characteristics and to exclude any data thought to be of poor quality or data whose errors cannot be determined. Typical data quality may depend on the country recruiting the ship, the method used to make the measurement or the environmental conditions at the time of the observation. The effect of some of these factors is discussed in the remainder of this section. Only nighttime data have been used as during the day there are extra problems such as diurnal warm layers and solar contamination of air temperature measurements. Reports that deviate by more than 4.5 standard deviations from the local 10° area monthly mean have been removed from the analysis. Reports within 100 km of the coast have also been excluded. The ship tracking and black listing has been applied but no other quality control has been used.

4.2 Data Quality by Country

An initial examination of data quality was performed using the paired file of co-located reports from ships reporting bucket and engine intake SSTs. The accuracy of a report was assessed using the difference of each data point in the SST bucket/engine intake pair from the monthly mean of the 10° area. This difference was measured in units of the standard deviation of the data in that area in that month. Histograms of these normalised differences were plotted by recruiting country separately for the bucket and for the engine intake report. The country of origin of each report was determined either from the COADS country code where available or estimated from the callsign where the country code was absent. On the basis of these histograms, data from eighteen countries were accepted as being of relatively good quality for both SST and air temperature and present in a large enough quantity in the paired dataset to allow error estimates to be made. Figure 12 shows the histograms from the eighteen countries whose data were accepted for further analysis.

Figure 12a shows SST data from January and Figure 12b for July. Data are plotted in black for the SST report in the pair from the bucket (T_{sb}) and in red for the SST report from the engine intake (T_{se}). Figure 12c shows the January air temperature data and Figure 12d that for July. Again, differentiation is by SST measurement method rather than the method used to measure air temperature itself. Preliminary analysis suggests that the most important factor in determining the quality of the air temperature measurements may be the exposure of the sensor. The exposure varies from country to country depending on the emphasis placed on sensor location by the appropriate National Meteorological Service. It seems that countries issuing buckets to measure SST may be more effective in ensuring the air temperature sensors are well exposed: hence our categorisation. The air temperature taken on a ship reporting bucket SSTs will be denoted T_{ab} and the air temperature from a ship reporting engine intake SST will be denoted T_{ae} .

Figure 12 shows that SST and air temperature data from, for example, France and the Netherlands show relatively small scatter. Data from the USA and Japan have, in contrast, broader distributions indicating larger random errors, but are present in large quantities in the dataset and are therefore still valuable. Particularly interesting are the SST distributions from Canada (Figures 12a and 12b) which show large numbers of relatively cold bucket SST reports. The accompanying distributions of T_{se} are more symmetrical. As Canadian ships may be preferentially reporting from high heat loss regions it is possible that these relatively cold values of T_{sb} may be showing the effect of heat flux on the SST sample.

4.3 Data Quality by Ship Parameters

Random errors are calculated using the semivariogram method (Kent et al. 1999) whereby mean square differences between pairs of ship variables are regressed against ship separation to zero separation in an attempt to remove the spatial component of variability. It should be noted that the error estimates presented in this and the next section may appear inconsistent. For example, most measurements of T_{ae} are made using a mercury thermometer exposed in a screen and the random errors for this type of measurement are about 1.5 °C (Table 3). However, when the data are stratified by ship speed (Table 4) typical random errors for T_{ae} are smaller than this. This can be explained if the random errors we are examining are in fact systematically varying with the independent variable. If this is the case then we may be preferentially selecting air temperature measurements that are more likely to be similar by requiring them to be taken under similar conditions thus reducing the estimated random errors.

Both the engine intake and bucket data are predominantly associated with screen air temperature measurements made using mercury thermometers (Table 3, Figure 13a). Electrical resistance thermometers may be more accurate but there are too few measurements to make a good assessment. There is an indication that ships using resistance thermometers to make air temperature measurements may also make more accurate SST measurements but again the quantity of reports is small. It is clear that T_{sb} reports contain smaller random errors than T_{se} and this difference dominates over that for all other parameters where there are enough data to make a good error assessment. The same is true for air temperature, that is random errors for T_{ab} are almost always smaller than those for T_{ae} .

Table 4 shows similar error estimates to those presented in Table 3 but graduated in ranges of the speed of the ship at the time the observation was taken. The same information is plotted in Figure 13b. T_{sb} and T_{se} both tend to have slightly smaller errors when the ship is moving slowly. The reverse is true for the errors in T_{ab} and T_{ae} . This can be explained by increased ventilation of the air temperature sensor at higher ship speeds. Again, although the effect of ship speed can be seen in SST and air temperature measurements from ships using both buckets and engine intakes, the effect is smaller than the differences between the bucket and engine intake measurements themselves.

Figure 13c and Table 5 show the random error estimates as a function of relative wind speed. The SST estimates from both buckets and engine intakes show little variation with relative wind speed. This suggests that the speed of the ship is a more important factor in taking good quality bucket SST measurements than the speed of the airflow over the ship. As expected the random error estimates for air temperature decrease with increasing relative wind speed (and hence increasing sensor ventilation).

Table 6 and Figure 13d show how the random error estimates vary as a function of observing platform height in metres. Lower observation platforms are associated with more scattered observations. It is possible this is a coastal effect as smaller ships are often concentrated close to coasts where variability is typically greater than in the open ocean. This is more striking for air temperature observations than SST and also more pronounced for ships using buckets to measure the SST. Air temperature measurements taken on observing platforms below 15 metres height have comparable values of σ_{ab} and σ_{ae} . In all other cases T_{ab} contains smaller random errors than T_{ae} (i.e. $\sigma_{ab} < \sigma_{ae}$).

Table 7 and Figure 13e show how the random error estimates vary with latitude in the North Atlantic. T_{sb} reports have the largest random error in low latitudes and their lowest error at about 60°N although the variation is not very large. T_{se} reports have relatively small errors in the tropics and larger errors in mid-latitudes. Again, the engine intake SST error estimates are all larger than those from buckets. For air temperature, error estimates for both T_{ab} and T_{ae} increase with latitude. It is possible that this results from increased temporal variability which is not removed by the semivariogram method.

4.4 Data Quality by Environmental Parameters

Considering SST, random errors in T_{sb} decrease slightly with increasing wind speed (Table 8, Figure 13f). There is little variation in T_{se} accuracy with wind speed although the largest errors are found in the lowest wind speed category. For air temperature the accuracy of T_{ab} initially worsens with increasing wind speed with the most error variability in the 6-8 ms^{-1} bin. Above this range of wind speed the accuracy in T_{ab} improves with increasing wind speed. The accuracy of T_{ae} is again worse than T_{ab} but air temperatures taken when the wind speed is in the range 10 - 20 ms^{-1} have slightly smaller random errors than those taken at wind speeds below 10 ms^{-1} . For air temperature relative wind speed is a more important factor in the ventilation of the instruments, but the number of relative wind speed reports is much more limited as many ship reports do not contain the ships' speed and direction information necessary to make the calculation of relative wind speed.

Table 9 and Figure 13g show how the error estimates vary with total cloud cover at the time of the observation. The SST reports from both sources improve in quality with increasing cloud cover and the air temperature reports worsen with increasing cloud cover. Remembering that these are nighttime reports there is no obvious reason why the accuracy should depend on cloud cover so this result is not understood at the moment.

Table 10 shows how the error estimates are affected by precipitation (calculated from the present weather code). Random errors for all variables are slightly smaller when it is raining. There were not enough data to stratify into different rain rates.

Table 11 and Figure 13h show how the error estimates vary with local sea-ice concentration taken from Cavalieri et al. (1997). The SST error estimates σ_{sb} and σ_{se} decrease with increasing ice concentration, presumably reflecting the low variability of SST in ice-covered regions. For air temperature T_{ab} error estimates are much higher (by a factor of 2) in regions where there is

sea-ice and are comparable with those from T_{ae} in these regions. However the decrease in accuracy is much less marked for T_{ae} as the error estimate for air temperature in the ice-free region is already 50% greater than T_{ab} .

5. ANALYSIS METHODS

5.1 Effect of errors on analysis

It is well known that VOS weather reports contain errors, both systematic (e.g. James and Fox 1972, Kent et al. 1993a,b, Kent and Taylor 1996) and random (e.g. Kent et al. 1999) some of which are described in Section 3. This has two implications for the analysis of biases. Firstly, a large volume of data needs to be analysed to reduce the random component of the error since for small numbers of observations the random error is likely to be much larger than any bias. Walden (1966) noted that "The difference in measurement of the water temperatures by the two methods is completely concealed by many ... large errors". This is still the case. Secondly, care must be taken to ensure that random or systematic errors either in the variable being analysed, or in other variables, do not contaminate the analysis. This is particularly important as ships recruited by a particular country are likely to have been issued with the same instruments and instructions on how to take the observations. Thus it may be that ships using buckets to measure the SST may, in the main, use thermometers housed in screens to measure the air temperature and take visual observations of the wind speed. This becomes particularly important when biases are being analysed as a function of variables such as latent or sensible heat flux which are parameterised from a combination of basic variables.

James and Fox (1972) found that the assignment of a significance level to any results is difficult. They resorted to non-parametric tests of significance as the data they were testing were from non-standard distributions. They also had problems with non-independence of measurements, because their measurements came from a small subset of ships. This should not be a problem in the present study, at least in relatively well-sampled regions, as measurements from most of the VOS fleet are being used. Saur (1963) used improved design buckets (actually a thermometer and plastic thermometer holder with a metal shield) and assumed that the bucket measurements of SST were perfect, an approach we obviously do not wish to take in this study.

5.2 Comparison of climatological fields

Previous analyses (e.g. Parker 1985, Folland et al. 1993, Quan et al. 1999) have compared SST climatologies derived separately for each measurement method. Further details are given in Section 6.5.2.

5.3 Comparison with model fields

The VOS Special Observing Project - North Atlantic (VSOP-NA, Kent and Taylor 1991, Kent et al. 1991, 1993a,b) analysed a subset of VOS measurements merged with the output of a numerical model. It was hoped that comparison of observation-model differences for different measurement types would be independent of any biases in the model. The dataset was small (two years data from 46 ships) and required a large effort to assemble. Extra metadata was collected for each ship including photographs of measurement sites and instrumentation, additional parameters were recorded with each observation by the ships officers and the special VSOP-NA logbook data had to be keyed-in. The ship reports and model output were merged and the metadata collated and studied. Results are reviewed in Section 6.5.3.

The success of the VSOP-NA study has led to the development of a new WMO initiative, the VOS Climate Project¹, to produce operationally a VSOP-NA-type dataset (WMO 2000). The project is just getting underway and so far only a small volume of data has been collected but should in the future prove a valuable resource.

5.4 Observational pairs

5.4.1 Advantages of Using Paired Data

Requiring that we only analyse data pairs that are nearly coincident in space and time has several advantages. We remove any uncertainty in the analysis due to the comparison of different regions or time periods covered by climatologies. No smoothing of the data is required, again removing uncertainty from the analysis. Another benefit is that the paired data is easily used to generate error estimates for the dataset, an important part of any analysis of bias in the

¹ <http://www.ncdc.noaa.gov/VOSclim.html>

data. It is felt that these advantages outweigh any disadvantages due to the loss of data inherent in requiring all observations to be paired.

5.4.2 The Observational Pairs Dataset

To reduce any uncertainties in sampling, and to help ensure that the measurements being compared were obtained in similar environmental conditions, this study uses observation pairs taken within an hour and 50 km of each other. As most VOS report on the synoptic hours (0, 6, 12, 18 GMT) and a few every three hours this means that most of the pairs will be nominally simultaneous. There will obviously be some temporal variability due to the observation not being made on the hour. These errors due to temporal variability are expected to contribute to the random component of the error and hence be reduced by averaging. Errors due to the spatial separation of the measurement pair may also average out in many regions. Unfortunately regions where this may not occur are also the regions where strong SST spatial gradients lead to large surface fluxes which are expected to impact on the bucket SST measurements (see Section 3.1). Care must therefore be taken in interpreting results in regions of strong SST gradient. These problems are however expected to be smaller than those that would be encountered if SST climatologies were compared for the different methods.

Various versions of the paired dataset are used in this study. For all the pairs had to be nominally simultaneous (i.e. at the same reporting hour) and had to have passed the quality control and track checking described in Section 2. For the analyses presented in Section 6.1 the pairs had to be within 50 km separation and from the North Atlantic. Each report in the pair had to use a different method of measurement, in this case bucket or engine intake. Data for both night and day were used and no selection by country was made. For the random error estimates shown in Section 4 data up to 300 km separation were used for ships from the countries which reported good quality data (see Section 4.2). For these random error estimates the paired dataset were common-type (i.e. one dataset containing paired bucket reports and another dataset containing paired engine intake reports). For the analysis of individual data pairs described in Section 6.2 a more rigorous selection was performed on the three datasets used (one mixed-type and two common-type). Only nighttime data from the North Atlantic between 20°N and 50°N were used. Data were restricted to reports from the countries which reported good quality data. Wind speeds were constrained to be between 4 and 12 ms⁻¹ and the wind speeds in a particular pair to be within 5 ms⁻¹ of each other. All four combinations of absolute air-sea temperature difference in the report pair were constrained to be less than 15 °C. For the analysis data pairs had to be within 50 km separation and for the error estimates upto 300 km separation.

5.4.3 Theory

The method described in this section was developed by Dr Alexey Kaplan, Climate Modeling Group, Lamont-Doherty Earth Observatory of Columbia University. It handles the regression of variables which have significant and correlated errors. The simplest case is considered here neglecting the influence of wind speed, evaporation and solar radiation and focussing on the air - sea temperature difference. In our example the variables to be regressed are [bucket - engine intake SST] against [air temperature - engine intake SST]. These variables are obviously correlated through the engine intake SST. It is unfortunate that this is also the variable that contains the largest random error.

Let T be the true SST, T_{sb} the bucket SST, T_{se} the engine intake SST and T_a the air temperature. Consider only nighttime data. Assume T_{sb} is in error by a fraction (α) of the air-sea temperature difference and T_{se} is on average in error by a constant amount (β).

$$T_{sb} = T + \alpha(T_a - T) \quad (1)$$

$$T_{se} = T + \beta \quad (2)$$

(Note: we expect both α and β to be positive if the engine intake reports are relatively warm. α must be positive and should be of order a tenth to be physically realistic, but β can physically take either sign). Substituting for the true SST, T :

$$T_{sb} = T_{se} - \beta + \alpha(T_a - T_{se} + \beta) \quad (3)$$

$$T_{sb} - T_{se} = \alpha(T_a - T_{se}) - \beta(1 - \alpha) \quad (4)$$

$$(T_{sb} - T_{se}) = \alpha(T_a - T_{se}) + \gamma \quad (5)$$

where the new constant, $\gamma = \beta(\alpha - 1)$. Equation (5) can be written in a simple linear form, equation (6):

$$y = \alpha x + \gamma + \varepsilon \quad (6)$$

where y is $T_{sb} - T_{se}$, x is $T_a - T_{se}$ and ε represents a combination of the Gaussian random errors in all variables.

We wish to work with differences (x' and y') from the mean values (\bar{x} and \bar{y}). We can write an equation for the mean values (7):

$$\bar{y} = \alpha\bar{x} + \gamma \quad (7)$$

and one for the differences from the mean (8):

$$y' = \alpha x' + \varepsilon \quad \text{or} \quad (y - \bar{y}) = \alpha(x - \bar{x}) + \varepsilon \quad (8)$$

From now on we will work with the 'primed' difference values x' and y' . Equation 8 can be represented by an equation of the form $Y=\alpha X$ (+ error) where X and Y are matrices (of size $n \times 1$, where n is the number of observations), α is a scalar, y' is $(T_{sb}-T_{se})'$ and x' is $(T_a-T_{se})'$:

$$\begin{pmatrix} y'_1 \\ y'_2 \\ y'_3 \\ \vdots \\ y'_n \end{pmatrix} = \alpha \begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \\ \vdots \\ x'_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (9)$$

At this stage we only need to determine α , once we know α we can easily calculate γ , and hence β , from the equation for the mean quantities (7).

The regression equation $y'=\alpha x'$ has random errors in both x' and y' :

$$\varepsilon_x = \varepsilon_a - \varepsilon_e \quad (10)$$

$$\varepsilon_y = \varepsilon_b - \varepsilon_e \quad (11)$$

where ε_e is the error in T_{se} , ε_b the error in T_{sb} and ε_a the error in T_a . The errors in x' and y' , ε_x and ε_y are correlated (in this case through ε_e) and may be of a comparable size (although we cannot assume they are equal in size). We are also assuming that ε_a , ε_b and ε_e are uncorrelated. To perform a regression on the data we need to account for the different sizes of the errors in x' and y' and also to account for correlations between the errors in x' and y' . We therefore aim to transform the data into a working space in which the random errors in both variables are of unit size and uncorrelated with each other. An orthogonal regression can be performed in this working space and the results of this regression converted back to give the true regression equation for the real system. We start by defining a data matrix Z containing the anomaly data x' and y' :

$$Z = \begin{pmatrix} x'_1 & y'_1 \\ x'_2 & y'_2 \\ \vdots & \vdots \\ x'_n & y'_n \end{pmatrix}^T = \begin{pmatrix} x'_1 & x'_2 & \dots & x'_n \\ y'_1 & y'_2 & \dots & y'_n \end{pmatrix} \quad (12)$$

Then equation (8) can be rewritten as:

$$(-\alpha, 1)Z = (\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n) \quad (13)$$

We need to transform the data matrix Z so that the errors in each column of Z are equal in size and independent. The first stage is to define an error correlation matrix C :

$$C = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix} \begin{bmatrix} \varepsilon_x & \varepsilon_y \end{bmatrix} = \begin{bmatrix} \langle \varepsilon_x \varepsilon_x \rangle & \langle \varepsilon_x \varepsilon_y \rangle \\ \langle \varepsilon_y \varepsilon_x \rangle & \langle \varepsilon_y \varepsilon_y \rangle \end{bmatrix} \quad (14)$$

where:

$$\begin{aligned} \langle \varepsilon_x \varepsilon_x \rangle &= \sigma_a^2 + \sigma_{se}^2 \\ \langle \varepsilon_x \varepsilon_y \rangle &= \langle \varepsilon_y \varepsilon_x \rangle = \sigma_{se}^2 \\ \langle \varepsilon_y \varepsilon_y \rangle &= \sigma_{sb}^2 + \sigma_{se}^2 \end{aligned} \quad (15)$$

$$\text{we then further define } D \text{ such that } D^T D = C^{-1}. \quad (16)$$

See Appendix for the method of calculating D from C . The data matrix (Z) is converted to a matrix, Z_1 :

$$Z_1 = DZ \quad (17)$$

the errors in Z_1 are equal and uncorrelated (see Appendix for proof).

We can now perform an orthogonal regression on Z_1 :

$$Z_1 Z_1^T = E \Lambda E^T = \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \end{pmatrix} \quad (18)$$

The regression line is by definition collinear to the first empirical orthogonal function (EOF) $\begin{pmatrix} e_{11} \\ e_{21} \end{pmatrix}$. As all EOFs are orthogonal we can determine the regression coefficients directly using the fact that the regression line must therefore be orthogonal to the second EOF. This means that the dot-product of the regression parameters with the transformed data must be zero:

$$\begin{pmatrix} e_{12} & e_{22} \end{pmatrix} \cdot Z_1 = 0 \quad \text{or} \quad \underbrace{\begin{bmatrix} (e_{12} & e_{22})D \end{bmatrix}}_{h=(h_1, h_2)} \cdot Z = 0 \quad (19)$$

we can then write this equation in the form:

$$h_1x + h_2y = 0 \quad (20)$$

$$y = -\frac{h_1}{h_2}x \quad (21)$$

and then recover the regression parameters α and γ :

$$\alpha = -h_1/h_2 \quad (22)$$

$$\gamma = \bar{y} - \alpha\bar{x} \quad (23)$$

and hence β :

$$\beta = \gamma / (\alpha-1) \quad (24)$$

This computation is only valid for a sample where α and β do not vary significantly within the sample. In our analysis $\alpha < 0$ is unphysical and $\alpha \sim 1$ would mean that the bucket measured SST and the air temperature were always equal within the error ranges of the variables.

5.4.4 Application of Analysis Method

5.4.4.1 ERROR ESTIMATES

The random error estimates required are given in Equation 15 and are σ_{se} (the random error for the engine intake SST reports, T_{se}), σ_{sb} (the random error for the bucket SST reports, T_{sb}) and σ_a (the air temperature random error, T_a). They were derived by applying the semivariogram method (Section 4.3 and Kent et al. 1999) to three different paired datasets to which the same quality control described in Sections 2.2 and 2.3 had been applied.

Three datasets are required to obtain all the information necessary to analyse the data. The first is the file described in Section 5.4.2 containing pairs of quality controlled bucket and engine intake-derived reports. Of the other files, one contained bucket-bucket pairs (to derive σ_{sb} and σ_a) and the other engine intake-engine intake pairs (to derive σ_{se}). The air temperature error was derived for the air temperatures on ships that used buckets (i.e. $T_a \equiv T_{ab}$) despite the SST in the air-sea temperature difference (x in Equation 6) coming from the engine intake report. This was because T_{ab} was better correlated with T_{se} than was T_{ae} despite the extra spatial variability due to the ship separation. This emphasises the much better quality of T_{ab} over T_{ae} . Figure 12c and 12d suggest little difference in the random errors between T_{ab} and T_{ae} for a particular country, suggesting that the best indicator of air temperature report quality may be the recruiting country (see Section 4.2).

The use of separate files to calculate the random errors and biases is not ideal. However the analysis of a restricted area and of reports from selected recruiting countries may have allowed a homogeneous overall data base. A supporting estimate of $\sigma_{sb}^2 + \sigma_{se}^2$ was obtained from the semivariogram intercept of the bucket-engine intake file and was similar to that based on the bucket-bucket and engine intake-engine intake files, to within the estimates of the uncertainty in each quantity. Table 13 contains the intercept values for the semivariogram estimates for each month for the data between 1980 and 1997 along with estimates of their uncertainty. These values are twice the error variance for the single-type data and the sum of the error variances for the mixed-type data.

5.4.4.2 PROCEDURE

Using the error estimates described in Section 5.4.4.1 the data ($x = T_a - T_{se}$ and $y = T_{sb} - T_{se}$) are transformed into the working parameter space where errors in both variables are equal and uncorrelated as described in Section 5.4.3. Error estimates for the resulting regression lines were made in the following way. Firstly the uncertainty in the regression estimate is calculated using the software pack ODRPACK (Boggs et al. 1992)². The uncertainties are the square root of the diagonal elements of the covariance matrix for the regression model parameters. 95% confidence limits are applied using the Student's t value for the appropriate number of degrees of freedom for a two-tailed distribution. This is an approximation which should work well for the simple case of orthogonal linear regression solved here. The regression lines representing the

² <http://www.boulder.nist.gov/mcsd/Staff/JRogers/odrpck.html>

limits of the 95% confidence region are then calculated. These lines are then transformed back into the real space to give an error range for the final regression estimate. The whole process is then repeated adjusting the correlation matrix to account for the estimated uncertainties in the error estimates (those given in Table 13). The maximum range of these gradients and uncertainties for both the gradient and intercept is then taken as the error range for the regression estimates which are presented in the next section. The error range quoted therefore accounts both for uncertainties in the regression and uncertainties in the correlation matrix used to transform the data.

Figure 14 shows the seasonal variation of the estimated errors in SST and air temperature as shown in Table 13. It is clear that the random errors in both air temperature and SST reports from ships using buckets to measure the SST are usually much smaller than those from ships using engine intakes. For these nighttime air temperature error estimates there is a clear seasonal cycle for both methods of measurement with maximum random errors in May and minimum random errors in late summer and autumn.

6. RESULTS

6.1 Comparison of fields generated from coincident pairs of reports

6.1.1 Differences between bucket and engine intake SST fields

Figure 15 is a bivariate plot, generated from the paired dataset described in Section 5.4.2, showing the SST difference between bucket and engine intake measurements as a function of wind speed and solar radiation. The bucket SSTs are relatively cool at low solar radiation. At low solar radiation the bucket SSTs become increasingly colder than the engine intake SSTs as the wind speed increases. This is consistent with heat loss from the bucket increasing with wind speed. The exception is at wind speeds greater than about 25 ms^{-1} when the amount of cooling seems to decrease. The very high wind speed region of the plot shows the bucket SST relatively warm compared to the engine intake SST. This is in contrast to a cold bias in the bucket SST in the moderate to high wind speed region of the plot. This is likely to be due to the difficulty of making a bucket measurement of SST at high wind speed. It is hard to immerse the bucket in the water at all, and it is unlikely that the bucket can be completely filled with water. It will therefore be very hard to leave the bucket in the water long enough to equilibrate with the sea temperature. This becomes more obvious under high solar radiation conditions where the bucket SSTs are much warmer than the engine intake SSTs. This is consistent with a bucket that has

been heated on the deck by the sun and then not immersed properly in the sea. The bucket SST can be very biased in these conditions, although there are not many pairs in this region. In high solar and low wind speed conditions we can probably see a combination of the bucket sampling a shallower warmer layer than the engine intake, along with some warming of the water sample (or of the bucket) by the sun. In the centre of the plot, where both cooling fluxes and solar radiation are important, there is a balance between the competing effects of warming by the sun and cooling by the heat fluxes. The picture is therefore more complicated than a simple warm bias in the engine intake SST.

Smoothed fields on a two-degree latitude-longitude grid were generated using a 5-degree area Gaussian smoother separately for each report type from the same data pairs used above. To remove data in regions where sampling is poor it was required that the mean latitude and longitude for the different methods within a 2-degree area box were the same to within 0.1-degree. This requirement ensured that large extrapolations from the well-sampled region were not made.

Figure 16 shows the difference between the bucket and engine intake SST fields for cases where the shortwave radiation calculated for both reports in the pair was less than 1 Wm^{-2} . The bucket SST is colder than the engine intake SST over most of the North Atlantic except for a region to the north of the UK (top panel). The centre panels show the SST field for each method extracted along 45°N and 35°N . The lower panel shows the difference between bucket and engine intake SSTs at these two latitudes. In these North Atlantic midlatitudes the bucket SSTs are always colder on average than the engine intake SSTs.

Figure 17 is of the same form as Figure 16 but the fields are constructed of pairs for which the solar radiation calculated for each report was greater than 500 Wm^{-2} . When the solar radiation is high, the bucket SSTs are warmer than the engine intake SSTs over most of the North Atlantic. However heat loss due to evaporation may still be offsetting any solar warming effect.

Figure 18 shows SST differences generated for report pairs where the wind speed is less than 3 ms^{-1} for both reports and the solar radiation is greater than 300 Wm^{-2} . Bucket SSTs again are relatively warm, more so than in Figure 17 where the solar flux was greater than 500 Wm^{-2} but all wind speeds are included. Figure 18 probably reflects a combination of the buckets sampling a shallow warm layer and with reduced cooling of the sample due to lower wind speed.

Figure 19 shows SST differences generated for cases where the wind speed is greater than 6 ms^{-1} and the solar radiation less than 1 Wm^{-2} . The results are similar to those for the full range of wind speed data shown in Figure 16.

6.1.2 Air-sea temperature differences between bucket and engine intake SST fields

Figure 20 shows the air-sea temperature difference for reports made with the two different methods of SST measurement under conditions where the solar radiation for both reports is less than 1 Wm^{-2} . The air-sea temperature fields for the bucket SST (top panel) and engine intake SST (centre panel) reports show, as expected, that the SST is typically $1\text{-}2 \text{ }^\circ\text{C}$ warmer than the air temperature. The bucket SST field (top panel) shows slightly smaller air-sea temperature differences which is consistent with the expectation that the buckets suffer from cooling (or warming) which will act to reduce the reported air-sea temperature difference. The lower panel in Figure 20 compares the SST fields for the two methods along 35°N . At this latitude, the mean air-sea temperature difference from the bucket reports is always smaller than that from the engine intake reports.

The situation is very different when the solar radiation is greater than 300 Wm^{-2} , Figure 21. It is well known that marine air temperatures are contaminated by solar radiation so it is not surprising that the air-sea temperature differences are now more positive. This is mainly due to the warming of the reported air temperature. What is unexpected is the large difference between the effect of solar radiation on the air temperatures from the ships using different methods of SST measurement. The air-sea temperature difference from ships using engine intakes is positive over almost all of the North Atlantic (centre panel), that from the ships using buckets is mostly negative (upper panel), but with significant areas where the difference is positive. Kent et al. (1993b) examined the air temperatures from the VSOP-NA ships and concluded that solar radiation affected the air temperature measurements from all ships in a similar way, regardless of air temperature sensor type. The exception to this was for those ships which had air temperature sensors which were very badly exposed, which were few in the VSOP-NA. Thus significant differences in the solar radiation effect were not expected. The differences in air-sea temperature are much larger than the differences in SST between the methods. This confirms the results described in Sections 4.3 and 4.4 showing that the air temperature errors from those ships that use engine intakes to measure the SST are consistently of poorer quality (i.e. more variable) than those on ships which use buckets. Table 3 shows that the difference in quality of the air temperature reports does not result from a difference in air temperature measurement method as

for the most common types (mercury thermometer exposed in a screen or sling) the error is greater for the ships that report engine intake SST.

6.1.3 Air temperature differences between bucket and engine intake report fields

Figure 22 shows a bivariate plot, similar to that in Figure 15, but for the air temperature difference between reports that use buckets and those that use engine intakes to measure the SST. As expected from Figure 21 this shows that the air temperature difference between buckets and engine intakes is greatest at low wind speeds and in conditions of high solar radiation. A possible conclusion is that the air temperature sensors on the ships using engine intakes tend to be more poorly exposed than those on ships using buckets. Another possibility is that the ships that use engine intakes are larger and have a bigger 'heat island' effect in sunny conditions. It should be noted that the solar contamination effect shown in Figure 22 is in addition to any solar contamination that affects the two subsets of reports in a similar way.

