



Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies

J.J. Kennedy ^{*}, R.O. Smith, N.A. Rayner

Met Office, Exeter, UK

ARTICLE INFO

Article history:

Received 26 August 2009

Received in revised form 17 October 2010

Accepted 1 November 2010

Available online xxxx

Keywords:

Sea surface temperature

SST

AATSR

ATSR

HadSST2

Observational uncertainty

Observational error

Observing system adequacy

Observing system assessment

Traffic light plot

Climate

Ships

Drifting buoys

Moored buoys

Bias

SST bias

Sea surface temperature bias

Observing network

Observing network adequacy

Observing network assessment

GOOS

ABSTRACT

In situ data are widely used to provide a ground truth for the calibration and validation of satellite sea-surface temperature (SST) retrievals. They are also used to monitor long-term changes in the climate. For these applications, and others, it is necessary to understand the uncertainties in the data. Near-coincident SST observations from the Advanced Along-Track Scanning Radiometer (AATSR) and in situ platforms were used to understand the characteristics of errors in the measurements. The mean random error on the AATSR retrievals was found to be 0.14 K. The in situ errors were modelled as a constant offset plus a random error. For ships, the standard deviation of the constant offset was estimated to be 0.71 K and the mean random error was 0.74 K. For drifting buoys, the standard deviation of the constant offset was estimated to be 0.29 K and the mean random error was 0.26 K. These results suggest that there is a need to revisit current assessments of the adequacy of in situ observing systems. The trend in global-average SST between 1991 and 2007 calculated from in situ data was compared to its counterpart calculated from the ATSR instruments. The in situ record warms more slowly than the ATSR record, probably due to a decrease in the fraction of relatively warm-biased ship observations contributing to the global-average SST over the period and a corresponding increase in the fraction of relatively unbiased drifting buoy observations.

Crown Copyright © 2011 Published by Elsevier Inc. All rights reserved.

1. Introduction

Sea-surface temperature (SST) measurements constitute one of the longest global climate data records and are used extensively in climate analyses. Skilful forecasts of the climate can be made at seasonal (Folland, Parker, Scaife et al., 2006) and decadal (Keenlyside, Latif, Jungclauss et al., 2008) lead times due to the persistence of SST anomalies from month to month and year to year. Consequently, it is essential to be able to measure accurately the state of the sea surface and to understand the uncertainties in those measurements. As the focus of climate monitoring and forecasting shifts to a more regional scale, a comprehensive understanding of the uncertainties in climate measure-

ments will become even more important, particularly in areas where observations are sparse and potentially unreliable (Goddard, DeWitt, & Reynolds, 2009).

The majority of the climate record for SST comes from measurements made in situ by ships, drifting buoys and moored buoys (Rayner, Brohan, Parker et al., 2006; Worley, Woodruff, Reynolds et al., 2005). Although the measurements can be used to estimate such metrics as the global annual average temperature as far back as 1850, they are often sparsely distributed, yielding little information on short period events or at small spatial scales. Since the 1970s, satellite measurements of the sea surface have provided this additional detail (e.g. Chiodi & Harrison, 2006) and almost-global coverage is achieved by infra-red instruments starting in the early 1980s.

An important question is whether the observations – both satellite and in situ – constitute a record of SST that is adequate for the purposes of climate analyses. Global Ocean Observing System (GOOS, 1999)

^{*} Corresponding author. Tel.: +44 1392 885105; fax: +44 1392 885681.

E-mail address: john.kennedy@metoffice.gov.uk (J.J. Kennedy).

requirements state that, for climate change detection, the benchmark accuracy of the processed SST signal should be 0.1 K on 200–500 km squares monthly. It places a constraint both on the measurement error of the instruments and on the long-term stability of the observing network. Another consideration is that the sampling for in situ observations should be sufficient to remove bias from satellite products, both for climate applications and when atmospheric aerosol loading is high. The GOOS requirement recommends an accuracy of 0.1 K in 500 km squares on weekly time scales. However, it should be noted that some satellite instruments, for example the Along Track Scanning Radiometers, do not need to be calibrated using in situ observations.

In order to establish whether the accuracy of the observed SST record is adequate for such applications, it is necessary to characterise the uncertainties in the data. Because of the difficulties involved in measuring the temperature of a micrometre-thin film of sea surface through several kilometres of atmosphere, some satellite measurements are calibrated against in situ data. Ideally, the calibration would be made against equivalent radiometric measurements, but the cost of such devices means that it is not practical to deploy them in large numbers. In situ thermometric data are available in large numbers but are of variable quality and introduce a further uncertainty because the skin temperature as measured by the satellite needs to be converted to an equivalent bulk temperature as measured by the in situ thermometer. Uncertainties are largest during the day particularly when wind speeds are low and insolation is high, thus allowing a near surface warm layer to form that can be several degrees warmer than the water at depth.