Figure 23 shows the difference between the air temperatures from bucket and engine intake reports when the solar radiation is less than 1 Wm^{-2} . Differences are generally small, as would be expected. Figure 24 shows the difference between the air temperatures from bucket and engine intake reports when the solar radiation is greater than 500 Wm^{-2} . Under these conditions the engine intake report air temperatures are consistently warmer than the bucket report air temperatures which is consistent with the results shown in Figure 22.

6.2 Statistical Analysis of Individual Observational Pairs

Figure 25 shows the results of the analysis outlined in Section 5.4.3 on the data as described in Section 5.4.2 month by month for data between 1980 and 1997. In each month there is a significant component of variability resulting from the correlation of errors through the engine intake SST. This component falls along the dotted line in the upper figure of each pair. It is any variability which does not fall along this line that we are interested in. Scatter away from the dotted line is due to random errors in the bucket SST and air temperature (which have been accounted for in the analysis) and any systematic errors (which we would like to identify). The lower panel in each pair shows the same data as in the upper panel but following the transformation to the working space in which the errors in the x and y variables are equal in size and uncorrelated. In this working space any points lying close to the dotted line in the real variable co-ordinates are attributed to correlations in the errors between x and y and down-weighted. Figure 26 may help in understanding the weighting and rotation of data that has taken

place. Figure 26 shows the same data as in the January panel but separated by quadrant defined in the real co-ordinate system. Each panel shows the same data points in the real and transformed co-ordinates and makes the relationship between the two co-ordinate systems more obvious. See the Figure caption for further details. From this Figure the suppression of data containing correlated errors, and the concentration by rotation of data where the errors are not correlated, is clear.

The regression parameters and their error estimates are given by month in Table 14. Excluding June, July and August the gradient (α) is approximately 0.2 ± 0.1 and the intercept (β) close to zero within the estimated error range. In the summer months the gradient estimates are poorly defined and α may be negative which is unphysical. Inspection of the panels of Figure 25 for June, July and August shows that in these months, in the North Atlantic latitudes chosen, the range of air - sea temperature differences is smaller than in the other months and the error range of the resulting regression large.

6.3 Referencing Individual Observations to Climatology

We referenced the individual observations in the paired dataset to 5 day-average SST fields to reduce the effects of spatial variability. However this did not reduce the estimates of σ_b or σ_c because the reference fields were noisy. In addition the 5-day fields were less complete than the blended SSTs being based on MOMDB only, so 20% of the January SSTs between 1980 and 1997 could not be used. We plan to repeat this using a smoother climatological field to reduce the effects of spatial variability.

6.4 Biases in Air Temperature Data

6.4.1 The Impact of Deck Height Metadata on Climatological Air Temperature Fields

The use of metadata with individual observations allows maps of corrections for the height of the air temperature sensor to be plotted for the first time. Figure 27 shows the annual 2° area average of observing platform height for each year of 1980 to 1997. The regional differences result from the different types of ship which are common in different regions. In the Arctic region small fishing vessels are most common and the observing heights are very low, typically around 15 metres. In mid-ocean regions bigger vessels dominate and the observing heights are much greater, typically around 30 metres. In addition the height of the observing platform is increasing with time in most regions. This is shown in Figure 28 which shows the linear trend

with time of the data shown in Figure 27. Figure 29 shows time series of the area average observation height for several ocean regions.

Prior to 1980 and after 1997 assumptions have to be made about the trends in air temperature observing platform height. In addition, if the results for the analysis period are to be applied to an already existing air temperature climatology the detailed results must be simplified for application to area averaged data. Figure 30 shows the results of such an application by David Parker who made assumptions as follows: All deck heights were assumed to be 6m up to 1890, then a linear increase to 15m in 1930, then 15m through 1970. For 1971-82, heights at each grid point were interpolated linearly between 15m in 1970 and 5-year annual average field of heights based Figure 27 for 1980-84 in 1982. For 1982-95, the 5-year annual average fields of heights were used. For 1996-2002, the 5-year annual average field of deck heights for 1993-7, already used for 1995, was incremented uniformly by 0.14m yr^{-1} in accord with global-average trends in deck heights since the early 1970s derived from WMO Report No. 47. The impact of the improved height information is shown in Figure 31. Figure 31a is taken from IPCC (2001) and compares global SST, night time marine air temperature (NMAT) and land air temperature. Figure 31b differs from Figure 31a only in the application of the new observation height corrections to NMAT. The global NMAT anomalies in recent years now slightly exceed those of SST, as might be expected in view of the higher land air temperature anomalies. Agreement with earlier SST anomalies is also generally improved (because of the higher observation height assumed in the reference period).

6.4.2 The Impact of Solar Radiative Heating on VOS Air Temperatures

Figure 32a shows a correction developed by David Berry (SOC) for the solar radiative heating of VOS air temperature sensors based on a reanalysis of the VSOP-NA dataset. Kent et al. (1993a) had suggested a correction which depended on the incident solar radiation and the relative wind speed. This was found to be inadequate to properly correct the data for the heat island effect of the ship which persists after the solar radiation has decreased (Aiguo Dai, pers. comm.). A revised correction is therefore being developed which allows for heating occurring prior to the observation time. This is work in progress but it suggests that effect of solar heating persists several hours after sunset, especially if the relative wind speed is low (Figure 32b). This could be important as nighttime marine air temperature climatologies aim to include only data uncontaminated by solar radiation effects (Parker et al. 1995).

6.4.3 Example of VOS Air Temperature Bias

Figure 33 shows a comparison by Alex Sen Gupta (SOC) of the air temperatures reported by the RRS James Clark Ross acting as a VOS and those recorded by research quality sensors well exposed on the ship's foremast. The UK Met Office screens are located on the Bridge Top but towards the back, not the best position but also not particularly poorly exposed. The 6-hourly VOS reports are much warmer than the research quality data and the influence of solar radiation (also plotted) on the difference between the two air temperature measurements is clear. It may be possible to analyse the VOS reports from other research ships using the research ship data stored in the World Ocean Circulation Experiment Surface Meteorology Data Assembly Center at Florida State University (Shaun Smith pers. comm.). Whilst the research ships are not typical of the VOS as a whole, they often carry the same instruments as the rest of the VOS and, as demonstrated here, can contain biases of a comparable size.

6.5 Comparison of SST Results with Previous Studies

6.5.1 Comparative Studies (Saur 1963, Walden 1966, James and Fox 1972, and Tabata 1978)

Early studies comparing bucket and engine intake SST reports (e.g. Roll 1951, Kirk and Gordon 1952, Åmot 1954 and Franceschini 1955) are not considered here as they either specifically refer to observations made using canvas buckets or make no mention of the type of buckets used. In this early period the use of uninsulated buckets was common and the results of these studies are therefore not relevant to the period studied here.

Saur (1963) concluded that engine intake SSTs were of variable quality and were not consistent within the limits one might expect from vertical temperature gradients. He found that the differences between the engine intake and reference bucket data varied not only from ship to ship but also between trips for a given ship and sometimes within a trip. For the twelve US Navy ships in the study the engine intake SST was biased warm by 0.7 ± 0.3 °C compared with the improved design buckets used as a comparison standard. The standard deviation of differences with the buckets was 0.9 °C. A contributing factor to the noise in the engine intake SST was the use of mercury thermometers with no remote readout which he suggested should be phased out. The result that engine intake SSTs are noisy and warmer than bucket SSTs is in agreement with the present study. However, it is not clear how the buckets used in the Saur (1963) study would compare to those used by the recent VOS. As the Saur (1963) study

assumed that the buckets were accurate no analysis was made of possible heat loss by the buckets (which would be hard to interpret for the VOS in general due to the unusual bucket design).

Walden (1966) concluded from a study of German merchant ships mainly in the period after 1960 that engine intakes were too warm. The difference increased with wind speed which was not expected as the ocean would become more mixed at higher wind speeds and he thought the two methods should therefore agree better. He was therefore probably seeing the effect of cooling on the buckets, which were probably similar to the type of German bucket used today. He also noted that increased solar radiation causes the buckets to become relatively warm. These results agree with the present study if a cold bias is attributed to the bucket data rather than a warm bias to the engine intake data.

Tauber (1969) reported that the Crawford-type bucket (Crawford 1969) cooled by 0.2°C in 3 minutes when the air-sea temperature difference was 3-4°C. The wind speed at the time is unknown but this does suggest that even insulated buckets of complex design can lose a measurable amount of heat, in agreement with the findings of the present study.

The global study of James and Fox (1972) compared simultaneous measurements of bucket and engine intake SST measured on the same ship. They found that on average the engine intake reports were 0.3°C warmer than the bucket reports which was significant at the 95% level. They state that "the 0.3°C difference between intake and bucket temperature is generally ascribed to engine room heating, but may be partially due to cooling of the bucket sample". At high latitudes they find no significant difference between the two methods of measurement which they ascribe to colder waters reducing the engine room intake temperatures and the positive air-sea temperature difference acting to warm the bucket sample. They found larger ships to have larger differences between the two methods, and the difference increased with both increasing intake depth and with distance inboard of the intake thermometer. This is surprising as one would expect the engine intake SSTs taken at greater depth would be cooler and therefore show smaller differences, but perhaps this indicates that the size of the ship and the distance inboard are more important and these effects dominate over the effect of the vertical thermal gradient in the ocean. The unimportance of the vertical thermal gradient is also suggested by Figure 15 in the present study which shows that the region where the buckets are relatively warm is confined to the low wind, and moderate to high incident solar radiation. James and Fox (1972) found little variation of the difference with increasing wind speed but differences were significantly larger in the highest wind speed category. They quote an unpublished document by Crawford who suggested using bucket measurements as a standard except where the environmental conditions prevail that

most strongly influence the bucket sample. James and Fox (1972) recommend that engine intake SST reports should be decreased as a function of ship characteristics whilst bucket readings would generally be increased when the winds are strong, atmospheric stability extreme or precipitation occurring. However they say that developing a correction procedure is not feasible and it would be better to stress the importance of improved intake thermometers closer to the hull and that observers should take extra care when measuring bucket SSTs in conditions of strong wind, large air - sea temperature differences and heavy precipitation. They also note that applying a generalised correction negates the efforts of observers aware of the issues and taking care making their measurements.

Tabata (1978) compared instantaneous merchant ship observations in the northeast Pacific Ocean with 3.5 day averages from OWS P and NOAA buoys. Comparisons were made where the merchant ship was within the same 2° square as the OWS or buoy. No consideration was made of the type of observation method but the difference is attributed to a warm bias in the merchant ship engine room intake SSTs. Interestingly in the region of the northwest Pacific used in the Tabata (1978) study, Parker (1985) shows that bucket SSTs are typically warmer than engine intake SSTs, so the results of the Tabata (1978) study are hard to interpret.

6.5.2 Comparison of Climatologies (Parker 1985, Folland et al. 1993, Quan et al. 1999)

Parker (1985) compared climatologies derived from "bucket" and "non-bucket" data (as described in Section 2.4) for the period 1975 to 1981. The mix of data should therefore be similar to that for the period of 1980 to 1997 studied here. He generated seasonal maps of bucket - non-bucket differences from monthly 5° x 5° climatology where there was sufficient data to be confident in the comparison. He found that on average buckets were 0.1°C colder than non-buckets, which is consistent with the results of both the present study and with the studies quoted in Section 6.5.1 when it is considered that the non-bucket category is known to contain a significant number of bucket reports leading to an underestimate of differences. The difference between the two methods was largest in the winter, again consistent with greater heat loss from the buckets in winter. His study contained data from both day and night so these results will also contain diurnal effects. Diurnal heating, both of the bucket and the ocean surface layer, can explain his further result that buckets are relatively warm in upper mid-latitudes in summer. All of these conclusions are confirmed by the bi-variate distribution of bucket-engine intake SST with solar radiation and wind speed shown in Figure 15. Moreover the pattern of difference shown in Figure 1a of Parker (1985) shows a strong similarity to that in Figure 16 of this report.

This suggests that the differences between climatologies found by Parker were indeed caused by the physical processes he suggests.

The Parker (1985) analysis forms the basis for the comparisons presented by Folland et al. (1993). They comment that the global difference they found of order 0.1°C might be smaller than those in previous studies due to the use of uninsulated buckets in the earlier studies. However Walden (1966) obtained a 0.3°C mean bias possibly from insulated buckets similar to the German bucket in Figure 10. Furthermore James and Fox (1972), describing measurements taken in 1968-1970 from buckets of known type, found that the mean difference from engine-intake SST for the German bucket was 0.18°C, the Crawford bucket (Crawford 1969) 0.23°C and “other” national buckets 0.25°C. It therefore seems likely that the Folland et al. (1993) and Parker et al. (1985) differences are underestimated due to the inclusion of buckets in the MOMDB “non-bucket and unknown category”. Their results should therefore be more significant than they estimate.

Quan et al. (1999) also compared separate climatologies based on different measurement methods, this time using the better defined COADS SST measurement type flags for the period 1980 to 1995. They found that globally buoy SST was about 0.1°C colder than the overall VOS SST and SST derived from eXpendable BathyThermographs (XBTs) was about 0.1°C warmer than the overall VOS SST. In the Gulf Stream region (75°W-40°W, 42°N-53°N) they find a strong seasonal signal in the difference between bucket and engine intake SST ranging from a warm relative bias of 0.35°C in the engine intake SST in January and a cold relative bias of -0.36°C in July. Again this strongly follows the air-sea temperature difference in the region and therefore supports the results of the present study. In the Kuroshio region (130°E-160°E, 45°N-55°N) the sense of the variation is the same but the engine intake remains relatively warm in all months (0.64°C in January and 0.15°C in July). They were unable however to exclude the possibility that differences in sampling, particularly in regions of strong SST gradient, caused these differences (Xiao-Wei Quan, pers. comm.).

6.5.3 Comparison with Model Output (Kent et al. 1993a)

The VSOP-NA study described in Section 5.3 (Kent and Taylor 1991, Kent et al. 1991, 1993a) compared a subset of the North Atlantic VOS with SST fields input to a Met Office operational forecasting model. They found a tendency for engine intake SST reports to become increasingly warm with increasing depth. This agrees with the findings of James and Fox (1972) which suggested that the distance inboard was more important than the measurement depth.

Hull sensors showed no large variations with measurement depth. All three methods of measurement (bucket, engine intake and hull sensor) became progressively warmer with increasing solar radiation, perhaps due to the lack of diurnal cycle in the model input SST field. Kent et al. (1993a) attributed the difference between the engine intake and bucket SST relative to the model field to a warm bias in the engine intake SST. The results of the present study suggests that at least part of the difference is due to a cold bias in the bucket SST.

6.5.4 Analysis of Individual Reports (Kent et al. 1999, Emery et al. 2001)

Kent et al. (1999) used data from COADS 1a for January and July 1980 and 1993 in a paired analysis of random errors in VOS observations using the semivariogram technique (see Section 4.3). They found that root mean square random errors calculated for 30° ocean regions varied from 0.4°C to 2.8°C with a mean value of $1.5 \pm 0.1^\circ\text{C}$. The present study shows that had they split their analysis by measurement type they may have explained some of the regional variation in random errors by the geographical distribution of measurement method. They also found larger random errors in regions containing the Gulf Stream and Kuroshio currents. These are regions where the air sea temperature difference can be large and variable which the present study suggests should lead to varying biases in bucket SSTs and hence increased random errors. They made no analysis of bias.

Emery et al. (2001) also make a paired analysis of reports within COADS. They look at data from 1990 and 1996 and make comparisons of drifting buoy, moored buoy and VOS SSTs. The paper does not describe in detail the procedures used in their analysis (for example the details of co-location and the quality control applied to the data) which makes some of the analyses hard to interpret. They however conclude that drifting buoys give similar SST measurements to nearby moored buoys (assuming a bias they quote of -0.96°C is in fact -0.096°C). They did not differentiate between bucket and engine intake SST reports and found an overall warm bias of ship SST relative to the drifting buoys of about 0.3°C . As there is no indication of the geographical distribution of the paired reports, or any stratification by time of day it is not possible to relate their findings to those of the present study.

7. DISCUSSION AND CONCLUSIONS

7.1 Accuracy of Bucket SST Measurements

The analysis of paired bucket and engine intake SST reports produces results that are usually easier to understand in terms of biases in the bucket measurements than in the engine intake measurements. Figure 15 shows how the mean difference between the two observation methods varies with wind speed and incident solar radiation. At zero solar radiation the difference between the measurement methods increases with increasing wind speed. As the wind speed increases we would expect the ocean to become better mixed, reducing any difference between the bucket measurements at the surface and the engine intake measurements at depth. The observed change is best explained by an increasing heat loss from the bucket with increasing wind speed (B1 and B2 in Section 3.1) or cooling of the thermometer by evaporation if the thermometer was removed from the water to read (B5).

At the highest wind speeds the bucket is warm, or the engine intake cold. This is again best explained by the difficulty of taking bucket measurements in high wind speeds as the difference between the two measurement methods increases with increasing solar radiation indicating that a bucket which has been warmed on deck has not had long enough to reach equilibrium with the sea temperature (B3). It has been suggested that in bad weather a ship normally reporting bucket SST may instead report an engine intake SST. If the method identification for this report was made using WMO Report No. 47, rather than the COADS metadata, the change of method would not be detectable. If the change in measurement method was missed then the difference calculated between the two methods would be reduced as the result would be a comparison of two reports with the same method. As the differences are large, it seems likely that there are ships attempting bucket measurements at high wind speeds. In these conditions the temperature of the bucket before the sample is taken probably has more effect on the reported temperature than the sea temperature itself. The number of these reports is small, but the errors can be large and it would be best to eliminate them from the dataset.

At low wind speeds the difference between the two measurement methods increases with increasing solar radiation, probably indicating the difference between the measurement depths of the two methods with the bucket measuring the temperature of the warm shallow surface layer which at these low wind speeds has not been mixed with the colder waters below (B8 and E2). There is a large region in the centre of the plot where the two methods agree to within $\pm 0.1^{\circ}\text{C}$. In this region there is too much wind or too little solar radiation to allow the formation of a

shallow warm layer and too little wind to cause significant cooling of the bucket sample. At higher wind speeds and low solar radiation cooling of the bucket dominates and at high solar radiation and low wind speeds the stratification of the ocean is the most important factor.

These results mean that the maps which have been presented of differences between bucket and engine intake SST in the North Atlantic can be understood (Figures 16 to 19). The bucket measurements are relatively cool in the regions where we expect strong heat loss to occur (Figures 16 and 19). Not only are the differences between the methods greater when the wind speed is higher, but the SST itself is lower for both methods (note change in scale of second panel between Figures 16 and 19). Under conditions of high solar radiation (Figure 17) the buckets become relatively warmer, this effect increases when the comparison is restricted to low wind speed cases (Figure 18).

This is a fairly straightforward picture which is easy to explain in terms of the physical processes we expect to be important under different environmental conditions. The problem comes when we wish to quantify these effects. For example the differences between the bucket and engine intake are expected to depend on both direct and evaporative heat loss from the bucket, the sensible and latent heat fluxes rather than on the wind speed. The fluxes are not measured by the VOS and we therefore need to parameterise the fluxes using mean meteorological measurements and bulk formulae (see Section 2.2.1 for the choice of parameterisations). Unfortunately we need to use the same parameters to calculate the fluxes that we know contain errors due to those fluxes. The problem is thus circular and needs careful analysis. In collaboration with Dr Alexey Kaplan a method to perform this analysis has been developed using information about the errors in the data and correlations between errors in different parameters. Initially a simple analysis has been performed using the air - sea temperature difference rather than the fluxes. For nighttime data at wind speeds between 4 ms^{-1} and 8 ms^{-1} the difference between the bucket and engine intake SST has been found to be related, for all but the three summer months, to the air - sea temperature difference (Figure 25 and Table 14). These results suggest that the bucket SST is biased by a fraction of about 0.2 (± 0.1) of the air - sea temperature difference. Once the cooling effect on the bucket had been taken into account there was no significant, consistent bias between the two measurement methods.

Many studies have shown that bucket SSTs are subject to heat loss. Walden (1966) found an increasing difference between bucket and engine intake SST with wind speed which he attributed to evaporation from the bucket or thermometer. Tauber (1969) measured heat loss from the Crawford (1969) bucket. James and Fox (1972) showed that bucket SSTs were colder

than engine intake SSTs, particularly in winter. Differences were greatest at night, at higher wind speeds, when there was precipitation and if the bucket observation was made on the windward side of the ship. Parker (1985) and Folland et al. (1993) interpreted maps of SST difference in terms of heating and cooling of the bucket observations but were hampered by relatively poor knowledge of the measurement method. Kent et al. (1991) showed that the bucket SST was cooler than engine intake SST and became relatively more cool as the sum of the sensible and latent heat fluxes (calculated using the model input SST) increased. Kent et al. (1993a) tentatively showed that the VSOP-NA bucket SSTs are affected by solar radiation. The present study has shown that more recent bucket SST measurements are also subject to heat loss. In addition a better knowledge of SST measurement method due to the use of external metadata has allowed a better quantification of the effects for the VOS as a whole rather than for a selected subset of ships.

Although many of the differences between the bucket and the engine intake SST can be explained in terms of biases in the bucket SST the random errors for the bucket SSTs are still much lower than those for the engine intake SST. Figure 12 shows that this difference in quality can be attributed largely to the recruiting country. The distributions of normalised deviations from the local mean typically show small differences between the bucket and engine intake reports for a given country, whereas the differences between countries are much greater. The relatively small random errors in the bucket SST observations seem therefore to be due to careful observing practice by ships from countries that have a national preference for bucket reports. Countries with a national reporting preference for engine intake SSTs tend to report poorer quality bucket SSTs for the minority of their ships using this method.

7.2 Accuracy of Engine Intake SST Measurements

The results of the previous section suggest that biases in the difference between bucket and engine intake SST measurements can be largely explained in terms of biases in the bucket SST. This is at variance with the widely held belief that engine intake SST measurements are biased warm (Saur 1963, Walden 1966, Tauber 1969, James and Fox 1972, Tabata 1977, Kent et al. 1993a and Emery et al. 2001).

Saur (1963) used the bucket as the standard for comparison and so attributed the difference of 0.7°C between the two measurement methods to a bias in the engine intake measurement. The engine intake observations in the Saur (1963) study were made using mercury thermometers

which were poorly mounted and maintained, but it is likely that on more modern ships the engine intake temperature is not measured in this way.

Walden (1966) found a bias of 0.3°C between the two methods and although finding evidence of cooling of the buckets and a strong diurnal signal due to the formation of a warm layer the author states that these two effects will at least partly cancel leaving "no doubt that at the inlet to the engines on average too high a temperature is read".

Tauber (1969) found warm biases of more than 0.5°C in 98% of the engine intake SSTs in a study on the "Academician Shirshov" and on other Russian vessels of greater than 1.2°C in 85% of measurements. He noted that the size of the error depended on the operating conditions of the engine room. When drifting with the engines stopped the flow decreased and errors of 8-10°C were observed. It is clear that engine intake SST measurements from the VOS are not in general quite this bad, so again perhaps the conclusions with reference to biases in engine intake SSTs are not relevant to more modern ships.

James and Fox (1972) had extensive metadata for the engine intakes in their study and concluded that the distance inboard of the inlet thermometer and the type of the thermometer were the most important indicators of engine intake temperature quality. Warm biases compared with the buckets reduced by 0.2°C to 0.1°C when only measurements from thermistors and precision thermometers were considered. For thermometers less than 3 metres inboard the difference was again 0.1°C compared to 0.9°C for thermometers more than 7 metres inboard. It seems possible therefore that the use of better quality thermometers has decreased the average bias in the engine intake SSTs (although a large scatter remains). Another possibility is that the intake pipes are now better insulated than in the past, leading to a smaller heating effect despite the probably increase in distance inboard of the thermometers. This is however only speculation.

Tabata (1977) compared VOS SST with OWS and buoy SST but did not consider the types of VOS measurement method separately. Tabata found VOS SSTs to be on average $0.2 \pm 1.5^\circ\text{C}$ higher than OWS and buoy SSTs and concluded "that the majority of the ships' observations were based on the engine intake method as this method has been shown to yield a slightly higher than the actual temperature (Saur 1963)". As shown by Parker (1985), the northeast Pacific where this study was conducted is a region where the bucket SSTs tend to be warmer than the engine intakes.

Kent et al. (1993a) attributed the difference between the measurement methods to errors in engine room intake SSTs based on the results of James and Fox (1972). They justified this on the increasing bias in engine intake SST with measurement depth (relating this to increased distance inboard of the thermometer, as in James and Fox, 1972) which was not observed in the hull sensor data.

Emery et al. (2001) compare individual VOS reports with drifting buoy SSTs having first verified the drifting buoys against moored buoy SSTs. They concluded that the VOS SSTs overall were 0.28°C warmer than the drifting buoys due to heating in the engine room, referencing Saur (1963). Emery et al. (2001) used COADS for their analysis so it should be possible to repeat their analysis taking into account the different measurement methods which would help in understanding their results.

It is therefore possible that studies which indicate that engine intake SSTs are warm are either based on outdated measurement methods (Saur 1963, Tauber 1969, James and Fox 1972, although the latter study contains a breakdown by engine intake thermometer type it is the overall bias that is usually quoted), ignored possible cooling in the bucket SSTs (Walden 1966, Kent et al. 1993a) or were influenced in their conclusions by outdated studies (Tabata 1977, Kent et al. 1993a and Emery et al. 2001). More work is however needed.

As discussed in the previous section the random errors in engine intake SST are larger than for buckets but the quality of the observations varies from country to country.

7.3 Accuracy of Marine Air Temperature Measurements

Like SST observations the quality of marine air temperature varies from country to country and the countries that have a national preference for bucket SST observations make the best quality air temperature observations (Figure 12). Air temperature observations are more consistent when they are taken at high relative wind speeds, true wind speeds or ship speeds, from a high observing platform or when it is raining (Figure 13, Tables 4, 5, 6, 8 and 10). Thermometers exposed in a sling may have smaller random errors than those exposed in a screen (Table 3). Air temperature observations need to be corrected for the effects of solar heating. Kent et al. 1993b suggested a correction scheme for daytime air temperature but improvements are needed to the algorithm to allow for the heat island effect of the ship (Aiguo Dai, pers. comm.). It is likely that much of the random error in the air temperature observations results from solar contamination (see for example Figure 33) and this limits our knowledge of the air -

sea temperature difference (see Figures 20 and 21) which is required to assess the biases in SST observations. Air temperature data will therefore need correcting before an assessment of bias in the daytime SST observations can be made. Data from the VSOP-NA project are being reanalysed in an attempt to improve the air temperature correction (see Figure 32). The effect of solar heating of the ship environment may be detectable several hours after sunset (see Figure 32b).

Another cause of bias in marine air temperature is varying observation height. The use of observing platform height metadata from WMO Report No. 47 in this study has allowed the regional and temporal distribution of observation heights, and therefore air temperature bias, to be calculated (see Figures 29 to 31).

7.4 Future Research

Planned future research includes a comparison of VOS SSTs with drifting and moored buoys (partly to help understand the results of Emery et al. 2001), an extension of the statistical analysis technique to include wind speed and fluxes and to include daytime SST. The analysis of daytime SST biases will require the correction of daytime air temperature data for the biases, work which is ongoing at SOC. A new contract has been proposed to include some of this work and also a further assessment of the impact of changing air temperature observing height on air temperature biases and investigating the feasibility of extending the observing height time series back into the 1970s.

Other studies beyond the scope on any extension to this contract are the analysis of research ship VOS data, the analysis of the VOSclim dataset and the development of an improved correction for VOS air temperature to remove solar radiation effects.

ACKNOWLEDGEMENTS

Jeffrey Blundell wrote the ocean mask code. David Berry implemented the code and produced the ocean mask. David Berry also produced the ice concentration dataset. Diane Stokes provided an alternative code for ship tracking and advice on its implementation. James P. Rigney of the US Naval Oceanographic Office helped in trying to track down the James and Fox (1972) dataset, unfortunately unsuccessfully. Discussions with Povl Frich (quality control), Sandra Lubker (duplicate elimination and track checking), Xiao-Wei Quan (methods of improved trimming and quality control), Scott Woodruff (data characteristics) and Steven

Worley (data access and storage) proved useful during this work. I must also thank Alexey Kaplan, Peter Taylor and Peter Challenor for their help.

COADS/MOMDB: Steven Worley of the National Center for Atmospheric Research Data Support Section provided the COADS/MOMDB data.

WMO Report No. 47: Data prior to 1995 was obtained from Joe Elms at the National Climatic Data Center, Asheville, NC, via Arlindo daSilva then of the University of Wisconsin, Milwaukee. Later data were obtained from the World Meteorological Organisation ftp site: <ftp://www.wmo.ch/wmo-ddbs/> Data after 1999 is currently not available, but should become so from this site in the near future (Teruko Manabe, pers. comm.).

Sea-Ice concentration: Provided by the EOS Distributed Active Archive Center (DAAC) at the National Snow and Ice Data Center, University of Colorado, Boulder, CO.

ETOPO5: Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado, 1988.

SST Climatology: Jim Arnott of the Hadley Centre provided HadISST and GISST2.2 data used in this study.

REFERENCES

- Åmot, A., 1955: Measurements of Sea Surface Temperature for Meteorological Purposes. Results of Observations from Ocean Weather Station M. Meteorologiske Annaler, Volume 4, No. 1, 11 pp.
- Boggs, P. T, R. H. Byrd, J. E. Rogers and R. B. Schnabel, 1992: User's Reference Guide for ODRPACK Version 2.01 Software for Weighted Orthogonal Distance Regression, National Institute of Standards and Technology, NISTIR 92-4834, U. S. Department of Commerce, Gaithersburg, MD, 99 pp.
- Cavalieri, C., L. Parkinson, P. Gloersen, and H. J. Zwally, 1997: Arctic and Antarctic sea ice concentrations from multichannel passive-microwave satellite data sets: October 1978 to September 1995, User's Guide, 17 pp.

- Christy, J. R., D. E. Parker, S. J. Brown, I. Macadam, M. Stendel and W. B. Norris, 2001: Differential trends in tropical sea surface and atmospheric temperatures since 1979. *Geophysical Research Letters*, 28, 183-186.
- Clark, N. E., L. Eber, R. M. Laurs, J. A. Renner, and J. F. T. Saur, 1974: Heat Exchange Between Ocean and Atmosphere in the North Eastern Pacific for 1961-1971., NOAA Technical Report NMFS SSRF-682, Department of Commerce, Seattle, WA., 108 pp.
- Crawford, A. B., 1969: Sea-surface temperatures: Some instruments, methods and comparisons, World Meteorological Organisation Technical Note No 103 (WMO No. 247.TP.135), WMO Geneva, 117-129.
- Dobson, F. W. and S. D. Smith, 1988: Bulk models of solar radiation at sea. *Quarterly Journal of the Royal Meteorological Society*, 114(479), 165-182.
- Emery, W. J., D. J. Baldwin, P. Schlüssel, and R. W. Reynolds, 2001: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, 106, C2, 2387-2405.
- Folland, C. K., R. W. Reynolds, M. Gordon, and D. E. Parker, 1993: A study of six operational sea surface temperature analyses. *Journal of Climate*, 6, 96-113.
- Franceschini, G. A., 1955: Reliability of Commercial Vessel Reports of Sea Surface Temperatures in the Gulf of Mexico, *Bulletin of Marine Science of the Gulf and Caribbean*, Volume 5, No. 1, 42-51.
- IPCC, 2001: Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment, Report of the Intergovernmental Panel on Climate Change, UNEP, WMO.
- James, R. W. and P. T. Fox, 1972: Comparative sea surface temperature measurements., Reports on Marine Science Affairs, No 5, (WMO336), 27 pp.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, E. 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. *Journal of Geophysical Research* 106, 3371-3380.

Josey, S. A., E. C. Kent, and P. K. Taylor, 1999: New Insights into the Ocean Heat Budget Closure Problem from Analysis of the SOC Air-sea Flux Climatology. *Journal of Climate*, 12, 2856 - 2880.

Kent, E. C., P. G. Challenor, and P. K. Taylor, 1999: A Statistical Determination of the Random Observational Errors Present in Voluntary Observing Ships Meteorological Reports. *Journal of Atmospheric and Oceanic Technology*, 16, 905-914.

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 Report No. 25, WMO/TD No. 456, WMO Geneva*. 123 pp.

Kent, E. C. and P. K. Taylor, 1996: Accuracy of humidity measurements on ships: Consideration of solar radiation effects. *Journal of Atmospheric and Oceanic Technology*, 13, 1317-1321.

Kent, E. C., P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993a: The accuracy of voluntary observing ship's meteorological observations - Results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology*, 10, 591-608.

Kent, E. C., R. J. Tiddy, and P. K. Taylor, 1993b: Correction of marine air temperature observations for solar radiation effects. *Journal of Atmospheric and Oceanic Technology*, 10, 900-906.

Kent, E. C., B. S. Truscott, J. S. Hopkins and P. K. Taylor, 1991: The Accuracy of Ship's Meteorological Observations: Results of the VSOP-NA., *Marine Meteorology and Related Oceanographic Activities Report No. 26, WMO/TD No. 455, WMO Geneva*. 123 pp.

Kirk, T. H. and A. H. Gordon, 1952: Comparison of Intake and Bucket Methods for Measuring Sea Temperature, *The Marine Observer*, Volume XXII, No. 155, 33-39.

Lindau, R., 1995: A New Beaufort Equivalent Scale. *Proceedings of the International COADS Winds Workshop*, Kiel, Germany, Institut fur Meereskunde, Kiel/NOAA, 232-252.

- Moat, B. I., 2002: The airflow distortion over Merchant ships. MPhil/PhD Upgrade Report submitted to the University of Southampton, 64 pp.
- NOAA, 1988: Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado.
- Parker, D. E., 1985: A comparison of bucket and non-bucket measurements of sea surface temperature. Met O 13 Branch Memorandum, Meteorological Office, Bracknell, UK., 16 pp.
- Parker, D. E and N. A. Rayner, 2002: Construction and Testing of the HadMAT Gridded Night Marine Air Temperature Analysis, Abstract presented at the International Workshop on Advances in the Use of Historical Marine Climate Data: Boulder, Colorado, USA, January 2002.
- Parker D. E., C. K. Folland and M. Jackson, 1995: Marine surface temperature: Observed variations and data requirements, *Climatic Change*, 31(2-4), 559-600.
- Quan, X.-W., H. F. Diaz, S. D. Woodruff, S. Lubker, and J. Eischeid, 1999: Comparison of Ship-Observed Sea Surface Temperature with Measurements from Drifting Buoys and Expendable Bathythermographs: 1980-95. WMO Workshop on Advances in Marine Climatology - CLIMAR99, Vancouver, Canada, Environment Canada, 319-323.
- Rayner, N. A., Horton, E. B., Parker, D. E., Folland, C. K. and Hackett, R. B. 1996: Version 2.2 of the Global sea-Ice and Sea Surface Temperature Data Set, 1903-1994. Climate Research Technical Note 74, unpublished document available from Hadley Centre.
- Rayner, N. A., D. E. Parker, C. K. Folland, L. V. Alexander and E. B. Horton, 2002: Globally complete analyses of sea surface temperature, sea-ice and marine air temperature, 1871-2000. To be submitted to *Journal of Geophysical Research*.
- Reynolds, R. W. and T. M. Smith, 1994: Improved Global Sea Surface Temperature Analyses using Optimum Interpolation. *Journal of Climatology*, 7, 929-948.
- Roll, H. U., 1951: Water Temperature Measurements on Deck and in the Engine Room, *Annalen der Meteorologie*, Volume 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. *Journal of Applied Meteorology*, 2, 417-425.
- Smith, S. D., 1980: Wind Stress and Heat Flux over the Ocean in Gale Force Winds. *Journal of Physical Oceanography*, 10, 709-726.
- Smith, S. D., 1988: Coefficients for Sea Surface Wind Stress, Heat Flux and Wind Profiles as a Function of Wind Speed and Temperature. *Journal of Geophysical Research*, 93, 15,467-15,474.
- Tabata, S., 1978: Comparison of Observations of Sea Surface Temperatures at Ocean Station P and NOAA Buoy Stations and Those Made by Merchant Ships Travelling in Their Vicinities, in the Northeast Pacific Ocean. *Journal of Applied Meteorology*, 17, 374-385.
- Tauber, G. M., 1969: The Comparative Measurements of Sea Surface Temperature in the U.S.S.R., World Meteorological Organisation Technical Note No 103 (WMO No. 247.TP.135), WMO Geneva, 141-151.
- Walden, H., 1966: On water temperature measurements aboard merchant vessels. *Deutsche Hydrographische Zeitschrift* [In German. English translation available in National Meteorological Library, Bracknell, UK], 19, 21-28.
- WMO, 1994: International List of Selected, Supplementary and Auxiliary Ships, WMO Report No. 47, WMO, Geneva, various pagination.
- WMO, 2000: Voluntary Observing Ships (VOS) Climate Subset Project (VOSCLIM): Project Document., JCOMM Technical Report No. 5, WMO/TD No. 1010, WMO, Geneva, 18 pp.
- Wolter, K., 1997: Trimming problems and remedies in COADS. *Journal of Climate*, 10, 1980-1997.
- Woodruff, S. D., S. J. Lubker, K. Wolter, S. J. Worley, and J. D. Elms, 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980-92. *Earth System Monitor*, 4, 4-8.

Woodruff, S., S. Worley, J. Elms, D. Parker, and H. Diaz, 1999: COADS Enhancements and the Blend with the U.K. Met. Office Marine Data Bank. Proceedings of OceanObs99: The Ocean Observing System for Climate, 18-22 Oct. 1999, Saint Raphael.

Yelland, M.J. and R.W. Pascal, 2001: Performance of the WHOI SST system during a five week trial on the RRS James Clark Ross, unpublished report, SOC JRD Meteorology Team.

TABLES

	[1] Bucket (WMO 47) no of reports	[2] ERI (WMO 47) no of reports	[3] Hull Sensor (WMO 47) no of reports	[4] % with SST sensor from data	[5] % with SST sensor from data and WMO 47	[6] % gain using WMO 47
Bucket (SI flag)	27358	7883	63	24.0	29.1	5.1
ERI (SI flag)	3343	13380	0	11.5	38.0	26.5
Hull Sensor (SI)	221	21	282	0.9	1.5	0.6
Unknown (SI)	7474	53723	1184	63.6	31.4	-32.2

Table 1: Number and percentage of reports with sensor information from COADS and WMO Report No 47 for April 1985.

Year	SST				Air Temperature		
	Bucket %	ERI %	Hull Sensor %	Unknown %	Screen %	Psychro- meter %	Unknown %
1980	30.63	17.73	0.24	51.40	15.40	11.78	72.77
1981	30.22	20.87	0.39	48.52	15.47	11.54	72.96
1982	28.54	36.16	1.33	33.98	28.92	18.32	52.25
1983	28.44	35.68	1.50	34.38	31.09	18.19	50.64
1984	29.11	36.21	1.35	33.33	31.34	19.32	49.02
1985	29.05	37.99	1.52	31.43	36.74	18.98	44.10
1986	27.18	40.10	1.56	31.15	40.70	17.88	41.34
1987	22.26	40.38	2.01	35.35	39.01	17.95	42.80
1988	21.87	38.69	1.54	37.89	42.07	19.20	38.26
1989	18.82	35.83	1.33	44.02	40.07	20.12	39.44
1990	20.18	37.22	2.00	40.61	42.28	22.38	33.97
1991	21.62	32.76	1.88	43.75	40.58	25.52	32.28
1992	24.84	29.83	2.03	43.30	36.17	26.08	37.02
1993	21.65	29.16	3.12	46.07	36.43	26.78	35.95
1994	17.93	30.35	3.25	48.47	31.55	29.05	37.94
1995	19.80	55.30	5.88	19.02	35.38	30.86	32.51
1996	24.40	55.94	3.45	16.21	34.38	31.54	33.18
1997	26.24	38.41	4.04	31.32	30.06	26.77	42.67

Table 2: Sensor information from COADS and WMO Report No 47 for SST and Air Temperature reports within COADS for Aprils between 1980 and 1997.

Type of thermometer	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
Mercury	461517	0.91	0.01	1.08	0.01	752380	1.53	0.00	1.72	0.00
Electric	412	0.49	0.18	0.63	0.18	3388	0.61	0.08	0.29	0.21
Alcohol	< 10					71325	1.48	0.02	0.76	0.02
Method of Exposure										
Screen	269270	0.86	0.01	1.15	0.01	532308	1.50	0.01	1.78	0.00
Ventilated Screen	< 10					12206	1.46	0.05	0.68	0.06
Sling	45789	0.97	0.03	0.71	0.03	92246	1.59	0.02	1.01	0.02
Whirling Psychrometer	1254	1.10	0.13	0.72	0.13	1305	1.08	0.15	0.72	0.13
Aspirated	1164	1.59	0.08	1.78	0.08	3838	1.48	0.06	1.47	0.06
Unscreened	< 10					354	1.73	0.23	2.33	0.23
Ship's screen	177	1.67	0.20	1.78	0.20	1125	1.95	0.13	1.57	0.15

Table 3: Error estimates ($^{\circ}\text{C}$) for SST and air temperature separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} . Error estimates are given separately for each type of thermometer (mercury, electrical resistance and alcohol) and for the method of exposure of the thermometer (screen, ventilated screen, sling, whirling psychrometer, aspirated psychrometer, unscreened and ship's screen). The information about instrumentation is merged into COADS from WMO Report No. 47.

Ship Speed ms^{-1}	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0 - 2	188490	0.61	0.01	0.77	0.01	100659	1.45	0.01	1.79	0.01
2 - 4	85336	0.91	0.02	1.09	0.02	136161	1.45	0.01	1.71	0.01
4 - 6	327300	0.95	0.01	0.64	0.01	403229	1.62	0.01	1.31	0.01
6 - 8	266928	0.96	0.01	0.55	0.01	324358	1.65	0.01	1.22	0.01
8 - 10	135131	1.01	0.02	0.52	0.02	262496	1.57	0.01	1.02	0.01
10 - 12	152627	1.00	0.01	0.52	0.01	342667	1.56	0.01	1.02	0.01
12 - 14	17584	0.89	0.04	0.58	0.04	80645	1.52	0.02	1.01	0.02

Table 4: Error estimates ($^{\circ}\text{C}$) for SST and air temperature, in ranges of ship speed in ms^{-1} , separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

Relative wind speed ms^{-1}	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0 - 2	4902	0.87	0.04	0.76	0.04	4090	1.56	0.06	1.50	0.05
2 - 4	20009	0.84	0.03	0.88	0.03	19510	1.53	0.03	1.59	0.02
4 - 6	31628	0.86	0.02	0.93	0.02	34421	1.53	0.02	1.62	0.02
6 - 8	41420	0.81	0.02	0.88	0.02	41686	1.55	0.02	1.57	0.02
8 - 10	46485	0.78	0.02	0.78	0.02	44651	1.53	0.02	1.48	0.02
10 - 12	46296	0.76	0.02	0.68	0.02	44607	1.49	0.02	1.32	0.02
12 - 14	38952	0.78	0.03	0.60	0.03	38491	1.52	0.02	1.24	0.02
14 - 16	29088	0.76	0.03	0.60	0.03	29550	1.43	0.03	1.23	0.03
16 - 18	20255	0.76	0.04	0.57	0.04	20190	1.43	0.04	1.15	0.03
18 - 20	13626	0.76	0.04	0.57	0.04	12936	1.51	0.04	1.14	0.04
20 - 22	8046	0.81	0.05	0.62	0.05	7431	1.58	0.05	1.12	0.06
22 - 24	3581	0.88	0.07	0.60	0.07	3465	1.60	0.08	1.22	0.07
24 - 26	1488	0.71	0.15	0.66	0.15	1367	1.67	0.12	1.14	0.14
26 - 28	547	0.71	0.18	0.33	0.18	478	1.43	0.23	0.89	0.36
28 - 30	184	0.81	0.35	0.35	0.39	151	1.33	0.50	0.93	0.34

Table 5: Error estimates ($^{\circ}\text{C}$) for SST and air temperature, in ranges of relative wind speed in ms^{-1} , separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

Platform Height	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0 - 5	2412	1.31	0.05	1.90	0.05	2704	1.48	0.06	1.69	0.07
5 - 10	49743	1.05	0.01	1.71	0.01	107432	1.50	0.01	1.87	0.01
10 - 15	94301	0.99	0.01	1.61	0.01	314007	1.44	0.01	1.89	0.01
15 - 20	569691	0.93	0.01	0.75	0.01	438717	1.54	0.01	1.41	0.01
20 - 25	504244	0.93	0.01	0.64	0.01	370504	1.55	0.01	1.01	0.01
25 - 30	25712	0.96	0.03	0.46	0.03	98915	1.45	0.02	0.86	0.02
30 - 35	5024	0.92	0.08	0.35	0.08	25619	1.37	0.04	0.84	0.04
35 - 40	3547	0.70	0.02	0.59	0.02	3235	1.27	0.10	0.74	0.13
40 - 45	1862	0.67	0.07	0.35	0.07	1673	1.16	0.09	0.78	0.08

Table 6: Error estimates ($^{\circ}\text{C}$) for SST and air temperature, in ranges of platform height in metres, separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} . Platform height is merged from WMO Report No. 47.

Latitude	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0 - 10	105621	1.02	0.01	0.78	0.01	95428	1.30	0.01	1.13	0.01
10 - 20	145768	1.08	0.01	0.81	0.01	188615	1.39	0.01	1.13	0.01
20 - 30	187768	1.08	0.01	0.69	0.01	286574	1.52	0.01	1.09	0.01
30 - 40	354662	0.98	0.01	0.54	0.01	557003	1.79	0.01	1.30	0.01
40 - 50	453284	0.85	0.01	0.67	0.01	641317	1.76	0.01	1.42	0.01
50 - 60	345530	0.83	0.01	0.99	0.01	351365	1.56	0.01	1.57	0.01
60 - 70	123841	0.81	0.01	1.18	0.01	120676	1.49	0.01	1.79	0.01
70 - 80	12483	0.91	0.02	1.42	0.02	40743	1.63	0.01	2.00	0.02
80 - 90	389	1.08	0.10	2.41	0.10	2494	1.55	0.04	2.19	0.05

Table 7: Error estimates ($^{\circ}\text{C}$) for SST and air temperature, in ranges of North Atlantic latitude, separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

Wind Speed (ms ⁻¹)	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0 - 2	34404	0.93	0.02	0.81	0.02	23225	1.60	0.02	1.60	0.02
2 - 4	141107	0.98	0.01	0.99	0.01	127981	1.53	0.01	1.66	0.01
4 - 6	189330	0.99	0.01	1.10	0.01	242816	1.49	0.01	1.66	0.01
6 - 8	172486	0.98	0.01	1.13	0.01	290092	1.49	0.01	1.61	0.01
8 - 10	129200	0.95	0.01	1.06	0.01	222862	1.50	0.01	1.50	0.01
10 - 12	93229	0.87	0.02	0.87	0.02	135018	1.48	0.01	1.42	0.01
12 - 14	45192	0.83	0.02	0.91	0.02	68409	1.48	0.02	1.48	0.01
14 - 16	29123	0.74	0.03	0.79	0.03	36952	1.51	0.02	1.46	0.02
16 - 18	25453	0.76	0.03	0.67	0.03	22235	1.53	0.03	1.45	0.03
18 - 20	14149	0.70	0.05	0.69	0.05	11435	1.48	0.04	1.50	0.04
20 - 22	7372	0.73	0.06	0.72	0.06	4531	1.42	0.07	1.50	0.05
22 - 24	2608	0.67	0.09	0.60	0.09	1722	1.40	0.10	1.81	0.08
24 - 26	1577	0.57	0.13	0.50	0.13	841	1.52	0.16	1.74	0.11

Table 8: Error estimates (°C) for SST and air temperature, in different ranges of wind speed in ms⁻¹, separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

Cloud Cover	Bucket					Engine Intake				
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0	49542	0.99	0.02	0.78	0.02	37091	1.66	0.02	1.22	0.02
1	98225	0.99	0.01	0.75	0.01	85722	1.60	0.01	1.18	0.01
2	92530	0.95	0.01	0.64	0.01	92041	1.45	0.01	1.07	0.01
3	90699	0.94	0.01	0.70	0.01	99366	1.45	0.01	1.13	0.01
4	65229	0.94	0.02	0.80	0.02	73940	1.45	0.02	1.23	0.01
5	61466	0.93	0.02	0.93	0.02	73989	1.44	0.02	1.32	0.01
6	78644	0.90	0.02	0.90	0.02	99822	1.42	0.01	1.34	0.01
7	305048	0.85	0.01	0.89	0.01	348556	1.43	0.01	1.35	0.01
8	311215	0.83	0.01	0.90	0.01	512823	1.50	0.01	1.65	0.01

Table 9: Error estimates (°C) for SST and air temperature separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} . Error estimates are given separately for observations if different ranges of cloud cover in octas.

Rain	Bucket				Engine Intake					
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
none	829837	0.97	0.01	0.99	0.01	1127576	1.53	0.00	1.54	0.00
some	33822	0.69	0.03	0.87	0.03	233252	1.46	0.03	1.46	0.02

Table 10: Error estimates ($^{\circ}\text{C}$) for SST and air temperature observations separately when precipitation is absent and present and for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

Ice Concentration	Bucket				Engine Intake					
	Number	σ_{sb}	σ_{sb} uncertainty	σ_{ab}	σ_{ab} uncertainty	Number	σ_{se}	σ_{se} uncertainty	σ_{ae}	σ_{ae} uncertainty
0.0 - 0.1	982073	0.95	0.00	0.95	0.00	1303524	1.53	0.00	1.53	0.00
0.1 - 0.2	4157	0.82	0.02	1.86	0.02	3565	1.18	0.03	2.01	0.05
0.2 - 0.3	4052	0.76	0.02	1.96	0.02	2378	1.16	0.03	2.13	0.05
0.3 - 0.4	2406	0.80	0.02	1.82	0.02	1669	1.19	0.04	1.93	0.07
0.4 - 0.5	1192	0.99	0.04	2.10	0.04	741	1.33	0.07	1.57	0.10
0.5 - 0.6	948	0.91	0.04	2.09	0.04	389	1.57	0.10	1.95	0.13
0.6 - 0.7	310	0.80	0.05	1.82	0.05	213	1.13	0.15	1.63	0.16
0.7 - 0.8	372	0.72	0.05	2.39	0.05	67	0.82	0.13	1.69	0.29
0.8 - 0.9	319	0.69	0.06	2.31	0.06	22	0.74	0.39	0.98	0.30

Table 11: Error estimates ($^{\circ}\text{C}$) for SST and air temperature for observations in different ranges of sea-ice concentration in the closest 1° area to the reports, separately for ships that use buckets to measure SST and ships that use engine intakes. σ_{sb} is the random error for T_{sb} , σ_{se} the random error for T_{se} , σ_{ab} the random error for T_{ab} and σ_{ae} the random error for T_{ae} .

		January	January	July	July
Code	Country	Bucket	Engine Intake	Bucket	Engine Intake
0	Netherlands	22244	13746	21842	13309
1	Norway	2552	2663	4174	3952
2	USA	7995	84307	11428	95574
3	UK	120453	14752	130837	15917
4	France	4782	17957	5999	21883
5	Denmark	906	186	2331	543
8	Hong Kong	4146	2653	3886	2856
9	New Zealand	646	447	556	572
13	Canada	15677	1971	20743	2761
14	Belgium	3213	607	2409	595
15	South Africa	4	2292	16	2278
16	Australia	13	958	24	725
17	Japan	19034	27337	23358	30824
20	Sweden	934	1460	1445	1233
23	Israel	1387	1496	1469	1582
25	Russia	49534	142983	50059	142898
35	Singapore	4123	542	4019	682
40	Germany	11539	1666	10335	1350

Table 12: Number of reports in paired data file for the eighteen countries selected as reporting large quantities of good quality data.

	$2\sigma_{sb}^2$	$2\sigma_{se}^2$	$\sigma_{sb}^2 + \sigma_{se}^2$	$2\sigma_{ab}^2$	$2\sigma_{ae}^2$	$\sigma_{ab}^2 + \sigma_{ae}^2$
January	1.26 ± 0.10	4.05 ± 0.16	3.00 ± 0.11	0.69 ± 0.05	2.10 ± 0.08	1.51 ± 0.06
February	1.51 ± 0.10	4.48 ± 0.16	3.57 ± 0.12	0.85 ± 0.06	2.47 ± 0.09	1.77 ± 0.06
March	1.29 ± 0.14	4.42 ± 0.17	3.42 ± 0.12	0.86 ± 0.06	2.75 ± 0.09	1.71 ± 0.06
April	1.96 ± 0.15	4.67 ± 0.21	3.77 ± 0.17	0.99 ± 0.07	2.85 ± 0.10	1.95 ± 0.08
May	1.41 ± 0.14	4.84 ± 0.22	3.84 ± 0.18	0.95 ± 0.07	3.08 ± 0.10	1.95 ± 0.08
June	1.78 ± 0.15	4.45 ± 0.25	3.13 ± 0.17	0.93 ± 0.07	2.72 ± 0.11	1.62 ± 0.08
July	1.71 ± 0.15	5.40 ± 0.26	3.66 ± 0.20	0.75 ± 0.07	2.53 ± 0.11	1.52 ± 0.09
August	1.69 ± 0.14	4.55 ± 0.28	2.62 ± 0.19	0.53 ± 0.07	2.10 ± 0.12	1.29 ± 0.09
September	1.21 ± 0.12	3.43 ± 0.23	2.67 ± 0.16	0.51 ± 0.06	1.70 ± 0.10	1.25 ± 0.07
October	1.55 ± 0.11	3.31 ± 0.17	2.90 ± 0.12	0.68 ± 0.05	1.50 ± 0.08	1.27 ± 0.06
November	2.22 ± 0.13	4.21 ± 0.18	3.07 ± 0.13	0.64 ± 0.06	1.81 ± 0.09	1.21 ± 0.06
December	1.30 ± 0.12	4.19 ± 0.17	3.27 ± 0.12	0.58 ± 0.06	2.06 ± 0.09	1.40 ± 0.06

Table 13: Semivariogram intercepts used to generate the correlation matrix used to transform each monthly data file.

		Gradient	Gradient	Gradient	Intercept	Intercept	Intercept
Month	Number of Observations	Best-estimate	Lower limit	Upper limit	Best-estimate	Lower limit	Upper limit
January	1476	0.15	0.10	0.20	-0.04	-0.12	0.04
February	1508	0.22	0.17	0.26	-0.11	-0.20	-0.03
March	1381	0.14	0.09	0.19	0.06	-0.01	0.12
April	990	0.24	0.13	0.34	-0.04	-0.17	0.06
May	1019	0.29	0.17	0.40	-0.01	-0.12	0.08
June	1041	-0.21	-0.44	0.01	-0.09	-0.14	-0.06
July	863	0.01	-0.28	0.28	0.06	0.01	0.09
August	775	-0.29	-0.60	-0.03	0.19	0.13	0.23
September	1074	0.18	0.03	0.33	-0.05	-0.20	0.06
October	1470	0.32	0.23	0.40	-0.27	-0.44	-0.12
November	1398	0.11	0.04	0.17	0.02	-0.10	0.12
December	1242	0.18	0.12	0.24	-0.02	-0.12	0.08

Table 14: Regression results by month. The best estimates are given for the gradient and intercept along with estimates of minimum and maximum values. See text for details of how the regression parameters and their uncertainties were calculated.

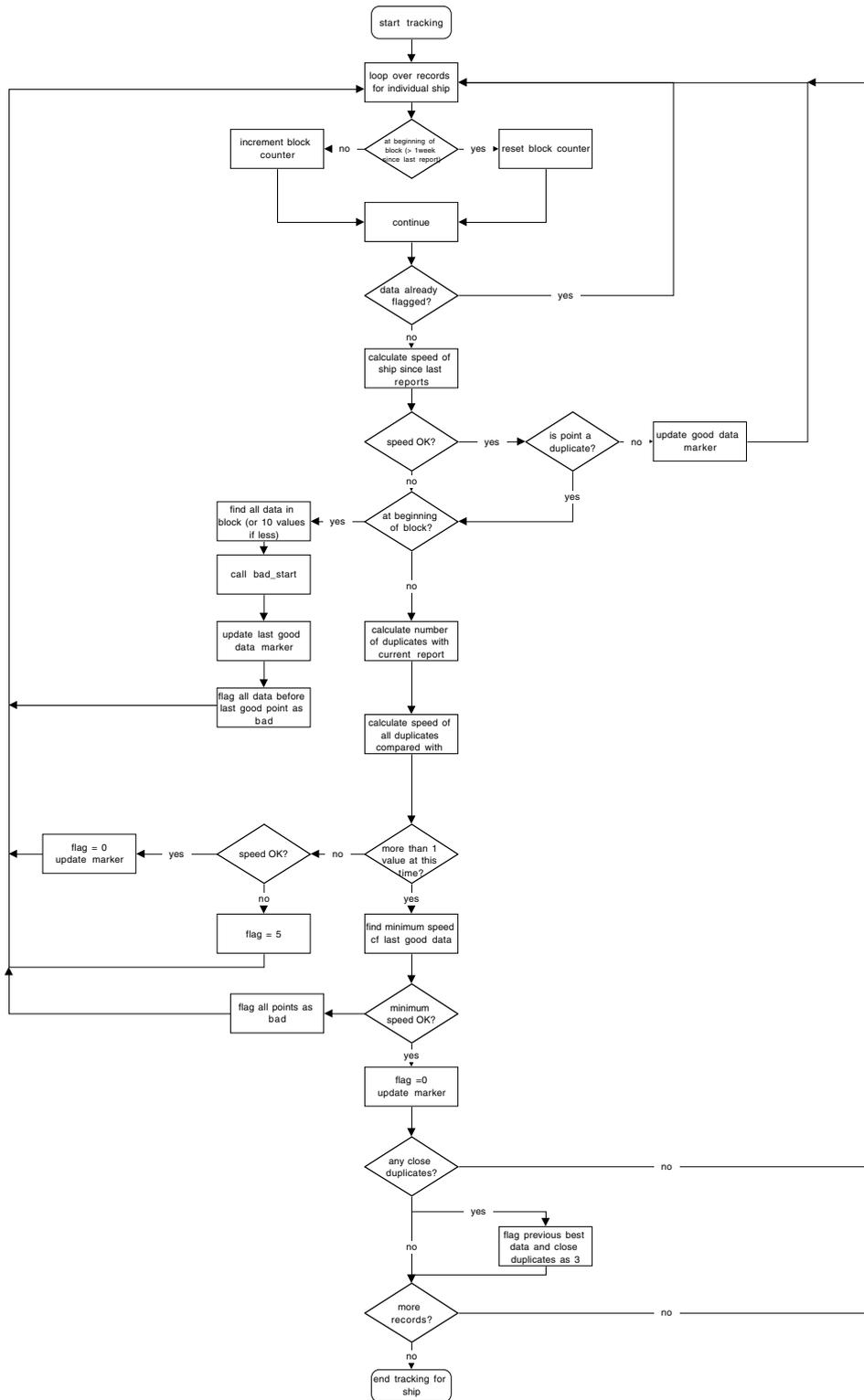


Figure 1 Flow chart of ship tracking program.