The ATSR (Along Track Scanning Radiometer) instruments were designed to deliver SST measurements that meet the demanding requirements of climate analyses (Llewellyn-Jones, Edwards, Mutlow et al., 2001). The dual-view configuration allows aerosol biases to be detected and removed from the data without recourse to in situ measurements; the on-board calibration ensures the long-term stability of measurements from the instrument. O'Carroll, Eyre, and Saunders (2008) have demonstrated that ATSR retrievals have relatively low measurement errors. Corlett, Barton, Donlon et al. (2006) and Barton (2007) showed that the biases and other errors in ATSR retrievals are generally lower than for other satellite SST retrievals making them an ideal reference data set. A reanalysis of the SST record produced by the ATSR instruments will provide the backbone for future climate quality analyses of SST (Merchant, Llewellyn-Jones, Saunders et al., 2008).

The adequacy of the global ocean observing system has been assessed by Zhang, Reynolds, and Smith (2006) and Zhang, Reynolds, Lumpkin et al. (2009). They calculated the number of buoys and ships needed in each 10° square to reduce satellite biases of 2 K to below 0.5 K and biases of 1 K to below 0.3 K. This approach has been used to target the deployment of drifting buoys since 2005 and thereby make best use of the available resources. Their approach depends on a thorough assessment of the uncertainties in the in situ data.

So, a key question for understanding both in situ and satellite records is, how accurate are the in situ data?

Until the 1990s, most in situ measurements were made by Voluntary Observing Ships (VOS). In the past, the measurements from VOS ships were not made with the needs of climate researchers in mind and the variation in bias and accuracy means that they do not often meet the stringent GOOS requirements. Several analyses have attempted to quantify the uncertainties associated with in situ SST measurements. The biases associated with the different methods used by ships to make SST measurements have been analysed by a number of authors, most recently Kent and Kaplan (2006). Emery, Baldwin, Schlusell et al. (2001), Smith and Reynolds (2002), Kent and Challenor (2006), and O'Carroll et al. (2008) made estimates of the random measurement uncertainties in ship and buoy data. More recently, Kent and Berry (2008) proposed a new error model

for in situ data that combines the usual random term (intra-platform error) with a bias term that varies from ship to ship (the inter-platform error). The basic rationale is that a ship's thermometer which is mis-calibrated such that it reads 0.5 K too high (inter-platform error) will never get closer than 0.5 K to the true SST other than by a chance combination of random errors (intra-platform error). In this scheme, aggregating the observations from many ships or buoys with independent inter-platform errors is the only way to reduce the average error to zero. Brasnett (2008) used this error model implicitly in a scheme for calculating bias corrections for individual ships.

The analysis of measurement uncertainties is complicated by the fact that in situ observations are typically widely separated. Ideally, high quality, well-calibrated, coincident observations would be used to characterise the data, but these are rarely available in large quantities and, where they are available, it is not given that the results will generalise to other times and locations. Usually, pairs of in situ observations are compared making it harder to disentangle the individual error components.

For the purposes of the analysis presented here, the ATSR instruments are used as a consistent reference field with low measurement error (O'Carroll et al., 2008) and low detector bias (Barton, 2007, Corlett et al., 2006) that is independent of the in situ record. Near-coincident pairs of in situ and AATSR observations are used to assess the error characteristics of the in situ data.

The in situ and ATSR data are described in Section 2. A basic error analysis is detailed in Section 3 for comparison with earlier studies and to further establish the low-error character of the AATSR retrievals. In Section 4, the error model is extended to include inter- and intra-platform errors. In Section 5 the consequences of non-normal error distributions are explored. The insights gained from the first five sections are used in Section 6 to reassess whether the in situ network is fit for purpose as assessed against the GOOS requirements for satellite calibration. In Section 7, attention is turned to the longer term and the ATSR instruments are used to verify the accuracy of long-term trends in the in situ global climate record. Summary and conclusions are given in Section 8.