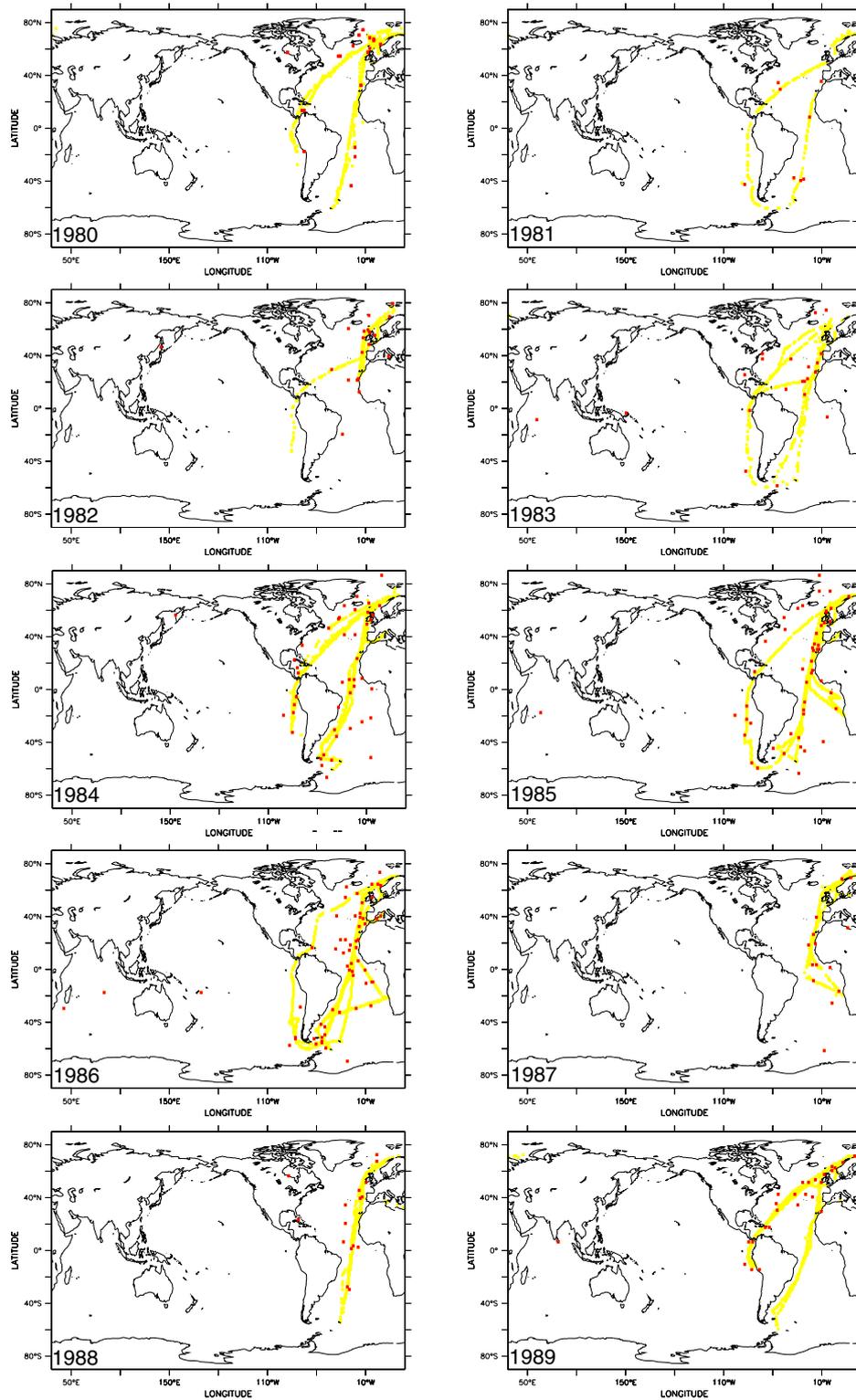


Figure 2 Example of Ship Tracking. Shown in yellow are the reports from the Karl Libknekht which passed the track check, shown in red are the reports that failed the track check.

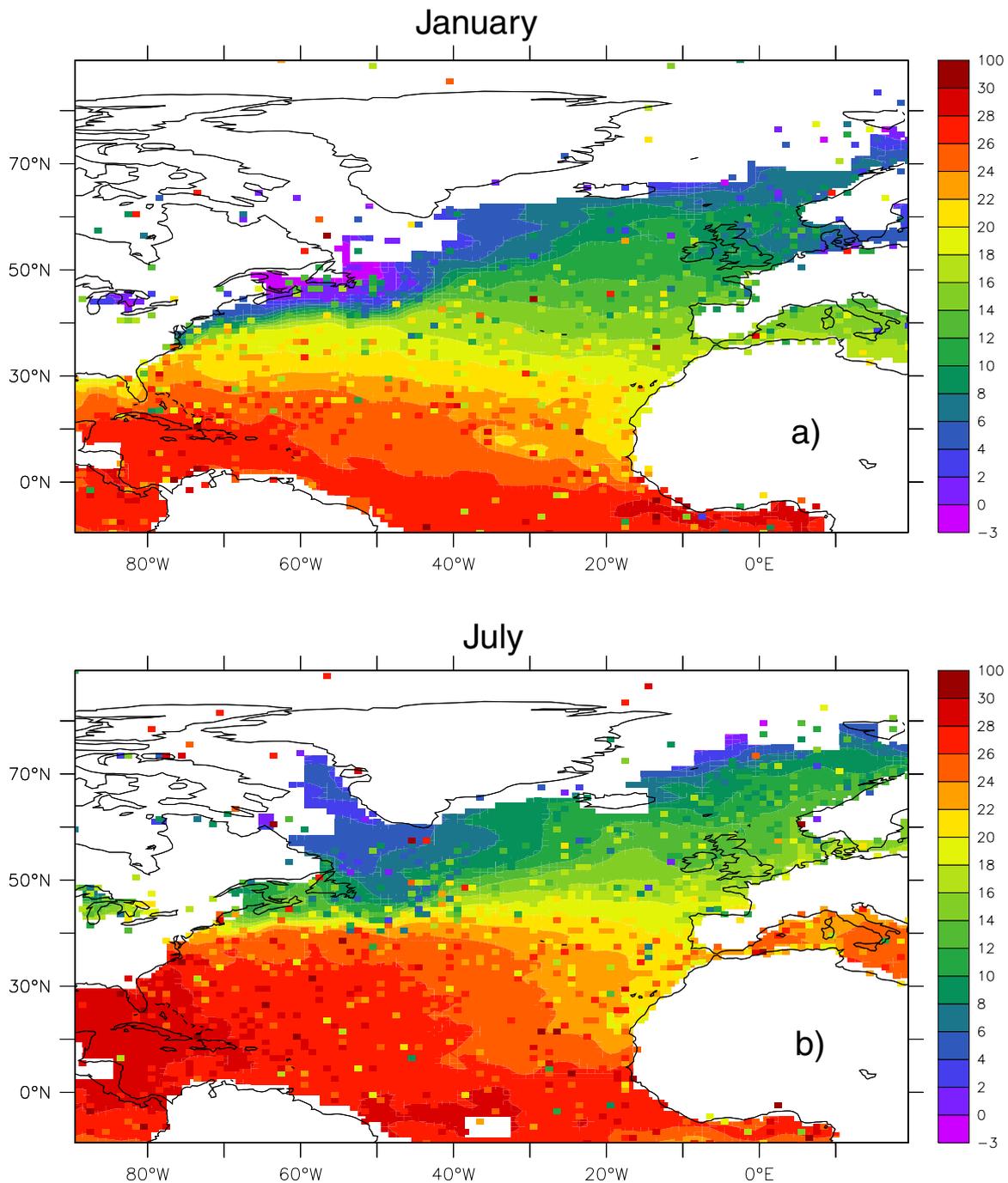


Figure 3 1° latitude-longitude average of SST reports removed by ship tracking overlaid on a smoothed field of SST reports passing the ship track test (units are °C). In many cases the SST is much different from the local smoothed SST and many of the reports removed are close to or on land or in the Arctic region. a) January 1990, b) July 1990.

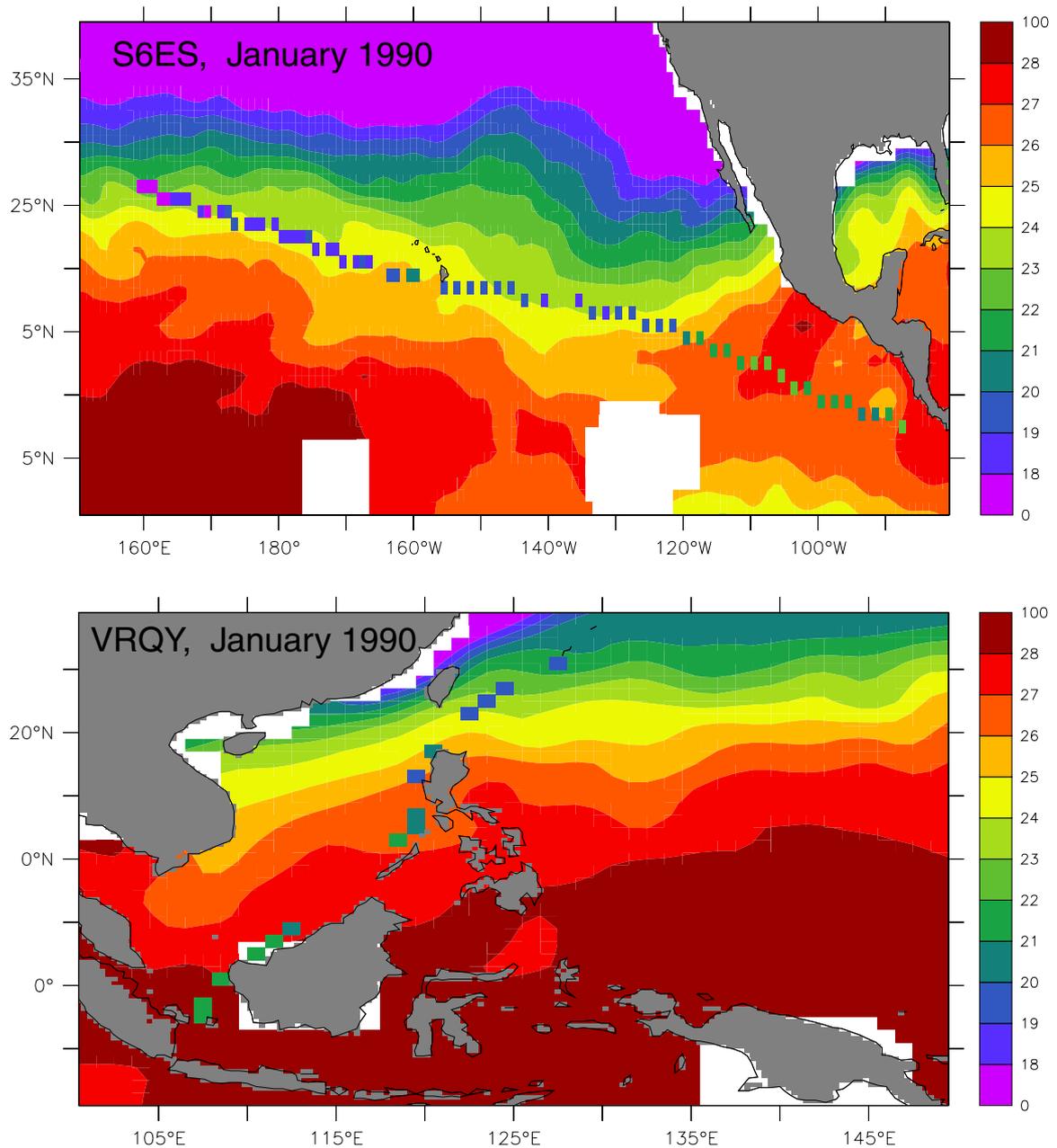


Figure 4 Example of ships identified by the blacklisting procedure in January 1990. The 1° latitude-longitude monthly mean SST ($^\circ\text{C}$) for the ship is overlaid on a smoothed monthly mean field. Each of the ships is reporting SST which appears to be significantly and consistently different from the surrounding field. Removing all the SST data from these and other blacklisted ships should improve the quality of the monthly mean field.

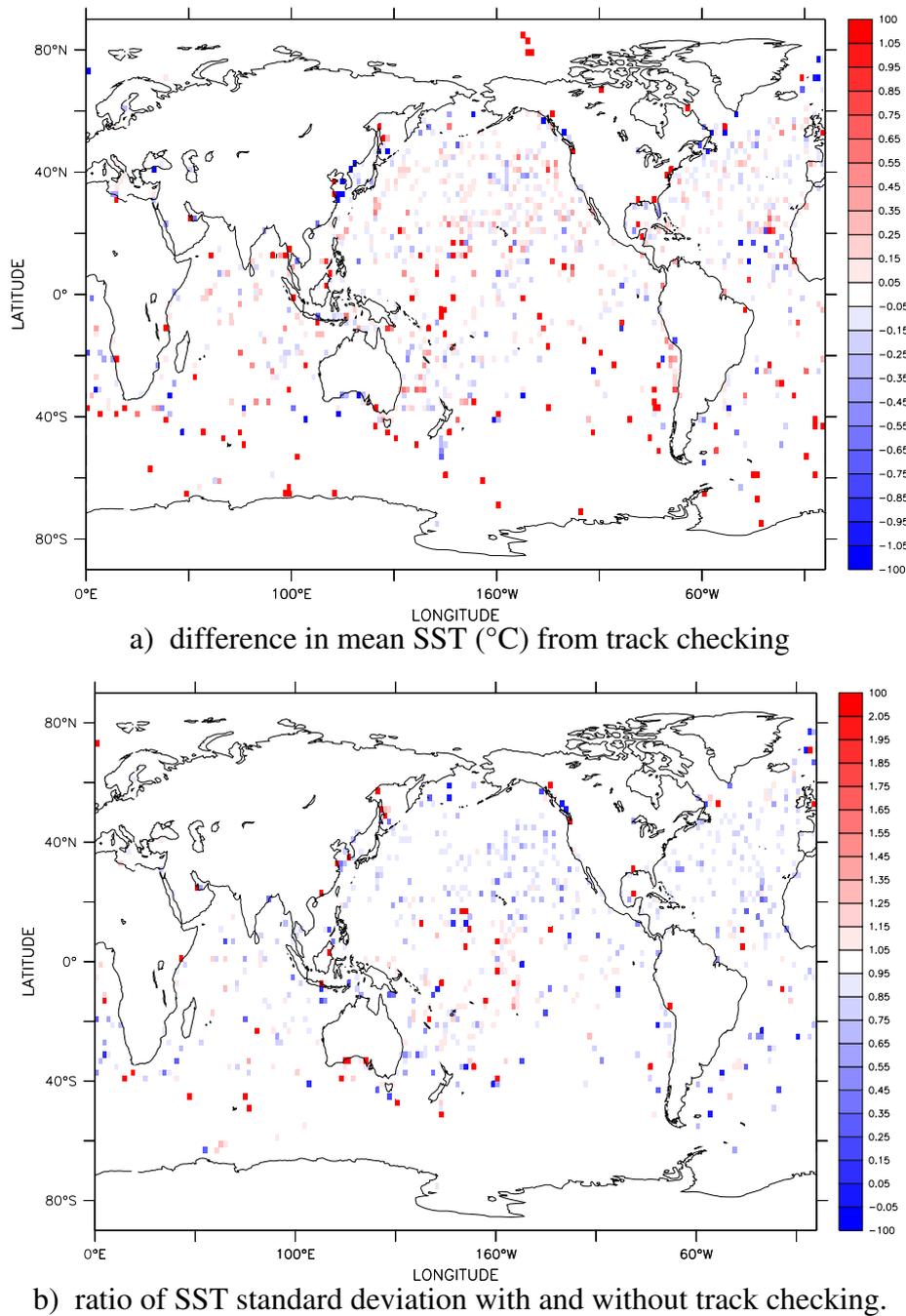


Figure 5 Difference in 1° latitude-longitude SST field following ship tracking for January 1980.

5a) Difference (with tracking - without tracking) in the mean SST field using the quality control flags (see Section 2.3). Note that quality control has been applied to all the data whether or not it failed the track check. 1° areas where all the data have been removed are shown as 99.

5b) shows the ratio of the SST standard deviation (with/without).

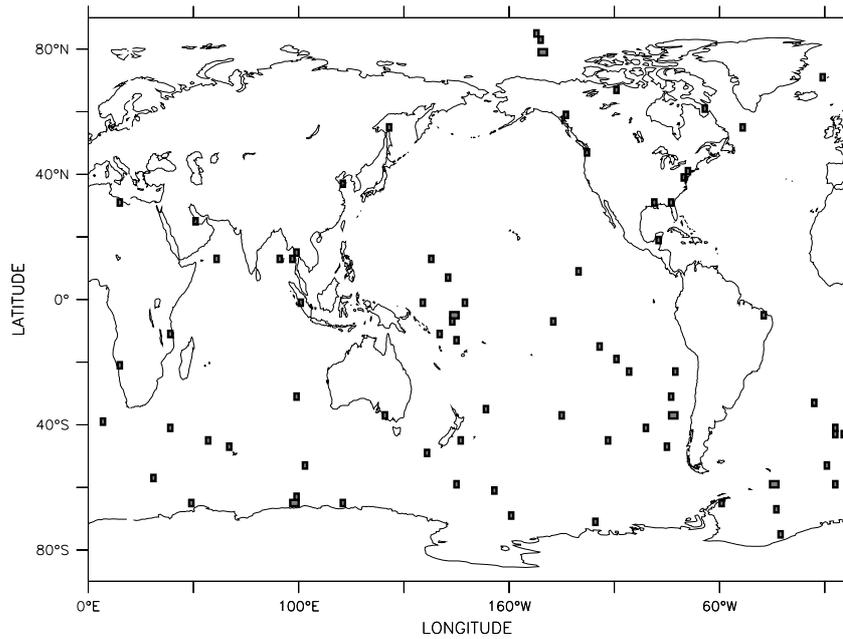


Figure 6a 1° regions in January 1980 where all the data within the square was removed by the track checking.

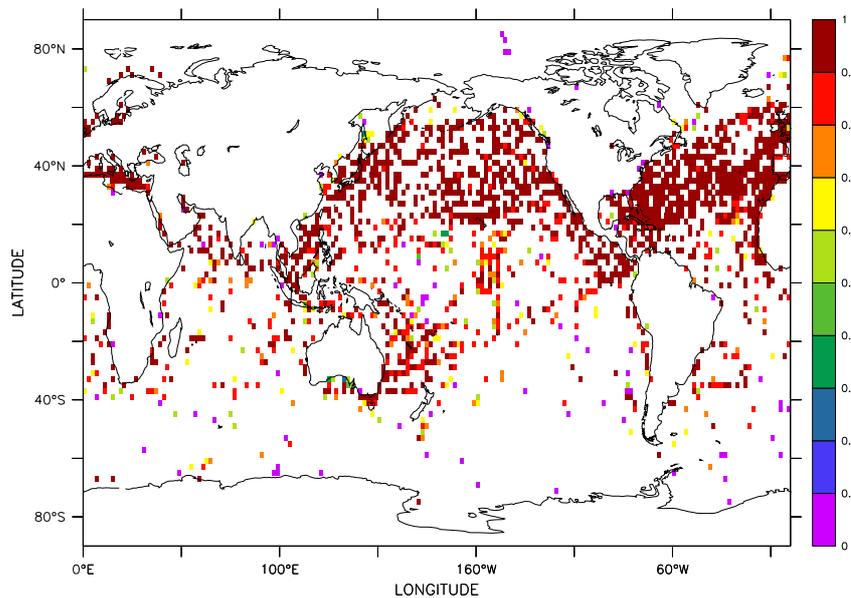


Figure 6b Proportion of data retained after the track checking within 1° squares for January 1980. Where no reports were removed the proportion has not been plotted. Note that the 1° areas shown in Figure 6a are plotted in this Figure in the lowest interval (0.0-0.1, coloured pink).

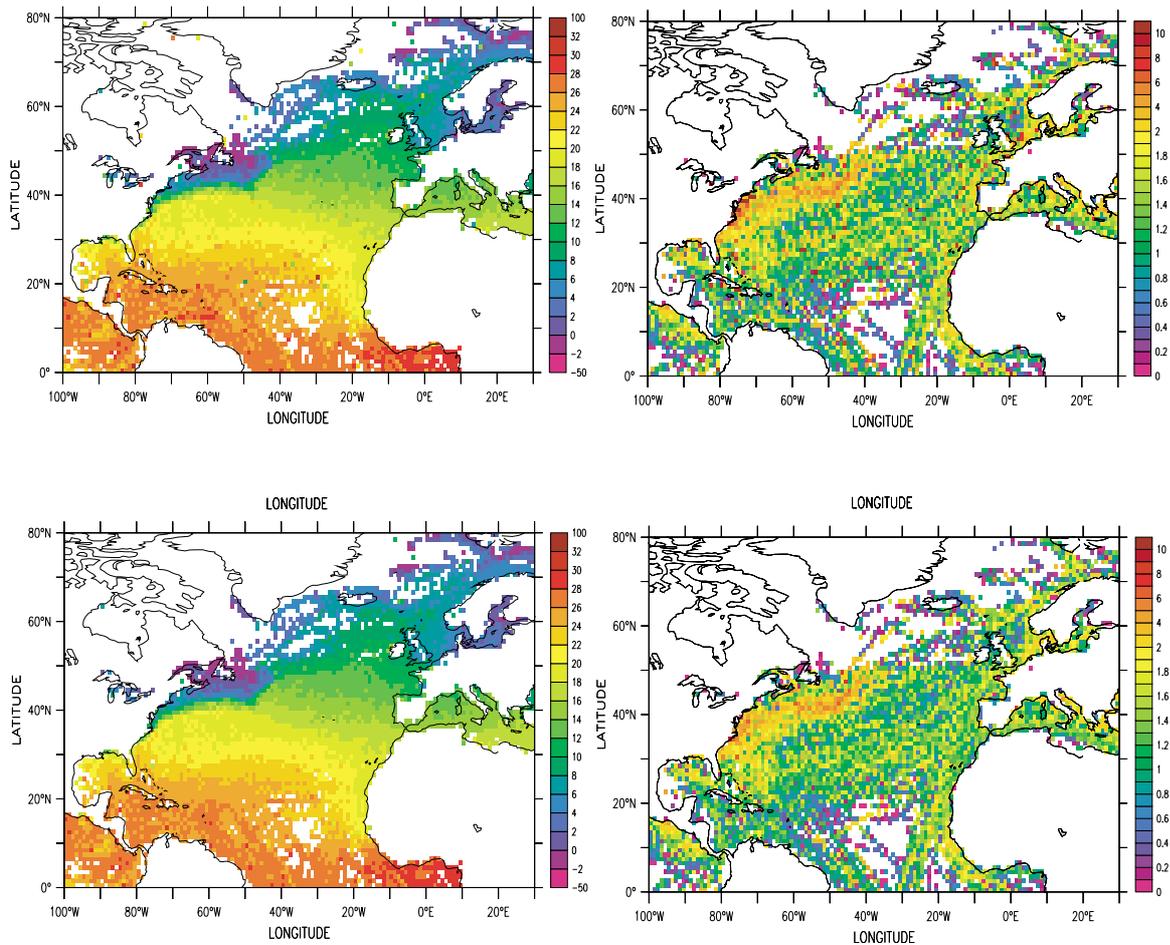


Figure 7 Mean and standard deviation of SST field in January 1985 before (upper plots) and after (lower plots) the quality control procedure outlined in Section 2.3.

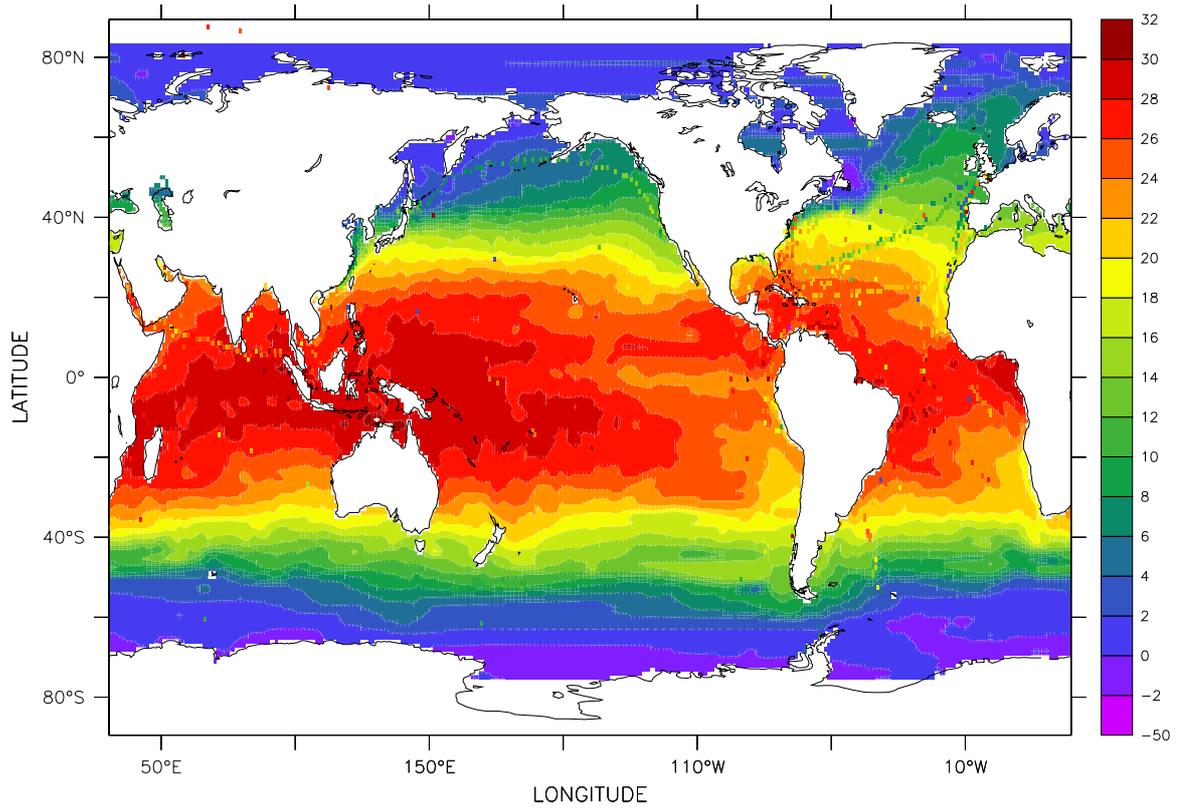


Figure 8 Plot of 1° latitude-longitude averages of SST data (°C) removed by blacklisting procedure plotted over a smoothed SST field for January 1985.

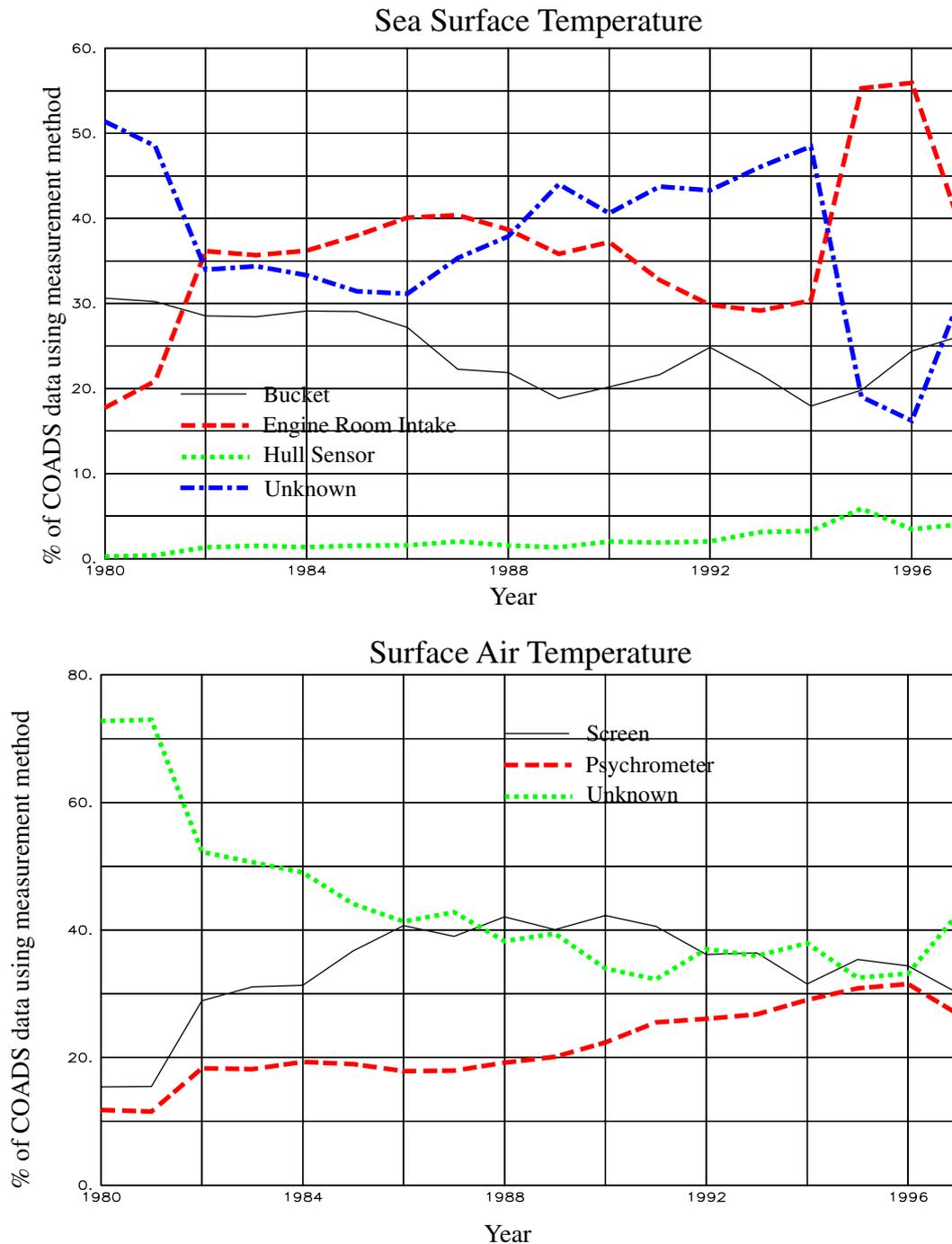
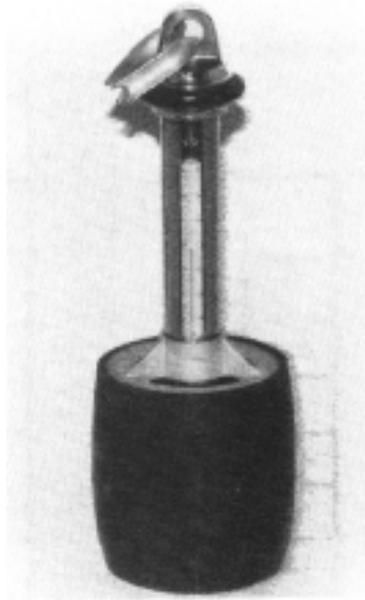


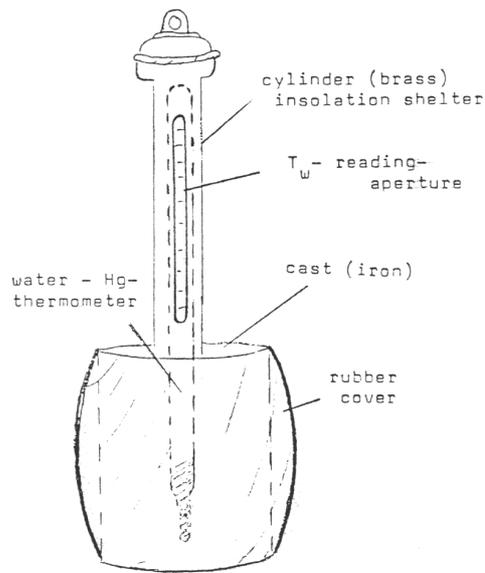
Figure 9 Percentages of VOS reports in April of each year of the dataset for which metadata could be associated
Top) SST measurement method from COADS and WMO 47 metadata combined.
Bottom) Air temperature measurement method from WMO 47 metadata.



Germany/France



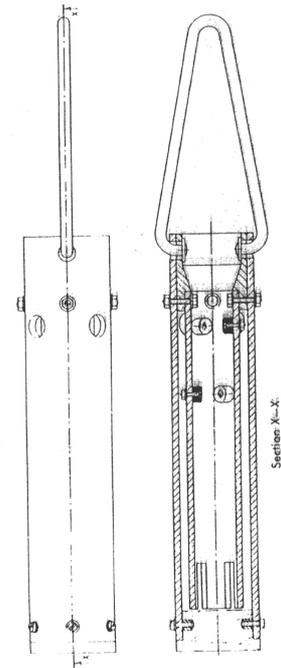
Netherlands



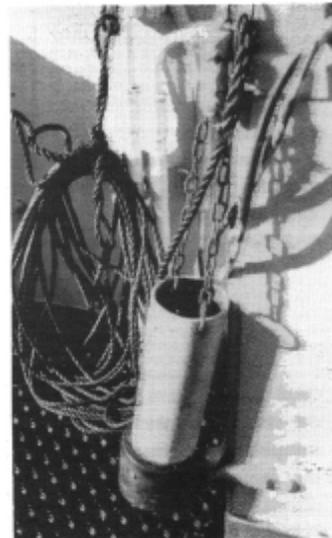
Germany/France



Netherlands



UK



Netherlands

Figure 10a Illustrations of some of the SST buckets in use on the ships participating in the VSOP-NA (from Kent and Taylor, 1991).



Figure 10b Buckets obtained at SOC for future analysis of heat loss characteristics.
From left to right: UK Met. Office bucket, Dutch Met. Office bucket,
German Met. Office bucket.

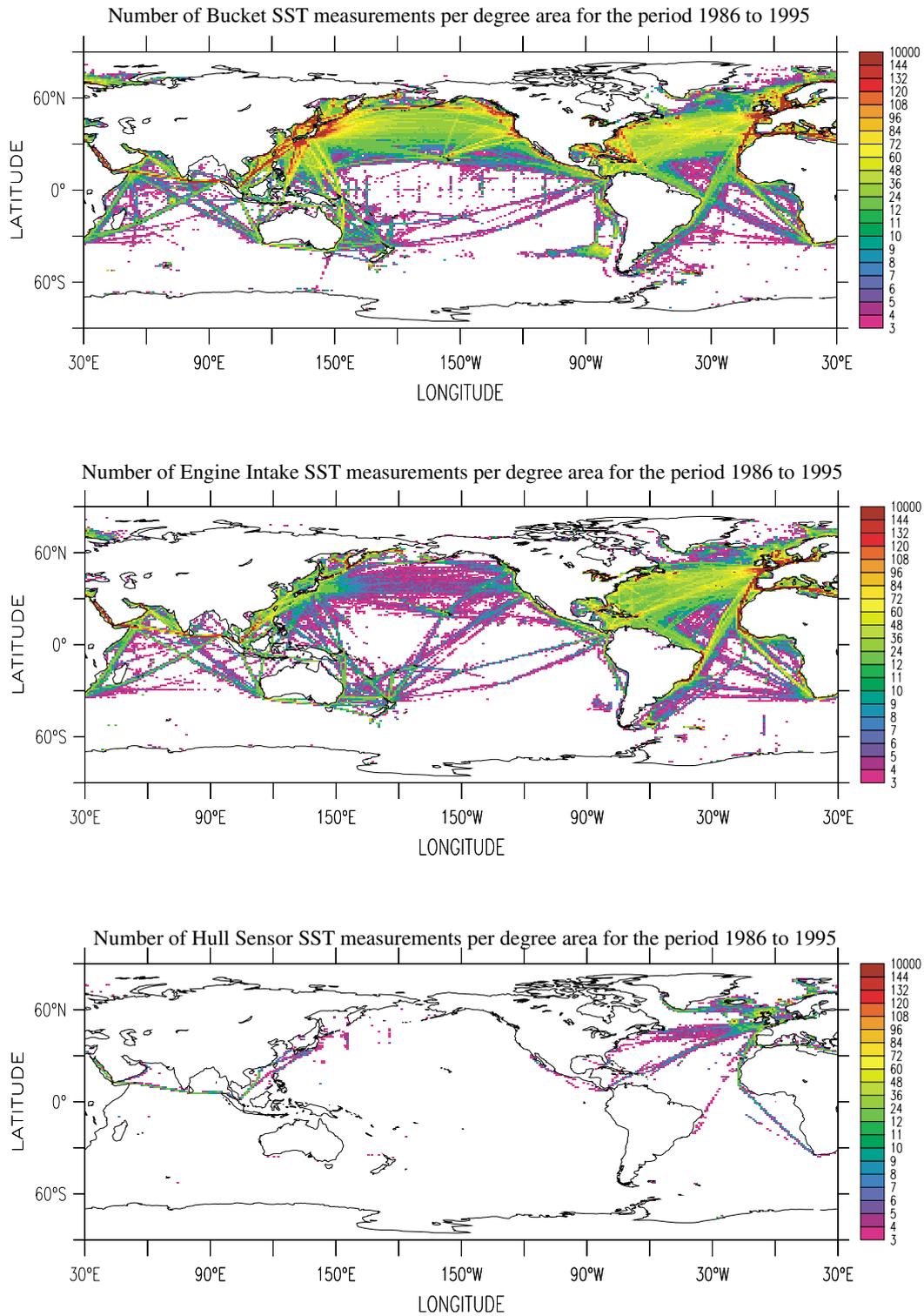


Figure 11 Number of SST observations per 1° area per year and by method for the period 1986 to 1995 for bucket measurements (top panel), engine intake measurements (centre panel) and hull sensors (lower panel).

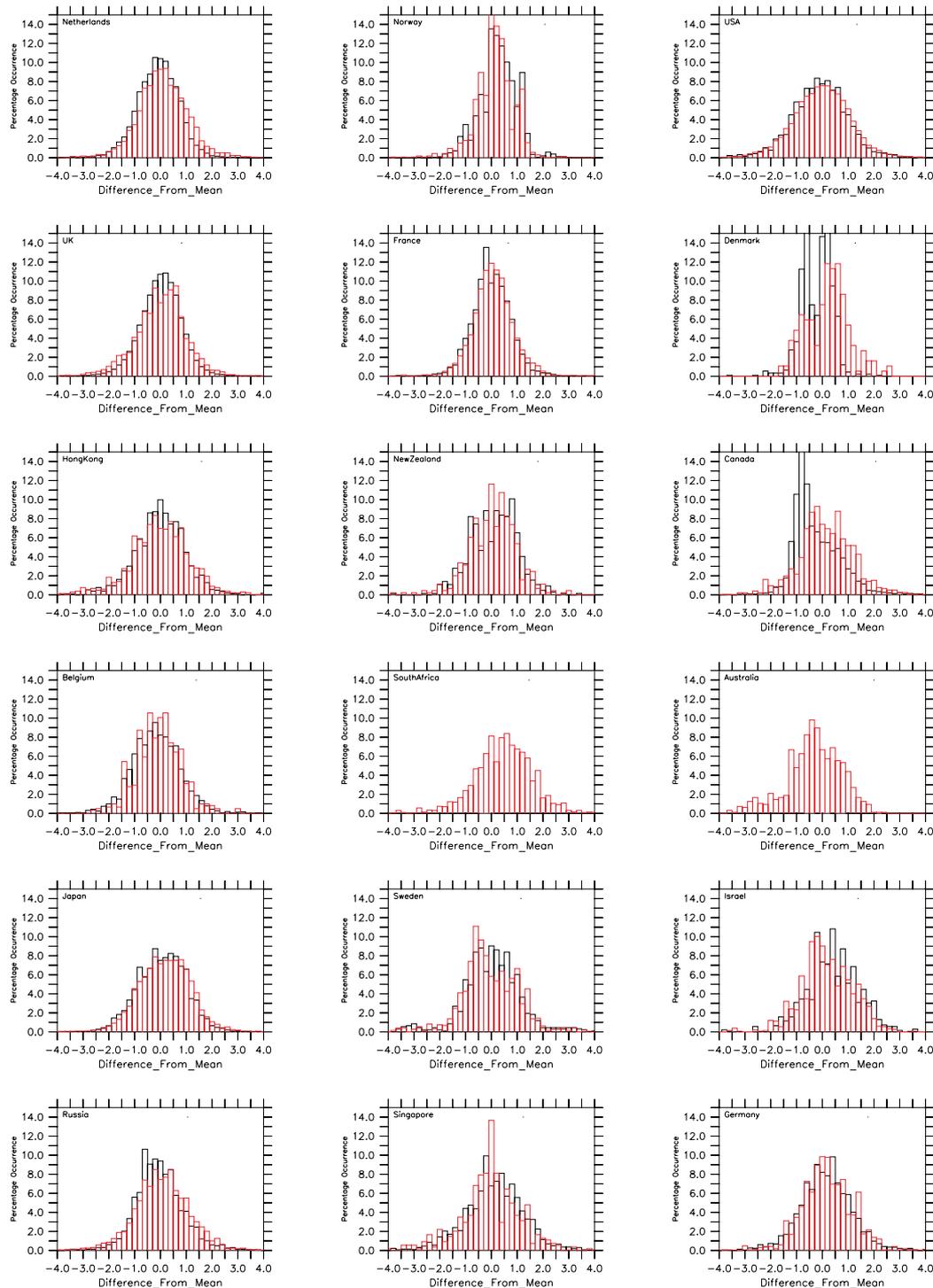


Figure 12a Histograms of deviations from local (10°) area mean of data points in the paired bucket (black) and engine intake (red) SST data file for January data from 1980 to 1997. Deviations are in units of local standard deviation and data are partitioned by recruiting country.

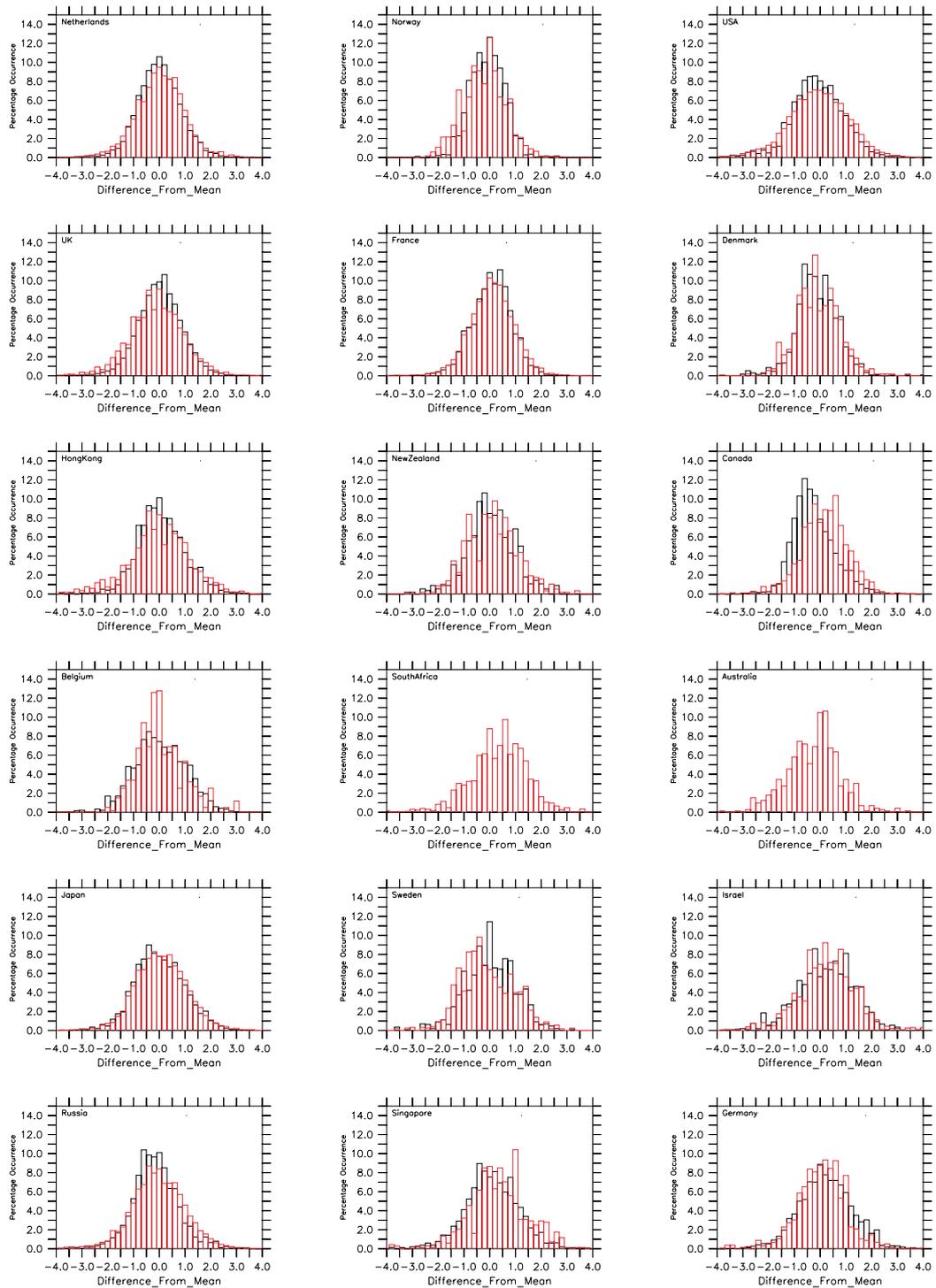


Figure 12b As 12a but for July.

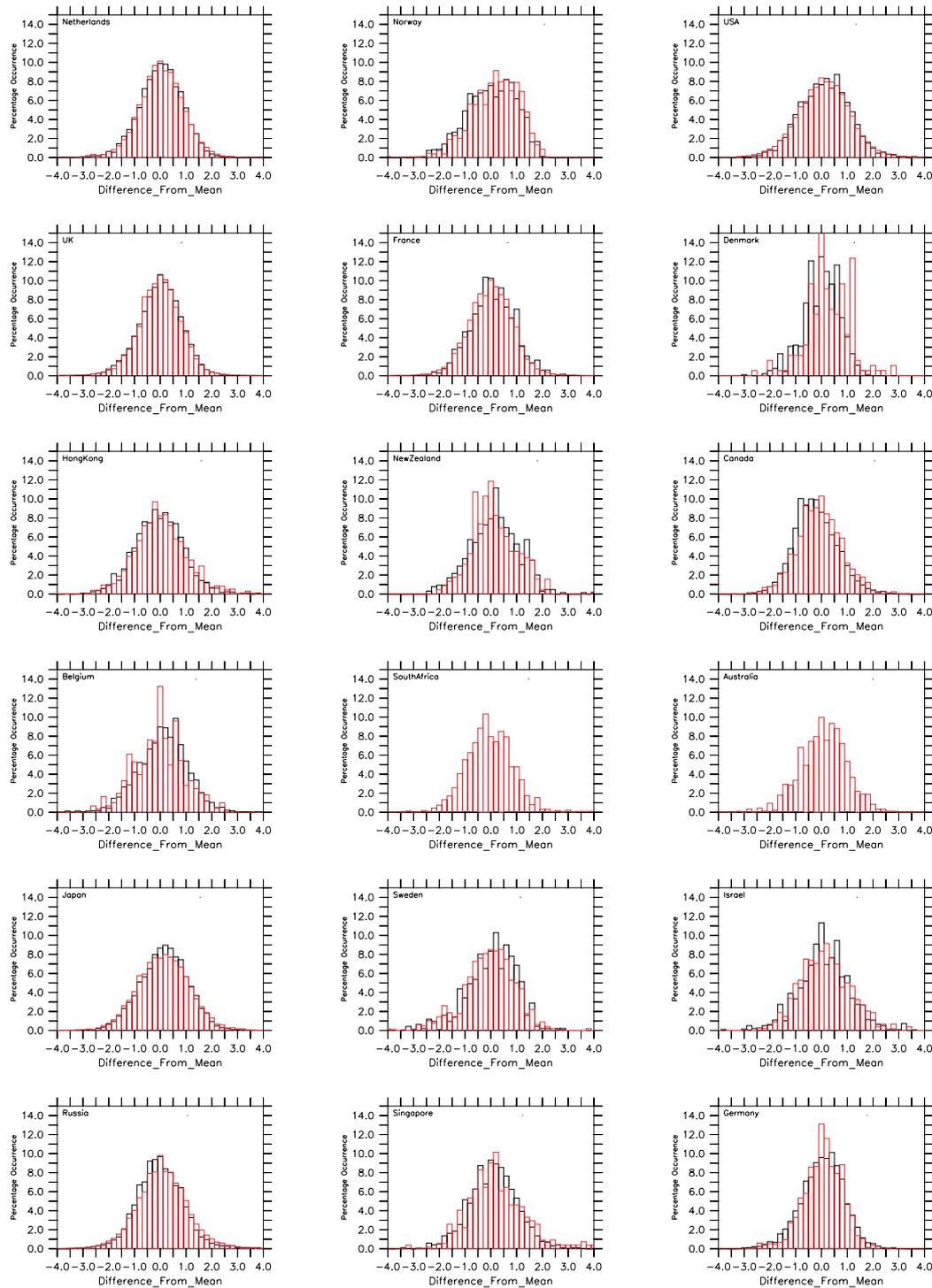


Figure 12c As Figure 12a but for January air temperatures reports associated with the SST data. Again red indicates data from ships that use buckets to measure the SST and black from those that use engine intakes to measure the SST. The method of air temperature measurement has not been taken into account.

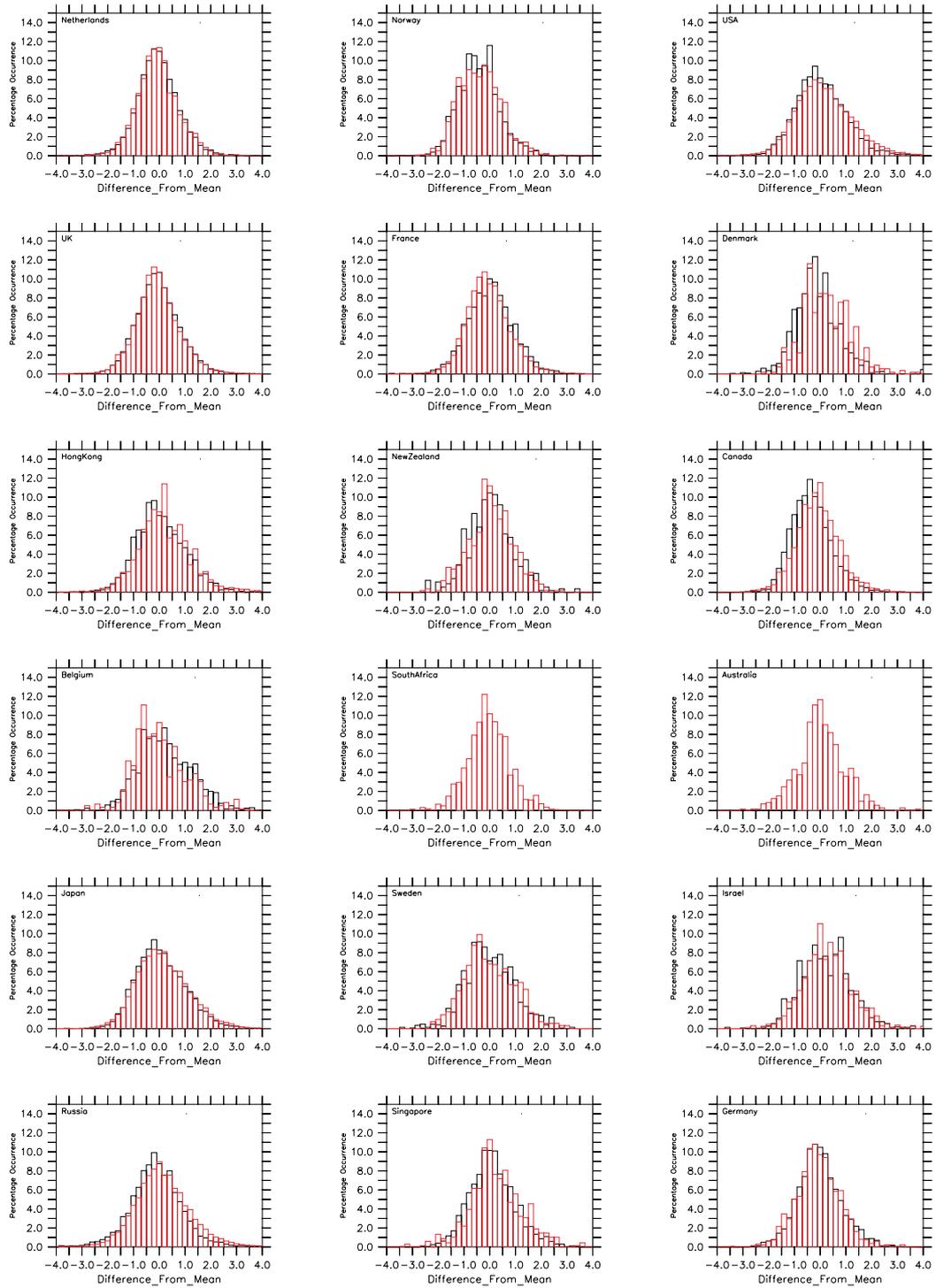
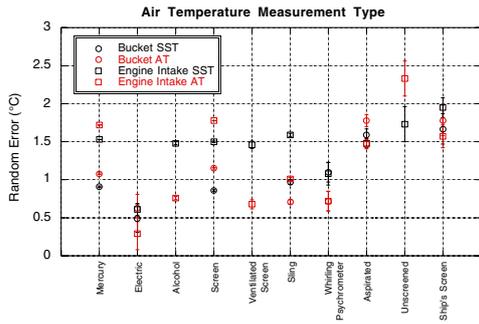
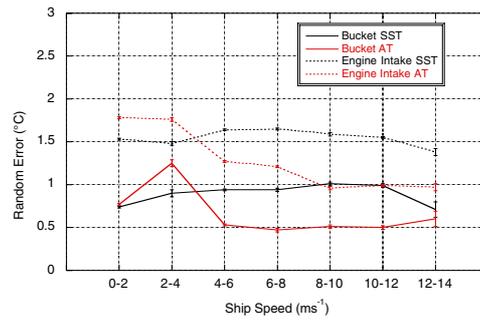


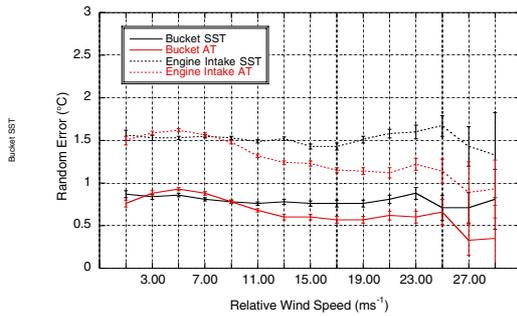
Figure 12d As Figure 12d but for July air temperatures.



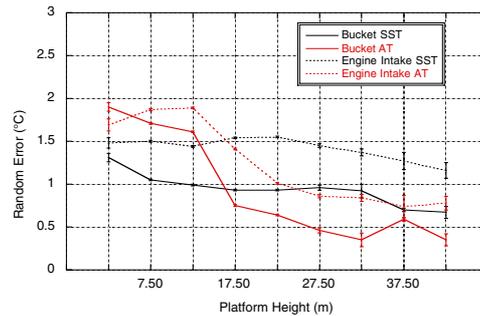
a) Error estimates by air temperature measurement type.



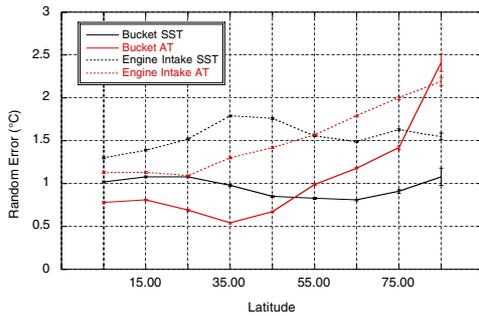
b) Error estimates by ship speed (ms^{-1}).



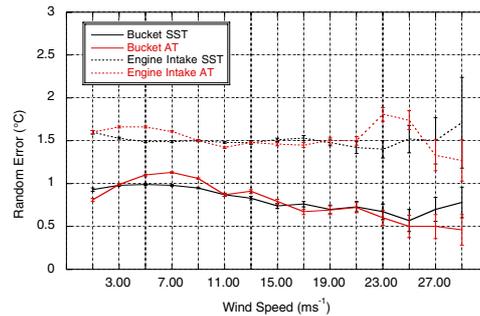
c) Error estimates by relative wind speed (ms^{-1}).



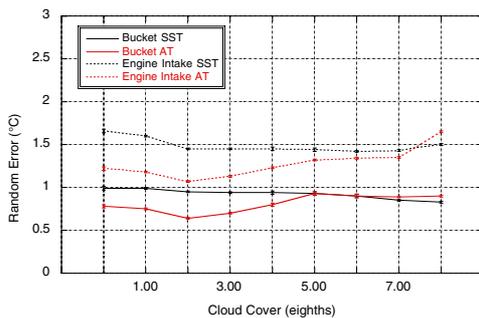
d) Error estimates by platform height (m).



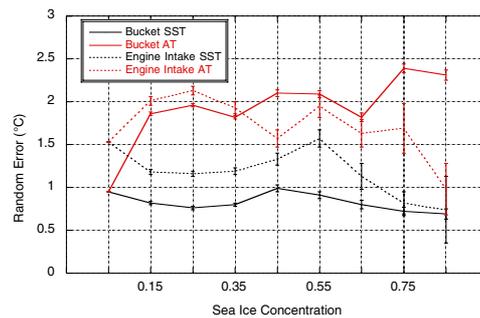
e) Error estimates by North Atlantic Latitude



f) Error estimates by wind speed (ms^{-1}).



g) Error estimates by cloud cover (eighths).



h) Error estimates by sea ice concentration (fraction).

Figure 13: Error estimates separately for ships that use buckets to measure the SST and those that report engine intake SST. Black lines are error estimates for SST and red lines for air temperature. Solid lines show bucket error estimates and dashed lines engine intake error estimates.

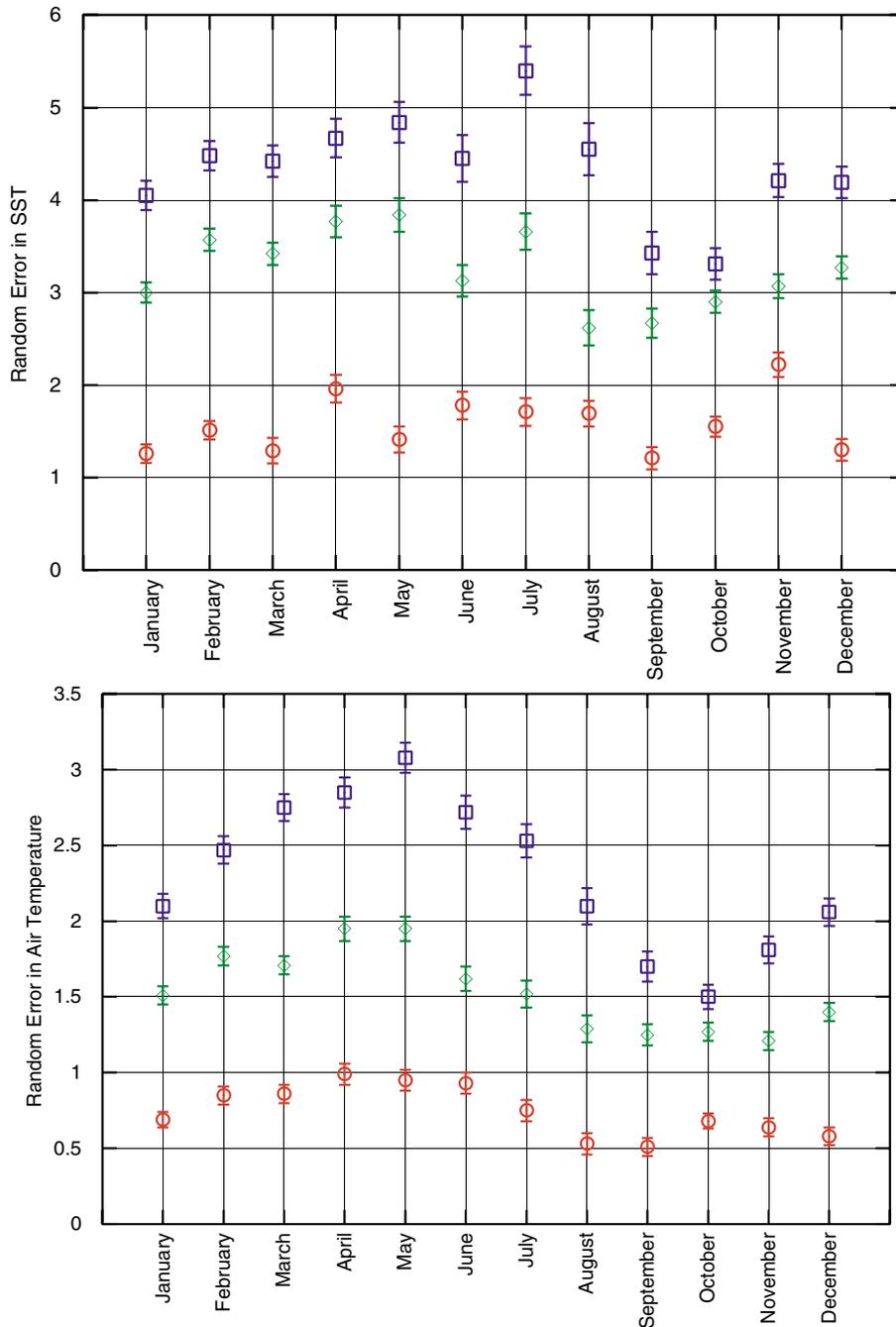


Figure 14 Seasonal variation in semivariogram intercepts used to generate the correlation matrix used to transform each monthly data file taken from Table 13. The upper plot shows SST error variances. Blue symbols are $2\sigma_{se}^2$, red symbols $2\sigma_{sb}^2$ and the green symbols $\sigma_{se}^2 + \sigma_{sb}^2$. The lower plot shows air temperature error variances. Blue symbols are $2\sigma_{ae}^2$, red symbols $2\sigma_{ab}^2$ and the green symbols $\sigma_{ae}^2 + \sigma_{ab}^2$.

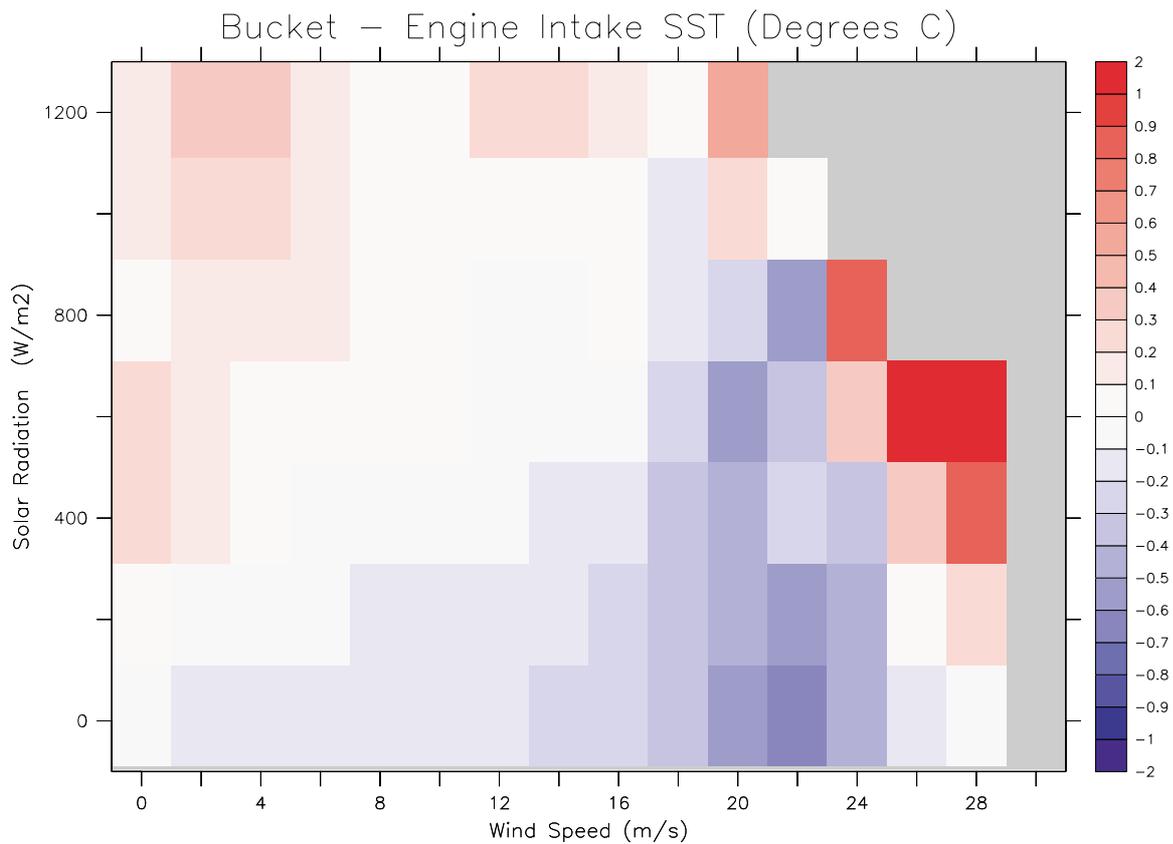


Figure 15 Bivariate plot of paired Bucket - Engine Intake SST as a function of wind speed and solar radiation averages for each report in the pair. Data are only plotted when the wind speeds from each method agree to within 2 ms^{-1} and the calculated solar radiation agrees to within 200 Wm^{-2} . Data are gridded in units of 2 ms^{-1} in wind speed and 200 Wm^{-2} in solar radiation using a Gaussian smoother with a range equal to the grid interval for each variable. In the grey region there are no data within a search radius of 0.5 ms^{-1} or 50 Wm^{-2} .

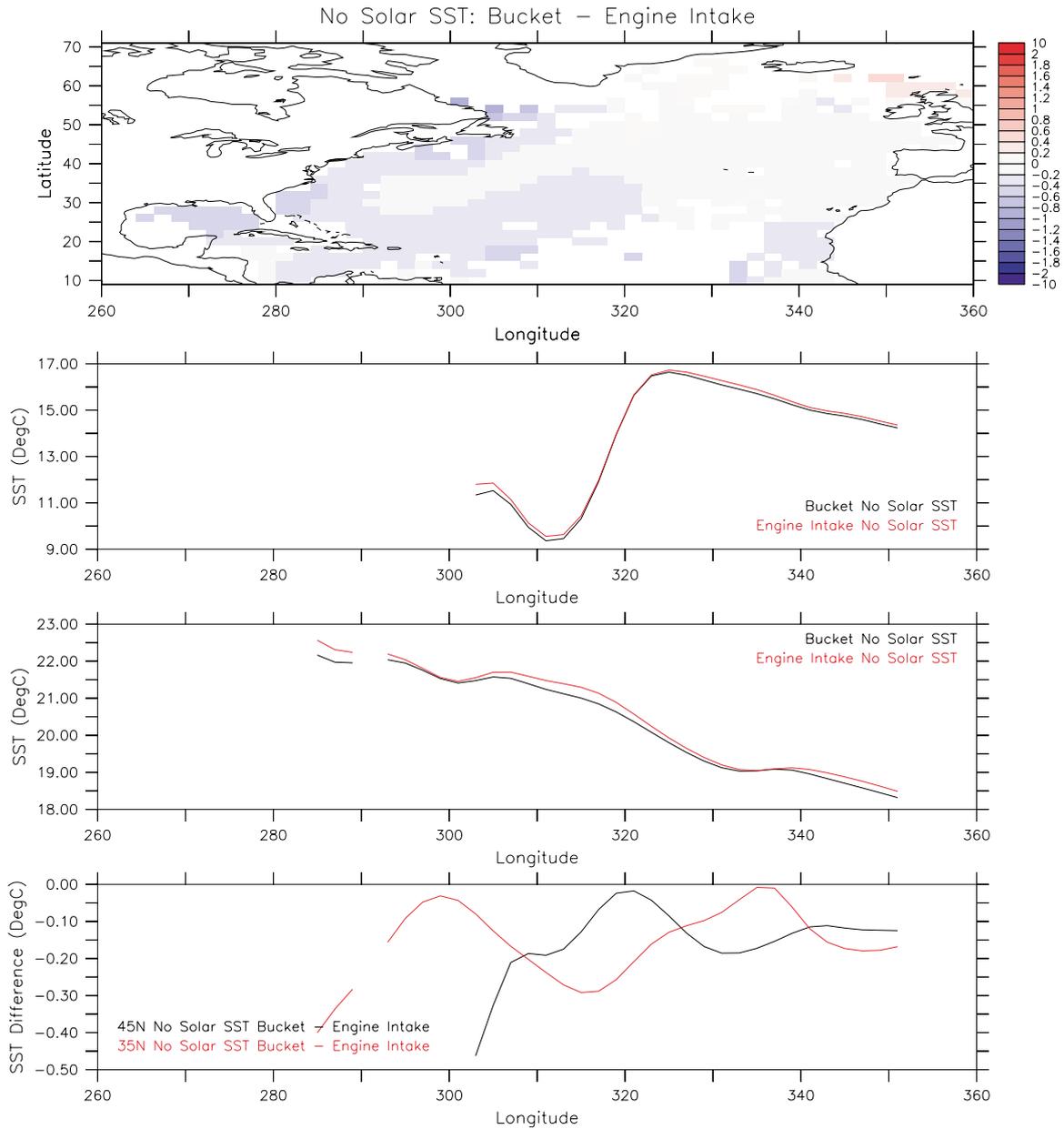


Figure 16 Difference between bucket and engine intake measured SST fields for conditions of very low ($< 1 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the SST fields (bucket SST - engine intake SST, top panel), the bucket SST (black) and engine intake SST (red) at 45°N (second panel) and the bucket SST (black) and engine intake SST (red) at 35°N (third panel) and the difference between the bucket and engine intake measured SST fields at 45°N and at 35°N (lower panel).

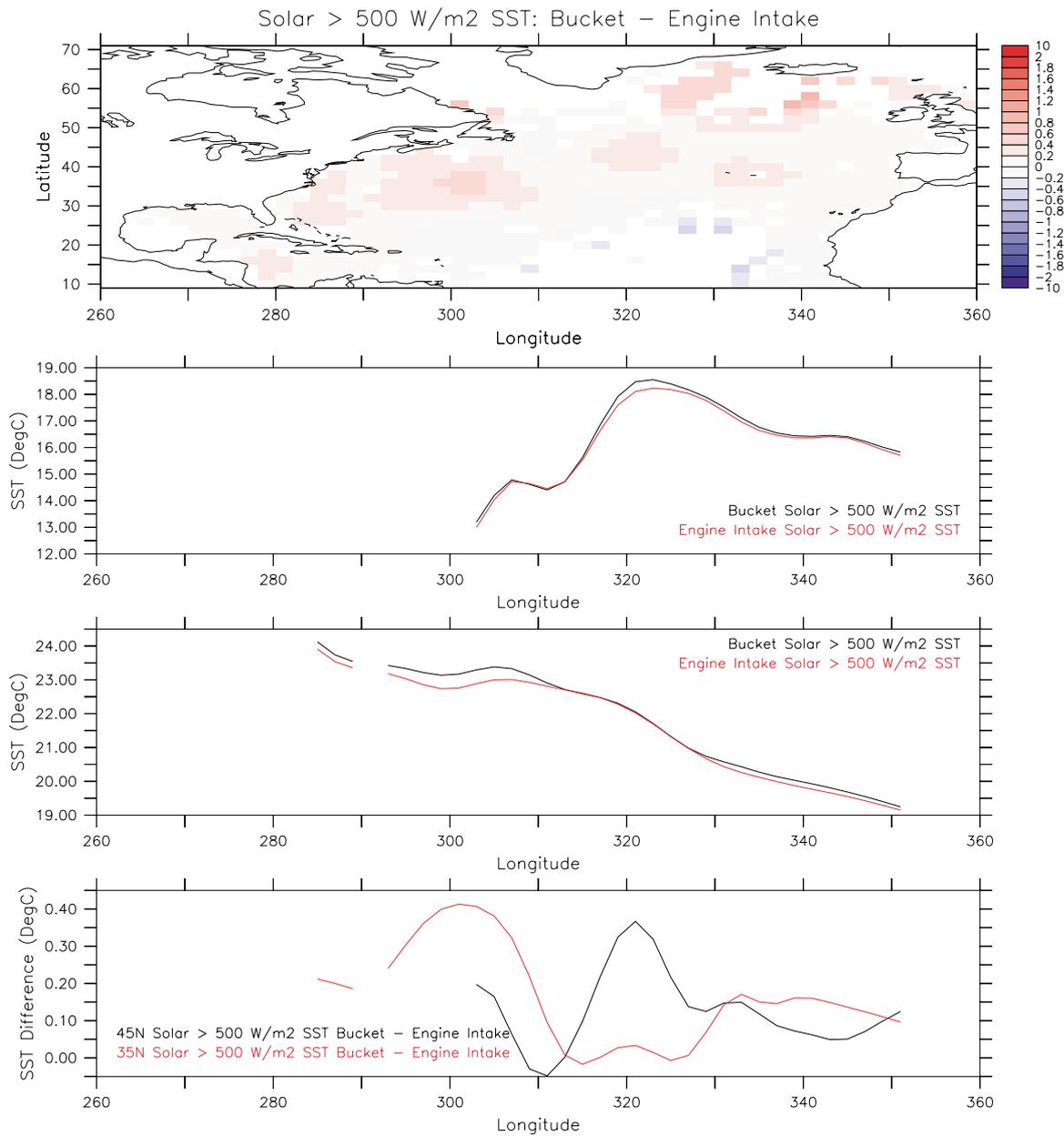


Figure 17 Difference between bucket and engine intake measured SST fields for conditions of high ($> 500 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the SST fields (bucket SST - engine intake SST, top panel), the bucket SST (black) and engine intake SST (red) at 45°N (second panel) and the bucket SST (black) and engine intake SST (red) at 35°N (third panel) and the difference between the bucket and engine intake measured SST fields at 45°N and at 35°N (lower panel).

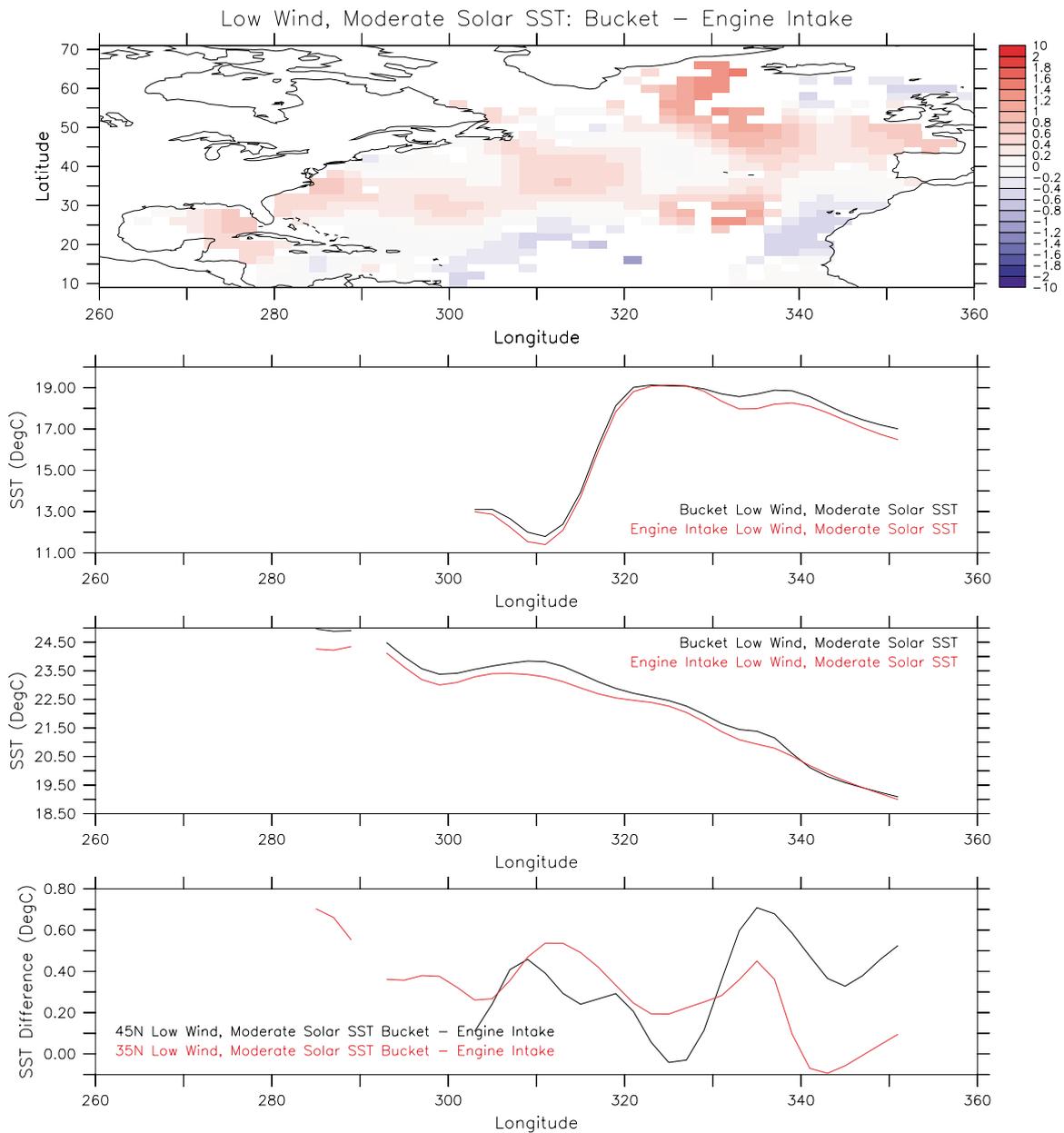


Figure 18 Difference between bucket and engine intake measured SST fields for conditions of low wind speed ($< 3 \text{ ms}^{-1}$) and moderate and high ($> 300 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the SST fields (bucket SST - engine intake SST, top panel), the bucket SST (black) and engine intake SST (red) at 45°N (second panel) and the bucket SST (black) and engine intake SST (red) at 35°N (third panel) and the difference between the bucket and engine intake measured SST fields at 45°N and at 35°N (lower panel).

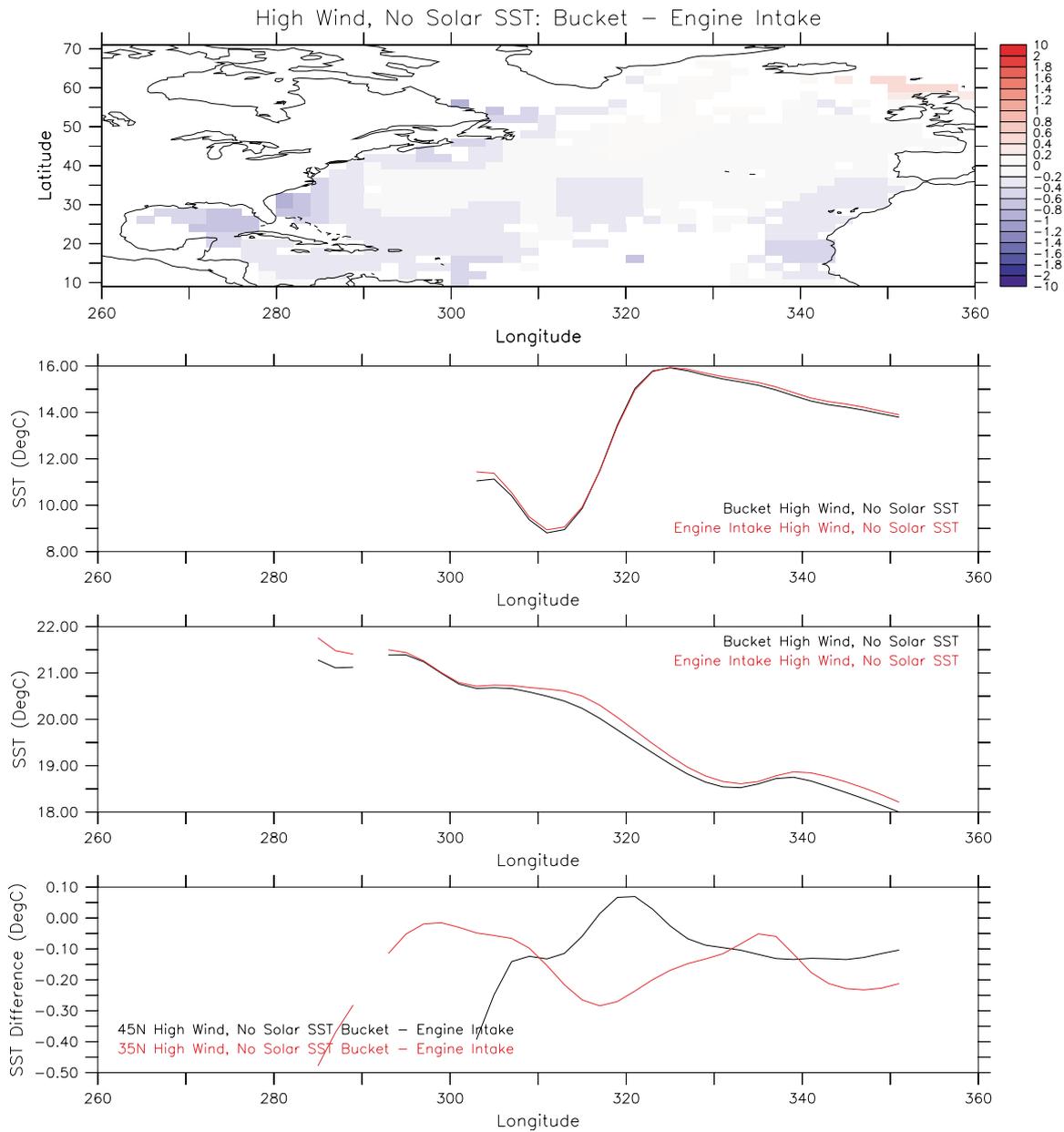


Figure 19 Difference between bucket and engine intake measured SST fields for conditions of high wind speed ($> 6 \text{ ms}^{-1}$) and very low ($< 1 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the SST fields (bucket SST - engine intake SST, top panel), the bucket SST (black) and engine intake SST (red) at 45°N (second panel) and the bucket SST (black) and engine intake SST (red) at 35°N (third panel) and the difference between the bucket and engine intake measured SST fields at 45°N and at 35°N (lower panel).

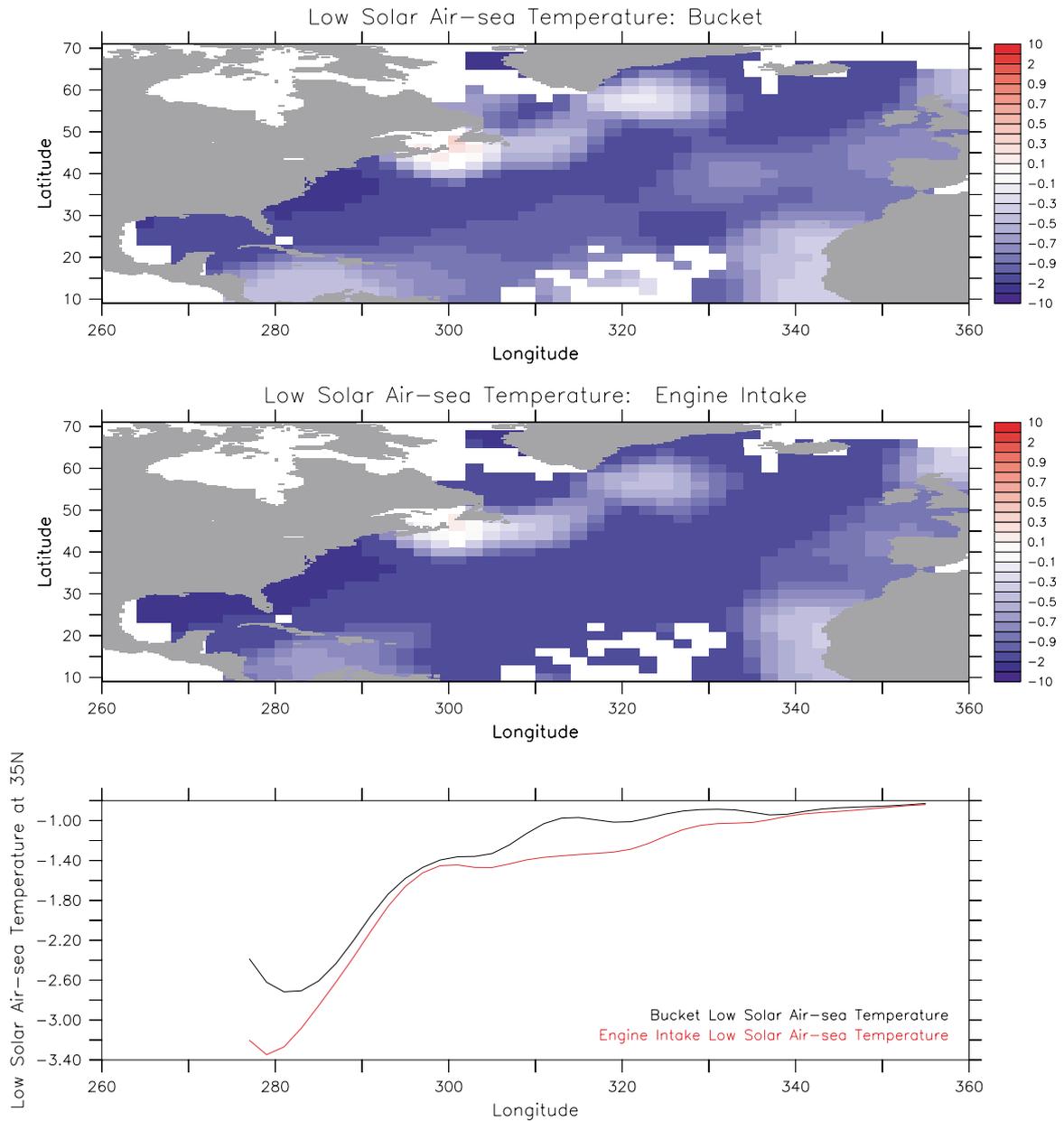


Figure 20 Top panel: Air-sea temperature difference for reports with bucket SST for conditions of low solar radiation (<math>< 1 \text{ Wm}^{-2}</math>).

Centre panel: Air-sea temperature difference for reports with engine intake SST for conditions of low solar radiation (<math>< 1 \text{ Wm}^{-2}</math>).

Lower panel: Air-sea temperature difference at 35°N for reports with bucket SST (black) and for reports with engine intake SST (red) for conditions of low solar radiation (<math>< 1 \text{ Wm}^{-2}</math>).

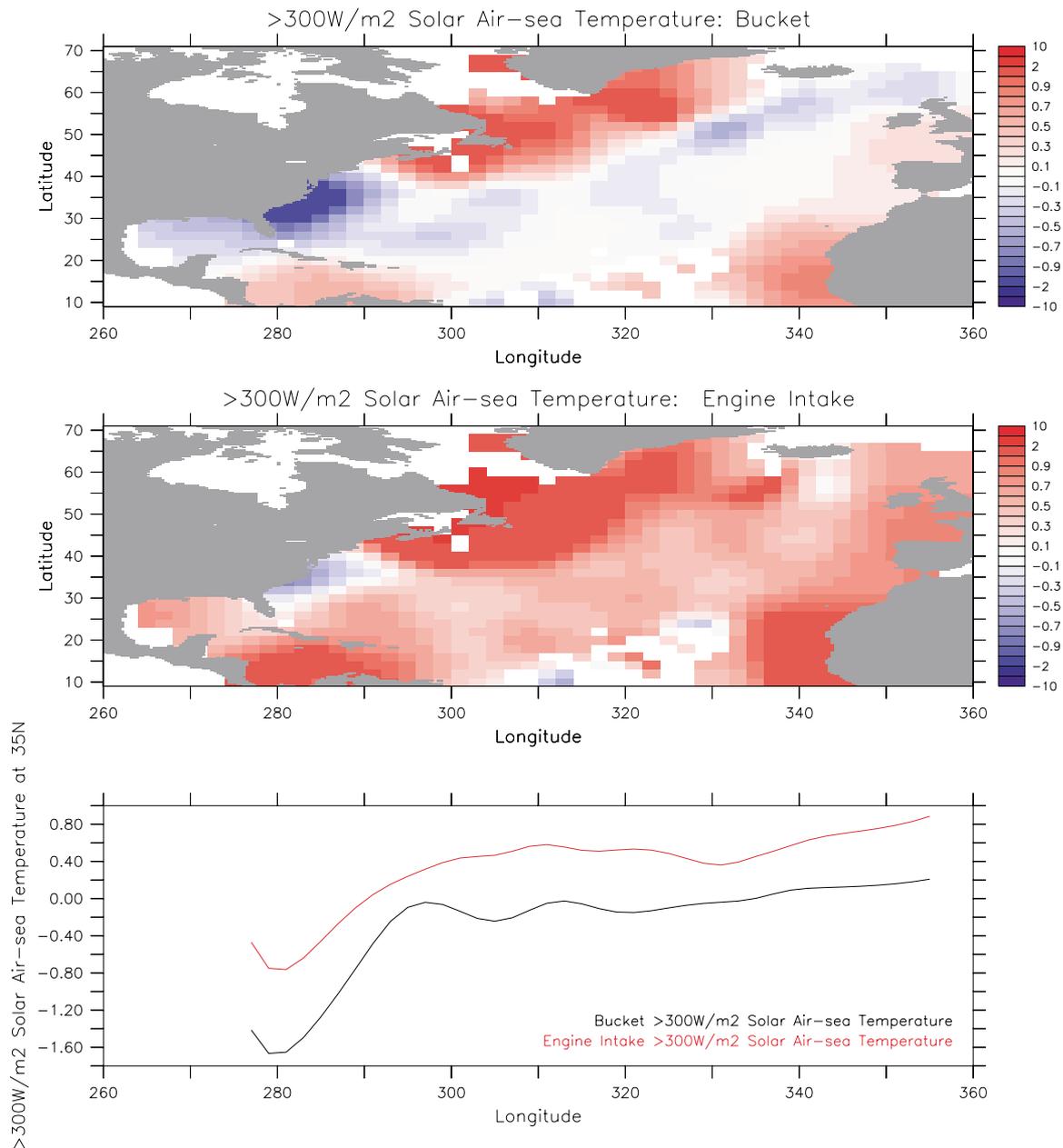


Figure 21 Top panel: Air-sea temperature difference for reports with bucket SST for conditions of moderate to high solar radiation ($>300 \text{ Wm}^{-2}$).

Centre panel: Air-sea temperature difference for reports with engine intake SST for conditions of moderate to high solar radiation ($>300 \text{ Wm}^{-2}$).

Lower panel: Air-sea temperature difference at 35°N for reports with bucket SST (black) and for reports with engine intake SST (red) for conditions of moderate to high solar radiation ($>300 \text{ Wm}^{-2}$).

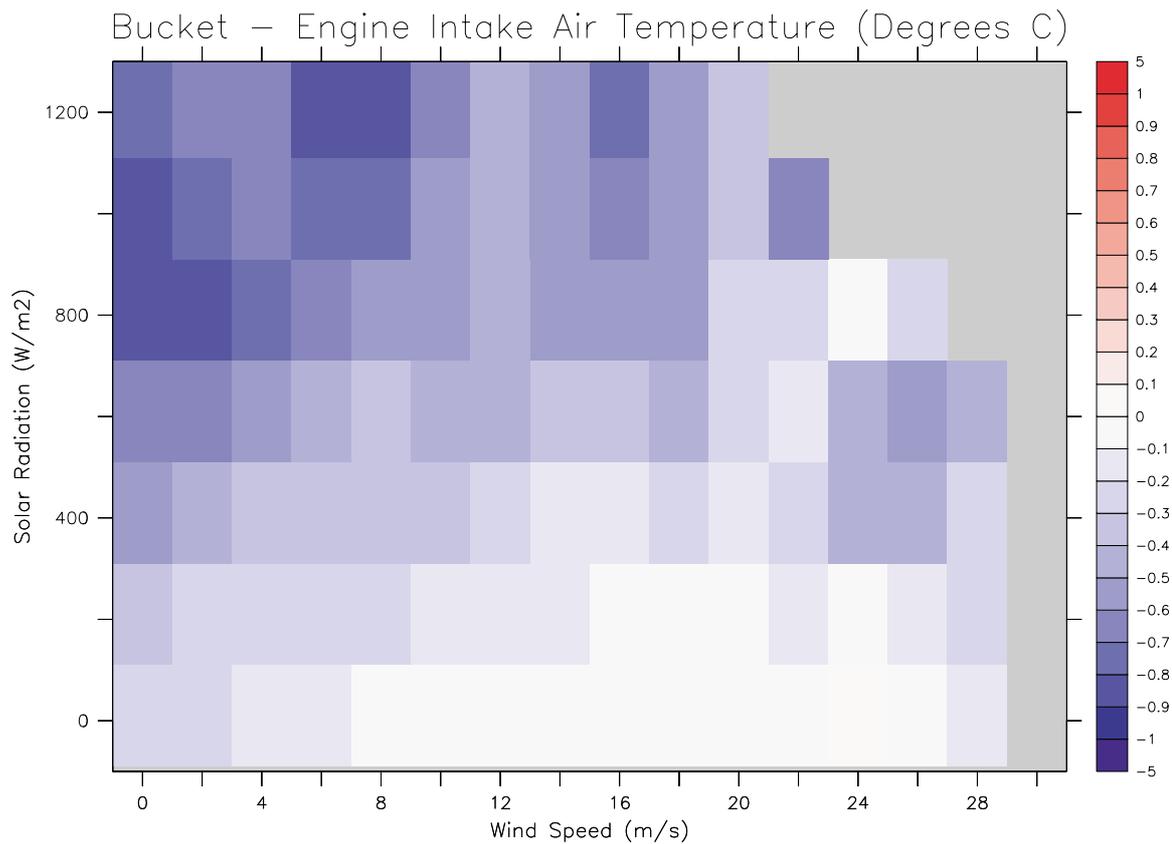


Figure 22 Bivariate plot of paired Bucket - Engine Intake report Air Temperature as a function of wind speed and solar radiation averages for each report in the pair. Data are only plotted when the wind speeds from each method agree to within 2 ms^{-1} and the calculated solar radiation agrees to within 200 Wm^{-2} . Data are gridded in units of 2 ms^{-1} in wind speed and 200 Wm^{-2} in solar radiation using a Gaussian smoother with a range equal to the grid interval for each variable. In the grey region there are no data within a search radius of 0.5 ms^{-1} or 50 Wm^{-2} .

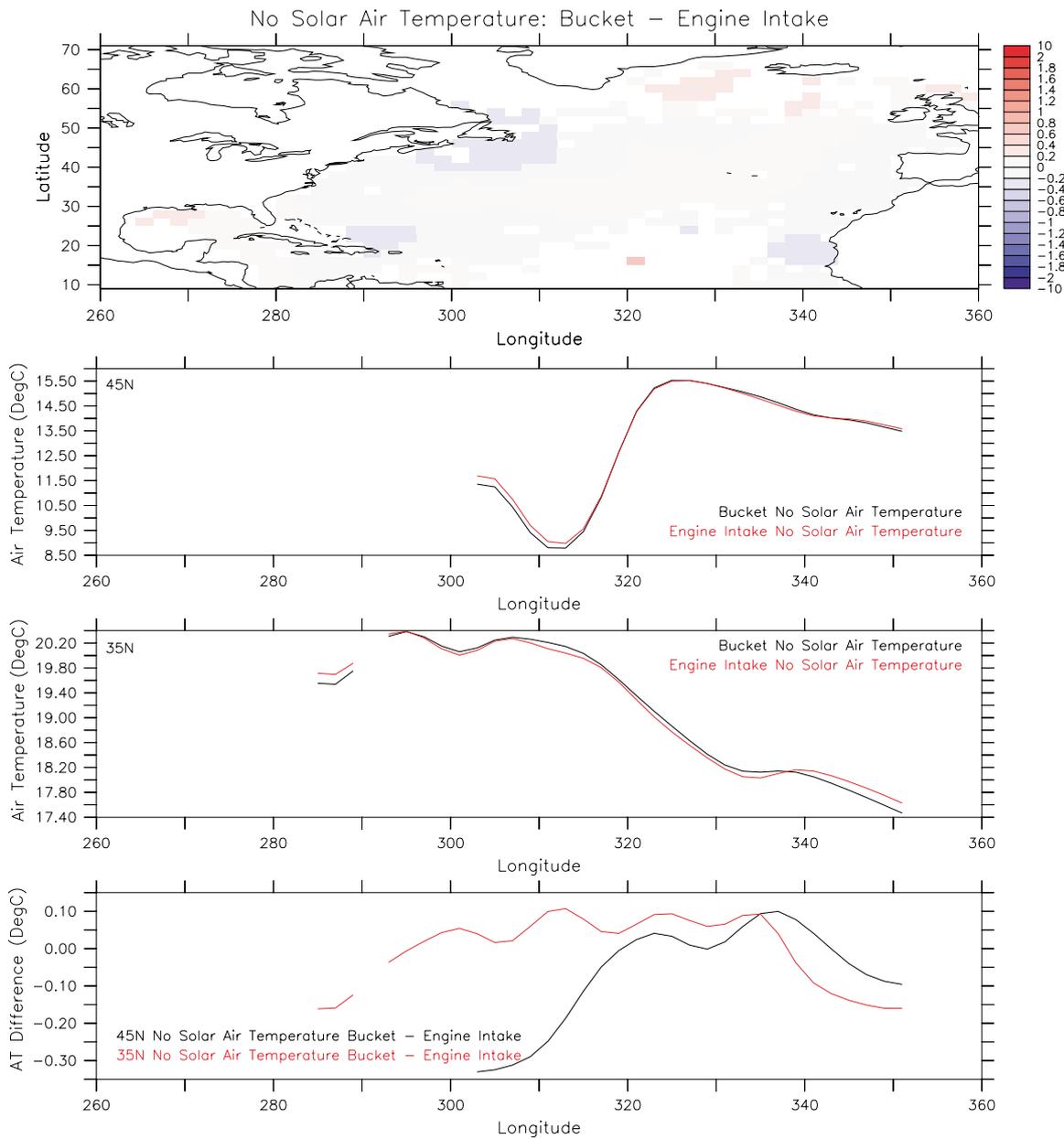


Figure 23 Difference between the air temperature from bucket and engine intake measured reports for conditions of very low ($< 1 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the air temperature fields (bucket report - engine intake report, top panel), the air temperature from the bucket report (black) and engine intake report (red) at 45°N (second panel) and at 35°N (third panel) and the difference between the air temperatures for the different methods at 45°N and 35°N (lower panel).

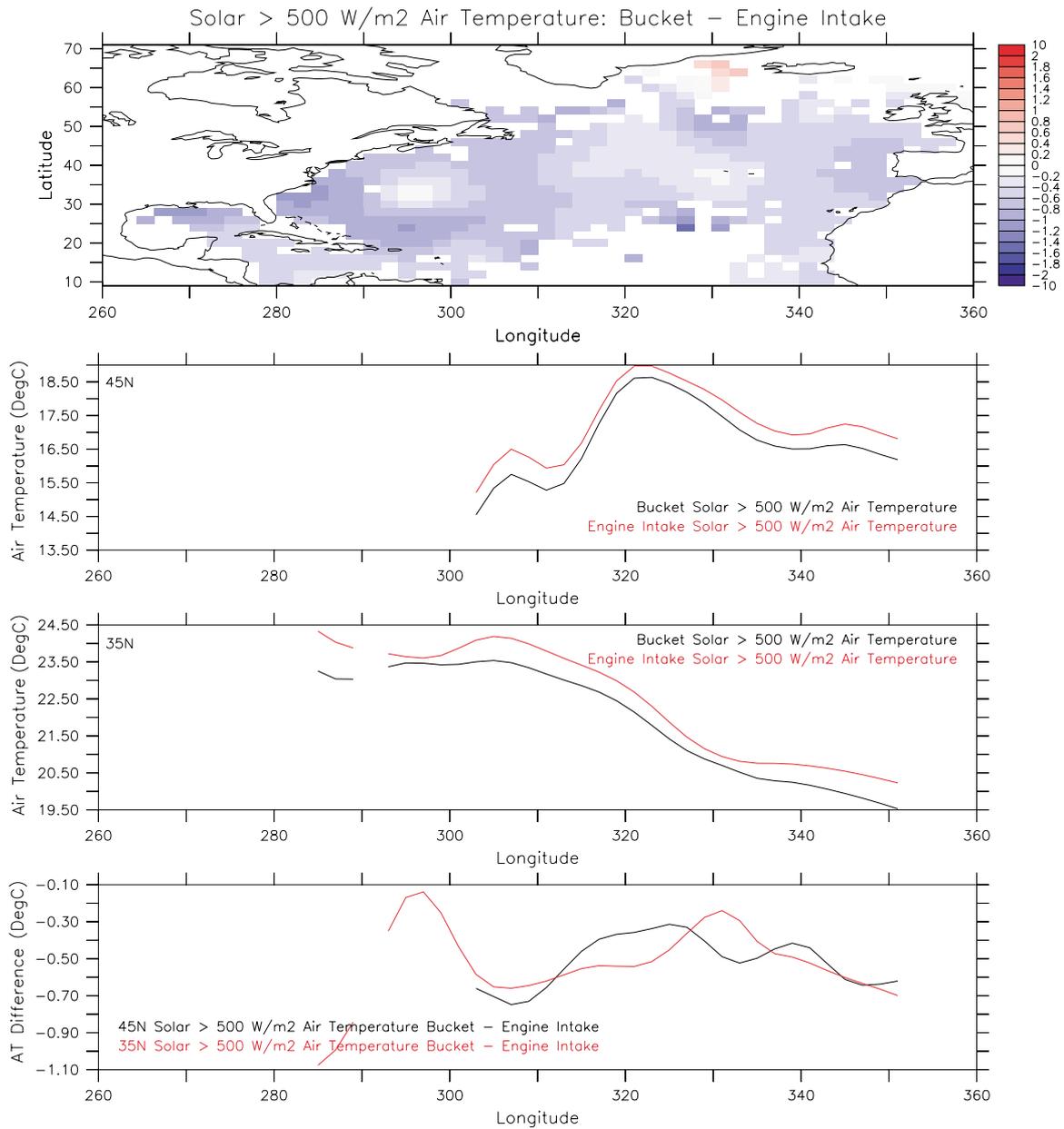


Figure 24 Difference between the air temperature from bucket and engine intake measured reports for conditions of high ($> 500 \text{ Wm}^{-2}$) solar radiation. Panels show the difference between the air temperature fields (bucket report - engine intake report, top panel), the air temperature from the bucket report (black) and engine intake report (red) at 45°N (second panel) and at 35°N (third panel) and the difference between the air temperatures for the different methods at 45°N and 35°N (lower panel).

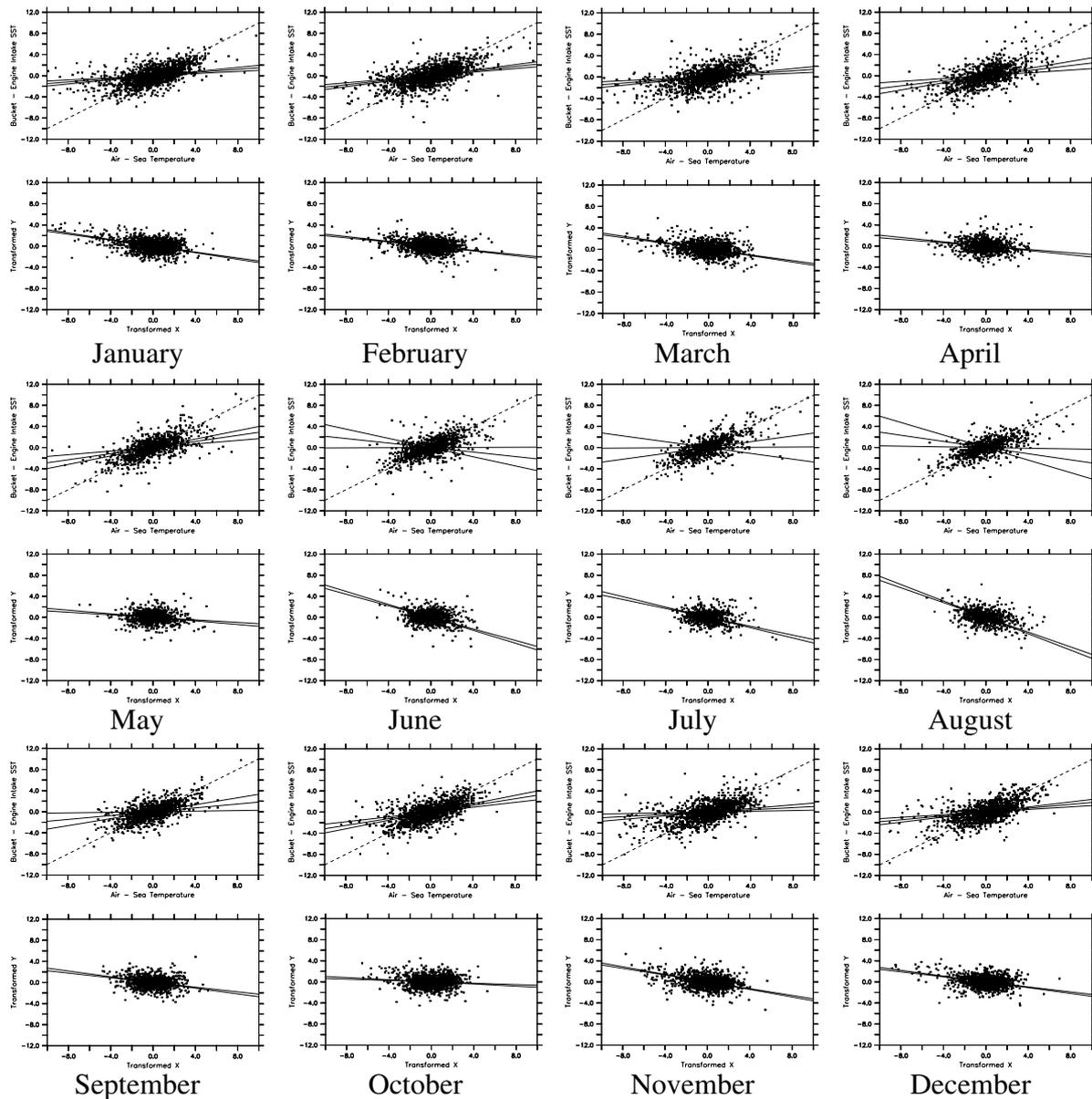


Figure 25 Regression lines in real and transformed co-ordinates. Mean values have been subtracted. The upper panel of each monthly pair shows bucket minus engine intake SST individual points plotted against air minus engine intake sea temperature. The dashed line is the relationship that would be expected if there was no relationship between x and y except for errors in the engine intake SST. The lower panel of each pair shows the same data in the transformed space along with two lines which span the uncertainty in the regression line. The best estimate regression line from the transformed data is converted back into real co-ordinates and plotted as the central line in the upper panels, along with the estimated range of uncertainty.

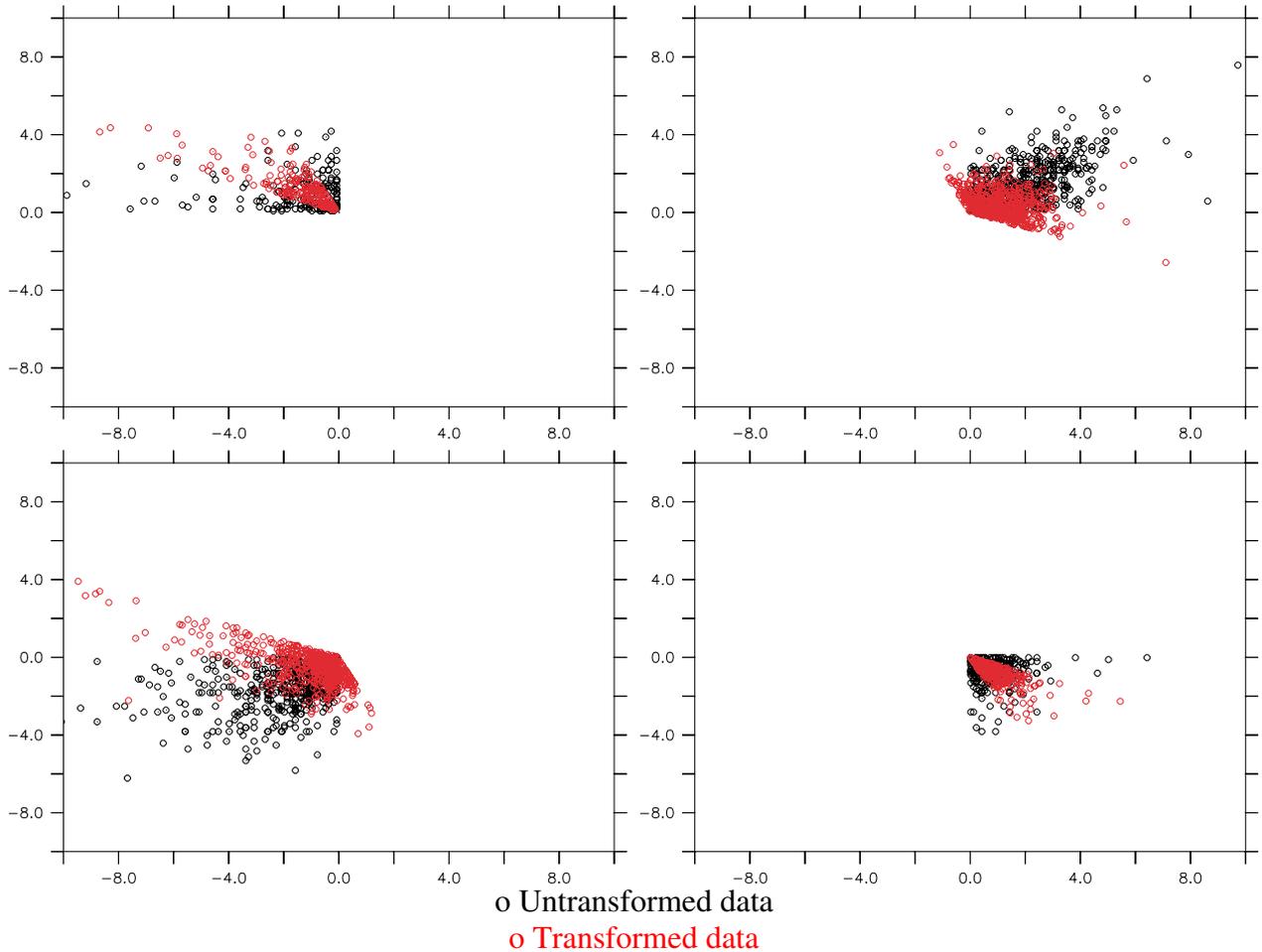


Figure 26 January data plotted by quadrant in both real and transformed co-ordinates by quadrant (defined by the real co-ordinates). This shows how the data is both weighted and rotated by the transformation. The upper right quadrant shows how data which is affected by the correlated errors in the x and y variables is down-weighted. The data from the upper left and lower right quadrants show data well away from the region of correlated errors being rotated so they are concentrated in a narrower band and mostly retain a similar weight. The lower left quadrant shows a combination of these two elements with suppression of the data containing correlated errors and a rotation of data in regions away from the correlated region.

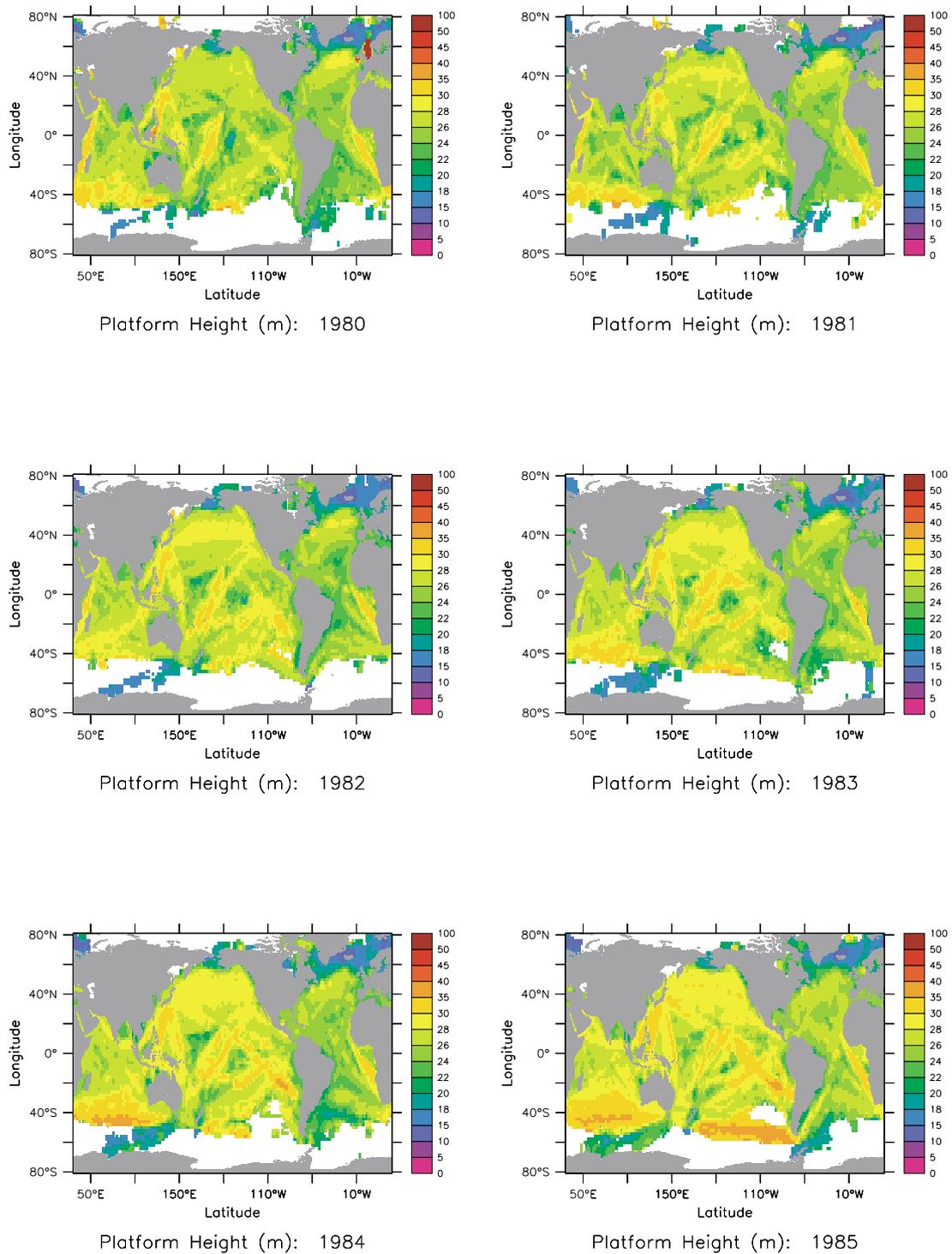


Figure 27a Global maps of the height of the observing platform for 1980 to 1985. Heights in metres have been gridded on a 2° area grid using a Gaussian smoother with width 2° and cut-off at 4° in each direction.

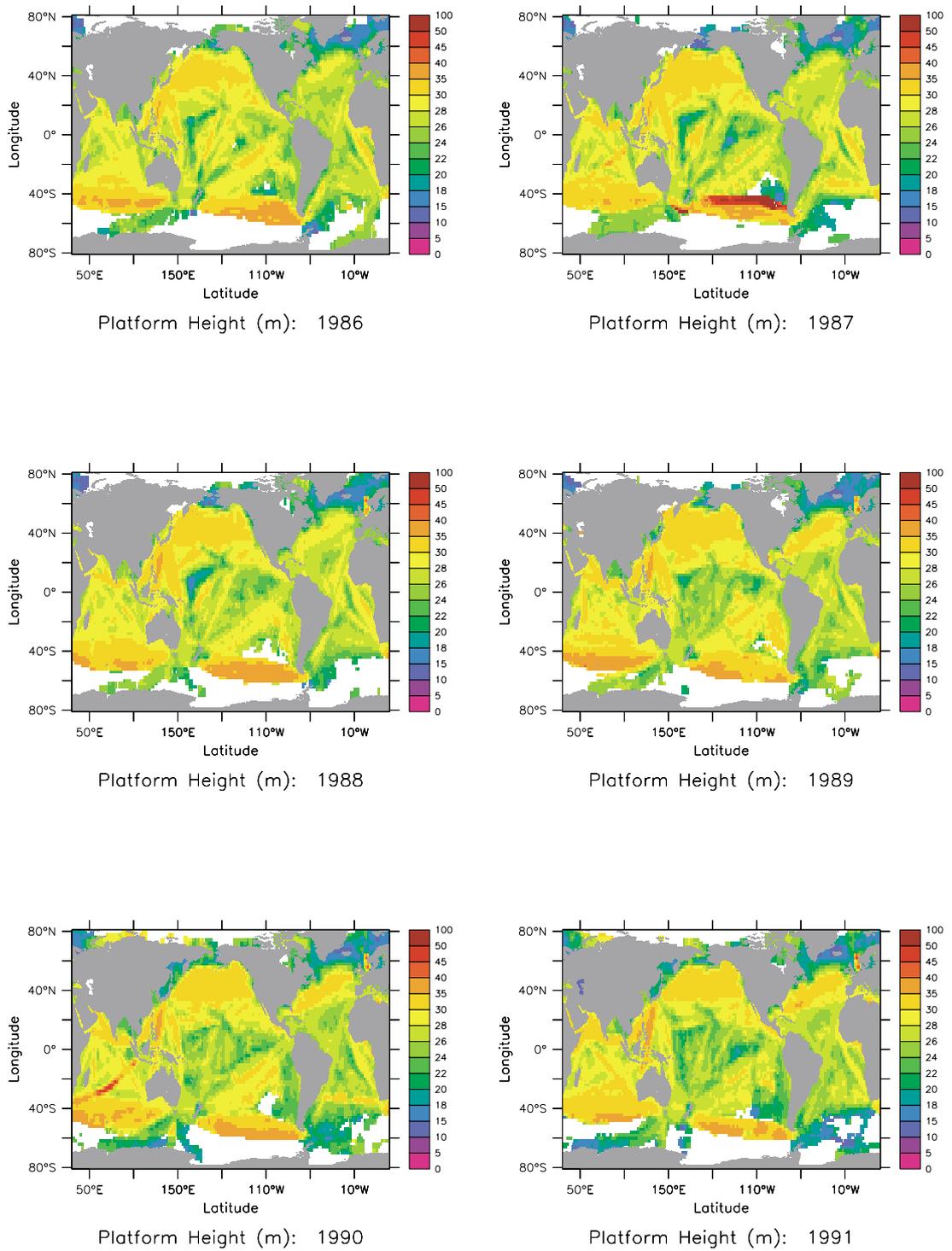


Figure 27b As Figure 27a but for 1986 to 1991.

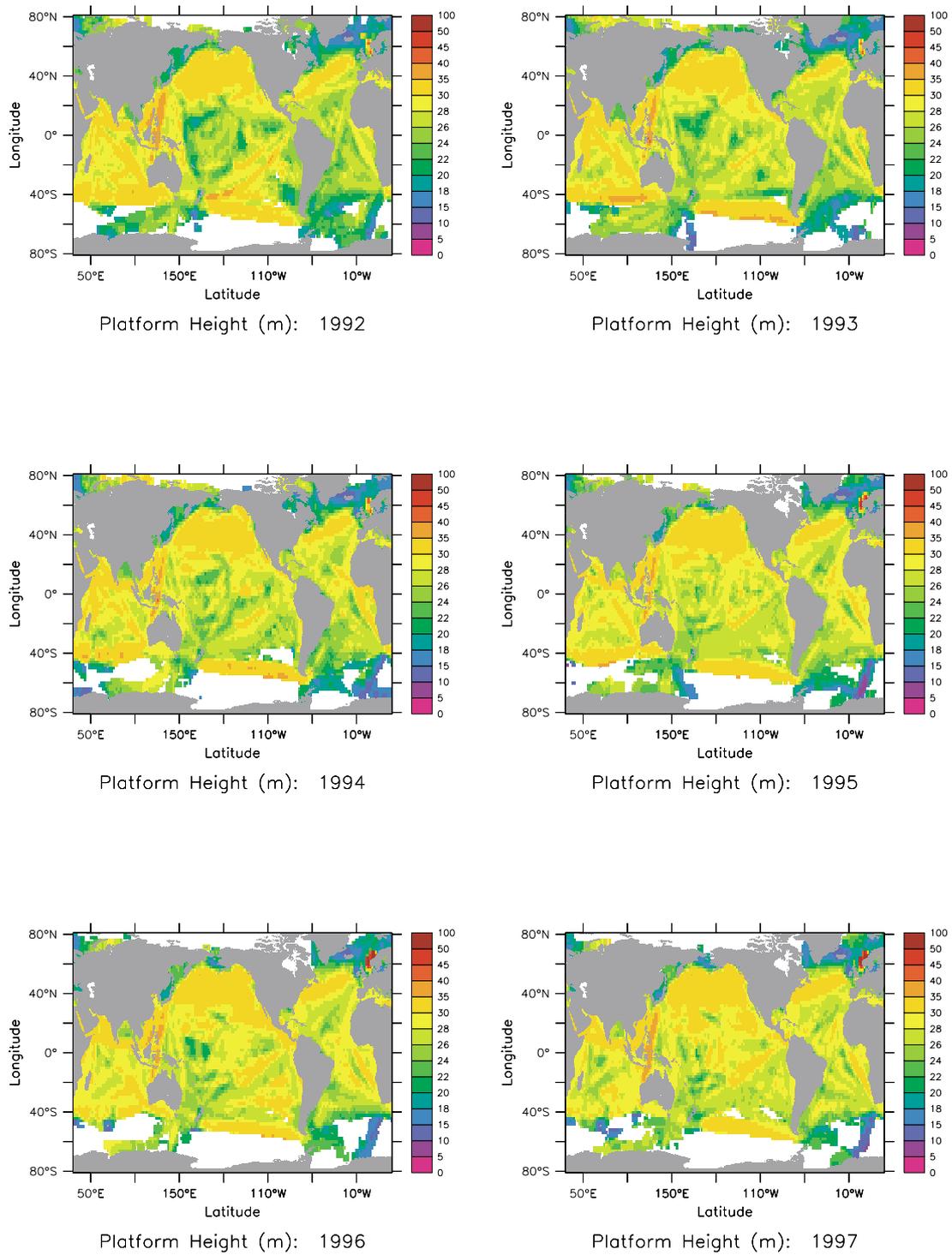


Figure 27c As Figure 27a but for 1992 to 1997.

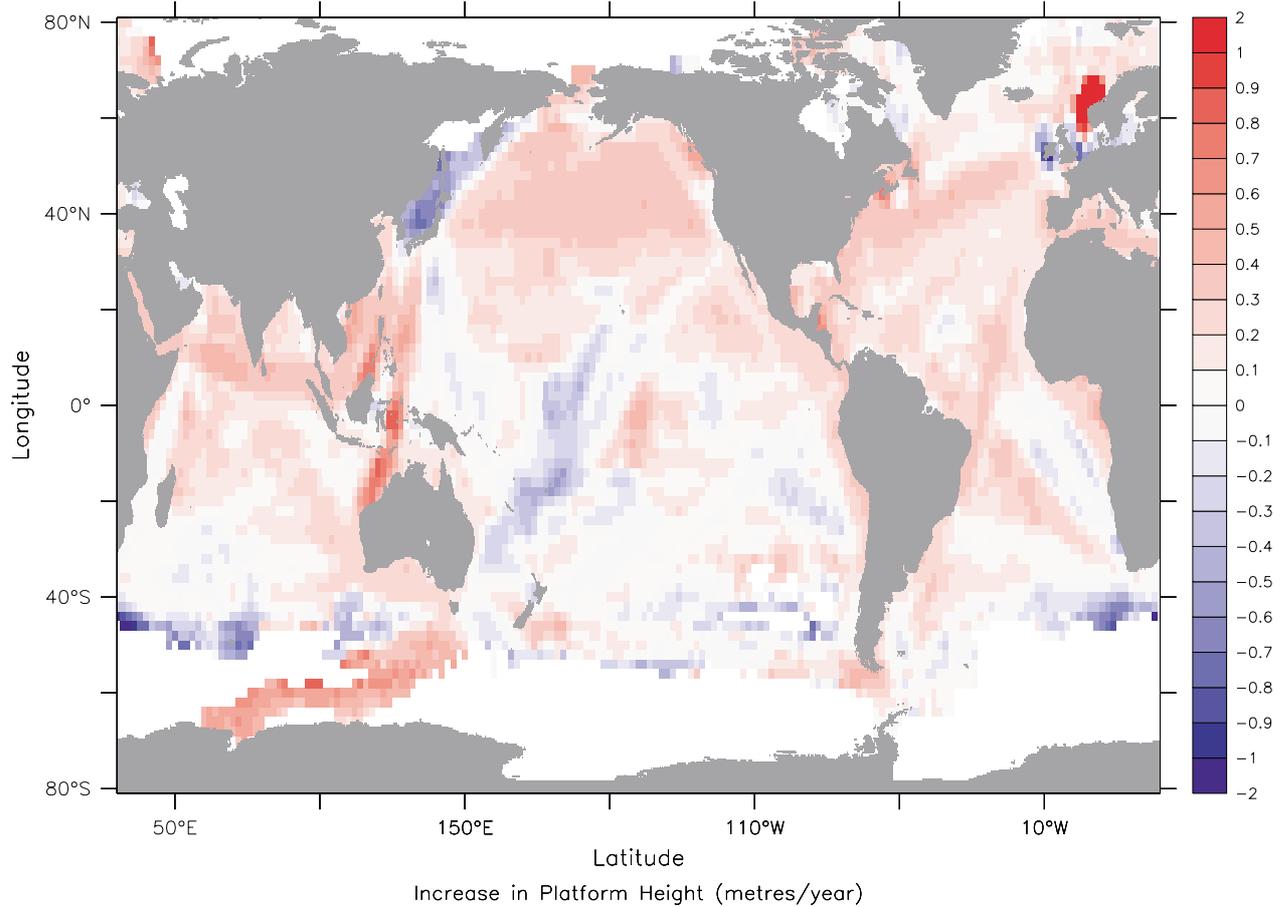


Figure 28 Global map of trend in observing platform height (m yr^{-1}). Trends are calculated from a linear regression of platform height in a 2° region (as shown in Figure 27) as a function of time.

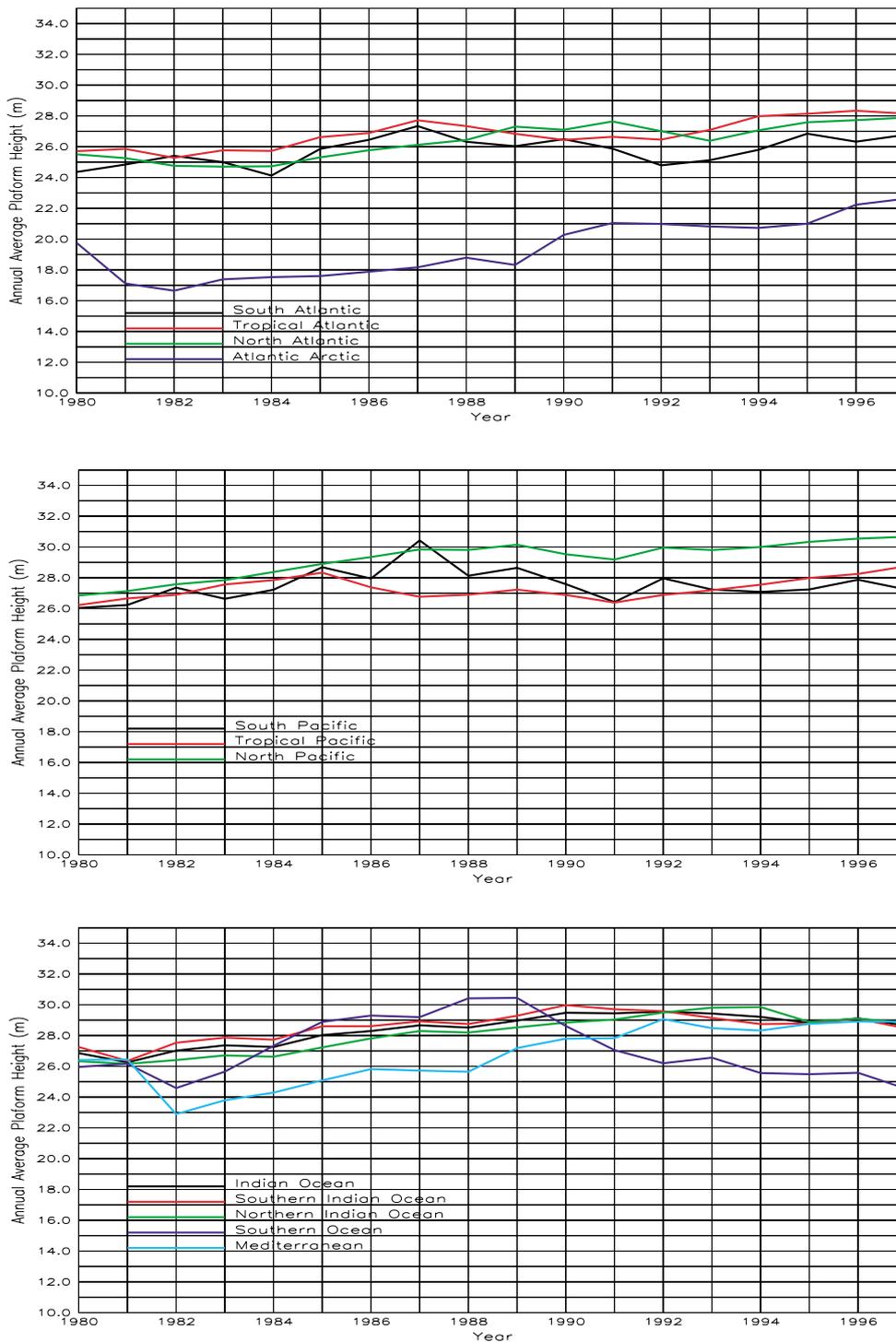
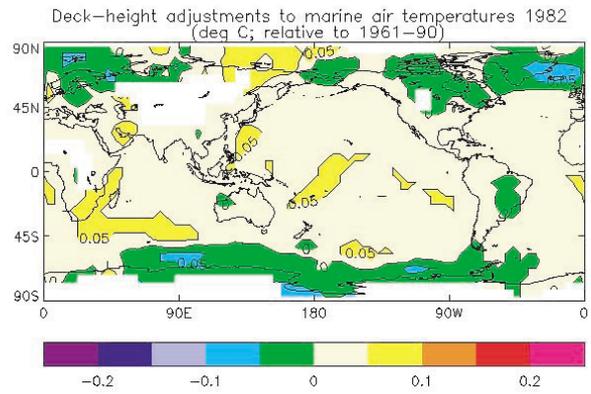
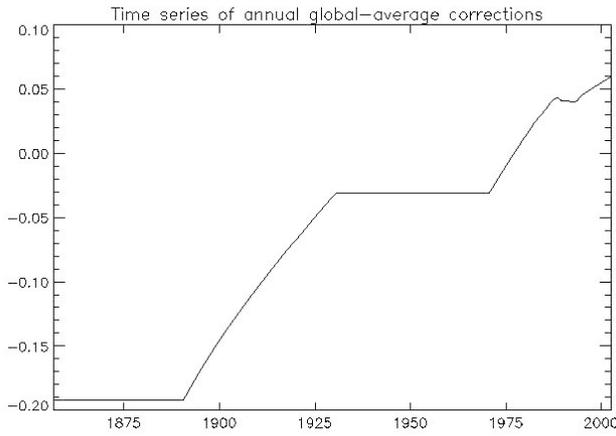
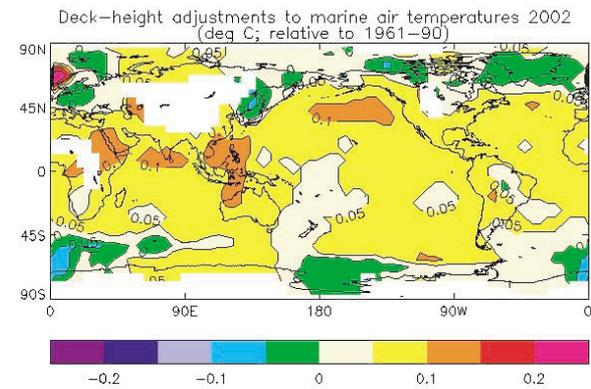
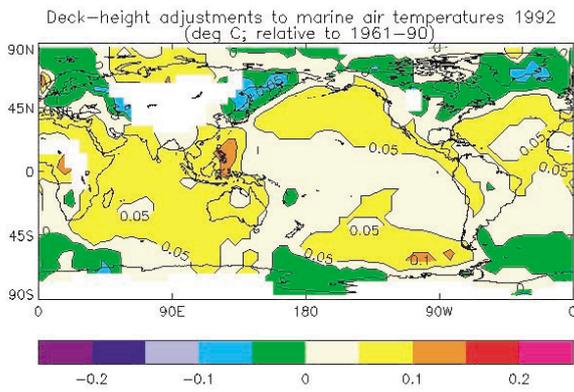


Figure 29 Average trends as shown in Figure 28 for different ocean areas.



a) Magnitude of global average correction

b) Regional Correction for 1982



c) Regional Correction for 1992

d) Regional Correction for 2002

Figure 30 The magnitude of the correction required to the air temperature to account for changes in observing height.

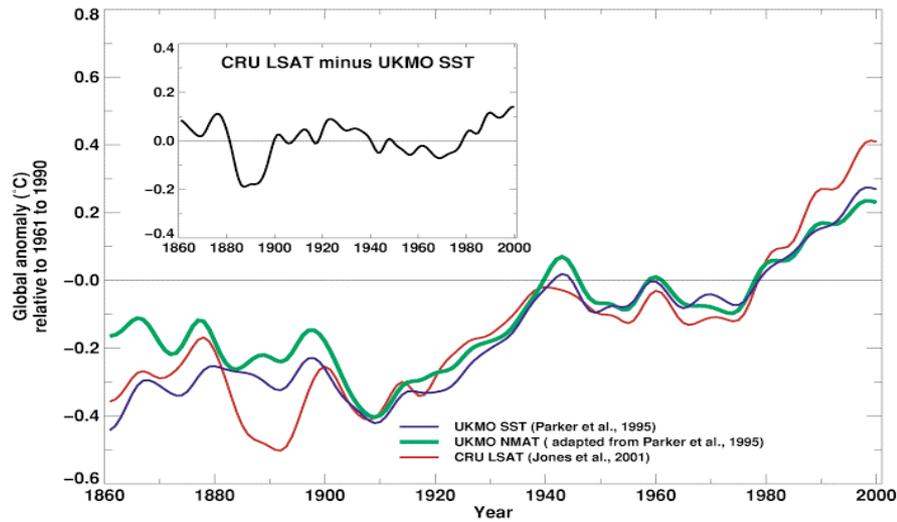


Figure 31a SST (blue line), Night Marine Air Temperature (green line) and Land Air Temperature (red line) global mean timeseries taken from IPCC (2001).

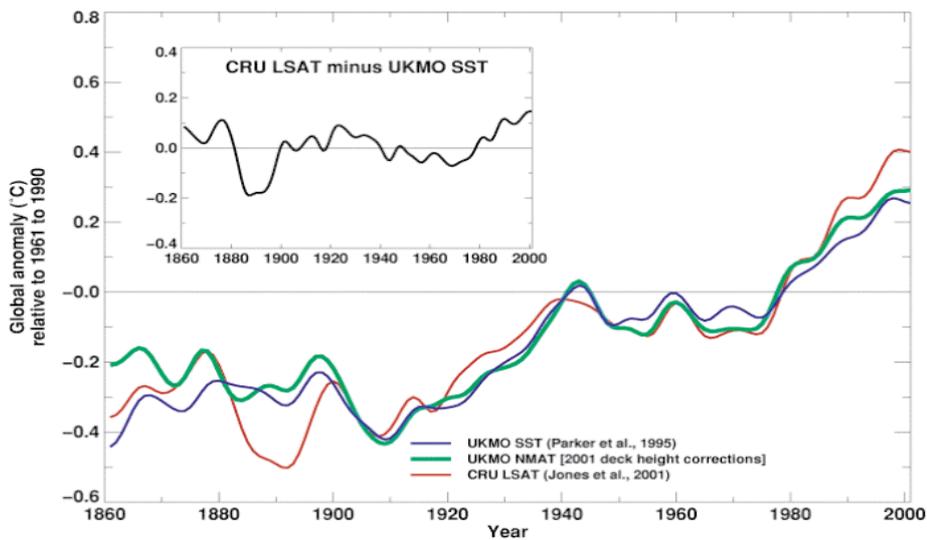


Figure 31b As Figure 31a but Night Marine Air Temperature with revised observation heights.

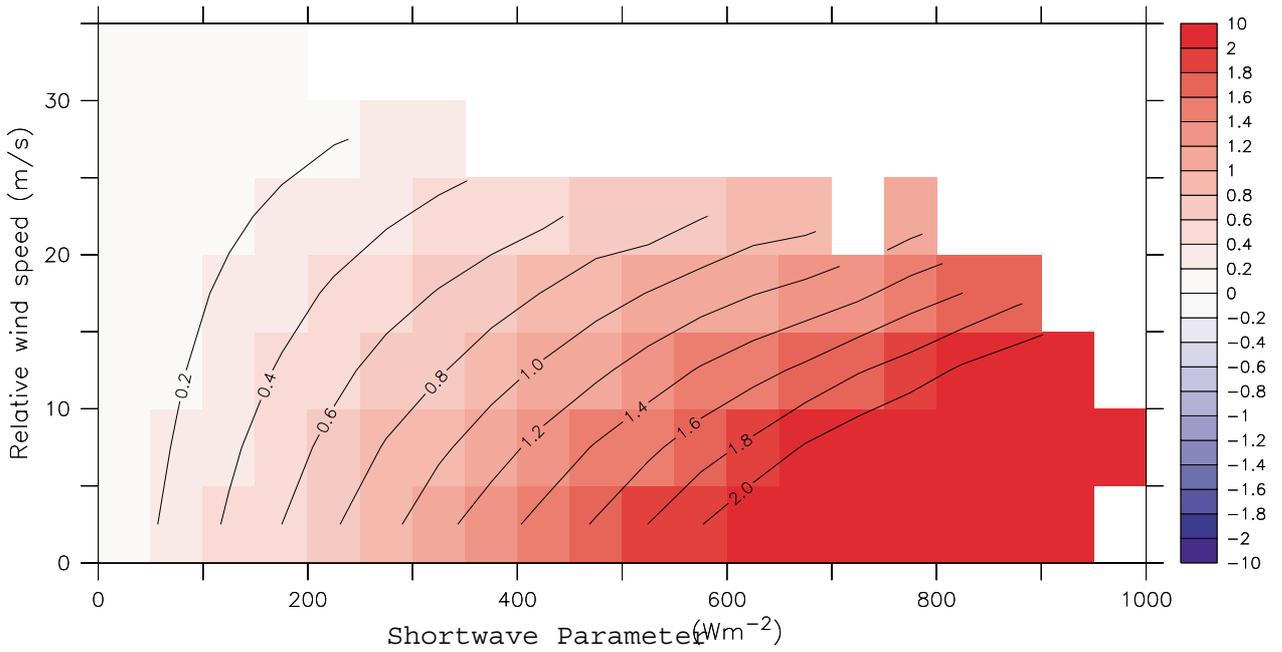


Figure 32a Magnitude ($^{\circ}\text{C}$) of correction required for VOS air temperatures within the VSOP-NA. The correction is a function of a shortwave parameter (defined as the incident shortwave after sunrise and before midday then a time decaying function of midday shortwave between midday and sunrise the following day) and relative wind speed (ms^{-1}).

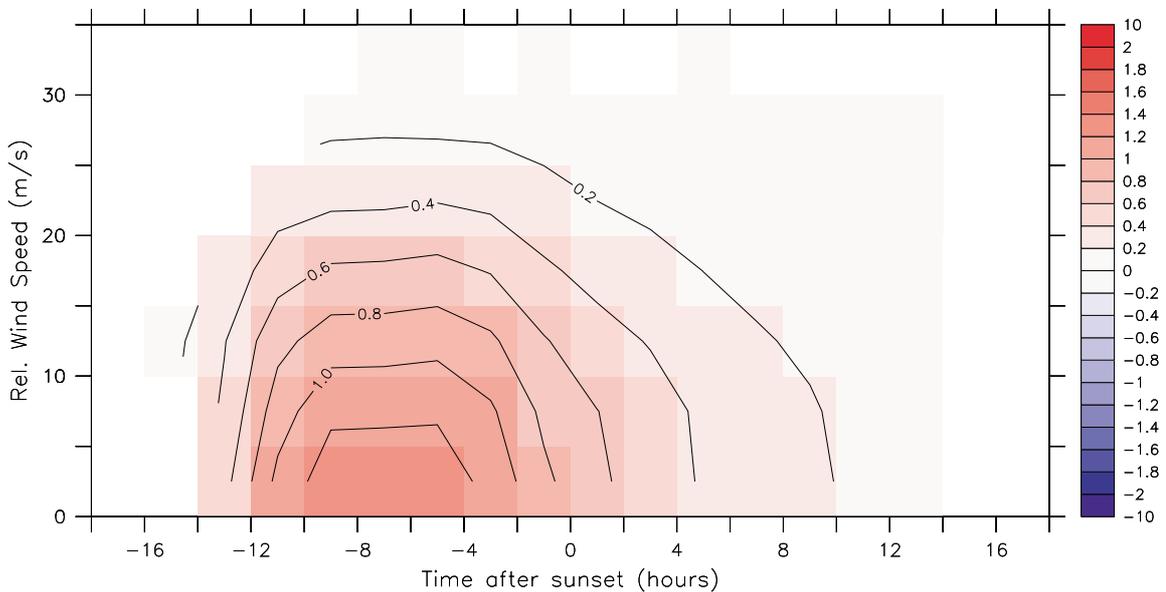


Figure 32b The magnitude of the correction shown in Figure 32a as a function of time after sunset and relative wind speed (ms^{-1}).

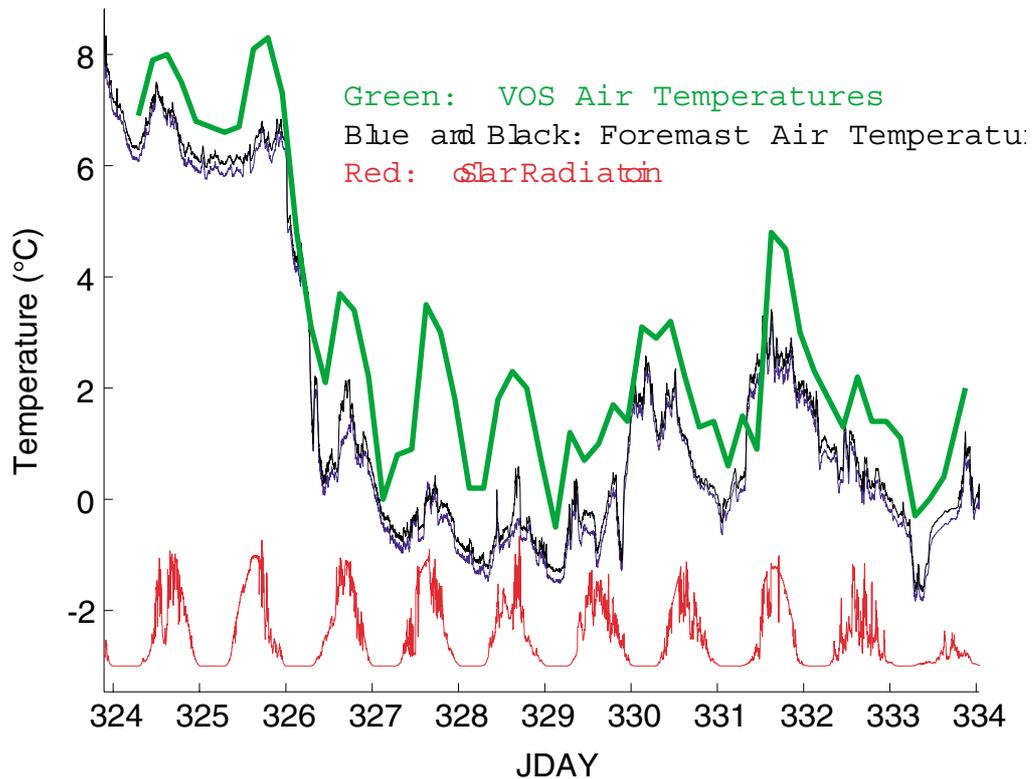


Figure 33 Air temperatures measured on the RRS James Clark Ross in December 2001. The blue and black lines are from the research quality sensors on the foremast. The difference between the two measurements is a function of the sheltering of one sensor by the junction box between them. The green line is the 6-hourly VOS reports from the same ship and clearly show the effect of solar heating (red line, arbitrary units).

APPENDIX**Method: Calculation of the inverse square root of a symmetric matrix**

To calculate D from the symmetric covariance matrix C firstly decompose matrix C using singular value decomposition (SVD):

$$C=USU^T$$

where U is orthogonal and S is diagonal. Note that this is a special case of SVD for a square symmetric matrix which means that the right and left singular vectors (U) are the same (normally SVD is in the form $C=USV^T$ where the right and left singular vectors, U and V are not equal). Then:

$$D = C^{-1/2} = US^{-1/2}U^T = U \begin{pmatrix} s_1^{-1/2} & 0 \\ 0 & s_2^{-1/2} \end{pmatrix} U^T$$

Proof: Errors in Z_1 are of unit rms and uncorrelated

$$\begin{bmatrix} D(\epsilon_x) \\ D(\epsilon_y) \end{bmatrix} \cdot \begin{bmatrix} D(\epsilon_x) \\ D(\epsilon_y) \end{bmatrix}^T = D \begin{pmatrix} \epsilon_x \\ \epsilon_y \end{pmatrix} (\epsilon_x \ \epsilon_y) D^T = DCD^T = D(D^T D)^{-1} D^T = DD^{-1}(D^T)^{-1} D^T = I$$