2. Data

The ATSR (Along Track Scanning Radiometer) series of instruments provide an almost complete record of SST from 1991 to the present. ATSR-1 flew from 1991 to 1995 and ATSR-2 from 1995 to 2008. The AATSR instrument is mounted on the ESA (European Space Agency) Envisat satellite which was launched in 2002. Envisat is in a near-polar sun-synchronous orbit, which crosses the equator at 10:00 local time on the descending node. The AATSR instrument has an inclined conical scan configuration, observing in the forward (55°) and nadir directions. Three infra-red channels centred in the atmospheric windows at 11 µm, 12 µm and 3.7 µm are used to retrieve the skin SST and additional channels in the visible spectrum are used for cloud clearing. During daylight hours, there is the possibility of solar contamination of the 3.7 µm channel. Therefore retrievals using all three channels are only available at night. The instrument was designed to give SST retrievals with an accuracy of better than 0.3 K and a long term stability of better than 0.1 K decade⁻¹.

The AATSR record from 2002 to 2007 was consistently reprocessed at the Met Office from the ESA AATSR METEO product and the level 1b brightness temperatures using the December 2005 Case C retrieval coefficients (http://earth.esrin.esa.it/services/auxiliary_data/aatsr/ATS_SST_AXVIEC20051205_102103_20020101_000000_20200101_000000) that were made operational on 7th December 2005 as recorded in the change log (<http://earth.eo.esa.int/pcs/envisat/aatsr/events/>) and are based on the HITRAN 2000 database. ATSR1 and ATSR2 data are processed using version 2 of the METEO product from NEODC (<http://neodc.nerc.ac.uk/?option=displaypage&Itemid=>

91&op=page&SubMenu=-1). The skin SSTs were converted to a sub-skin temperature using the scheme of O'Carroll, Watts, Horrocks et al. (2006). The sub-skin SST is more directly comparable with the in situ data. Because they are likely to be of higher quality than other retrievals, dual-view, three-channel (D3) retrievals were used in this analysis. As D3 retrievals are only available at night, only night-time observations were used in the analysis and complications arising from diurnal heating of the near-surface layer are not considered further. O'Carroll et al. (2008) estimated the measurement error on the D3 product to be about 0.16 K.

The spatial resolution of the ATSR data was 10 arc min in latitude and longitude. The 10 arc min product was used in preference to the 1 km full resolution product because its error characteristics had already been established by O'Carroll et al. (2008).

For the greater part of the analysis in this paper, the D3 night-time retrievals are used, but in the final section, D2 night-time retrievals from the ATSR-1 instruments are also used. There is a residual bias in the D2 retrievals for which a latitude-dependent correction has been developed. The data used here were corrected using the values given in Table 1, which are an older version of the corrections estimated for AATSR retrievals: http://envisat.esa.int/pub/ESA_DOC/ENVISAT/AATSR/Latitude_Dependent_Bias_Correction.pdf.

Monthly files of in situ data, covering the period August 2002 to December 2007, were downloaded from the National Centers for Environmental Prediction (NCEP) near real time marine observations web page (<http://icoads.noaa.gov/nrt.html>). The NCEP files contain Global Telecommunication System (GTS) meteorological reports from ships, moored buoys and drifting buoys. Drifting buoys typically take hourly observations (0000, 0100, 0200...2200, 2300 UTC) as do moored buoys. Some ships take hourly observations, but most take four observations a day at the synoptic reporting hours (0000, 0600, 1200 and 1800 UTC). The data were quality controlled using the method described in Rayner et al. (2006). Sea-surface temperatures were stored to the nearest 0.1 K and latitudes and longitudes to the

nearest tenth of a degree of arc. The observation times were stored to the nearest 0.01 of an hour UTC.

In the following analysis the moored buoys are split into two distinct groups: the equatorial moorings consisting of the Indian Ocean moorings, TAO and PIRATA buoys; and all other moorings.

3. Basic analysis and comparison with previous work

The first step was to perform a simple analysis that could be compared to earlier estimates of errors in in situ data. This provides a point of comparison with the existing literature and forms the basis for the more detailed error analysis performed in the later sections.

A data base was created containing paired near-simultaneous in situ and AATSR observations made between August 2002 and December 2007. For each in situ observation all AATSR observations taken within 1 h and within 1° of latitude and longitude were found. The closest one was selected. If the in situ observation came from a ship and no AATSR observation was available within 1 h, the closest AATSR observation taken within 2 h was used instead. This was done because many ships take only four measurements each day at fixed UTC hours (0000, 0600, 1200 and 1800 UTC) and these coincide with the Envisat overpass times, which occur at fixed local hours, only in certain limited regions. With a one-hour matchup criterion the interaction with the Envisat overpass times meant that the coverage was strongly biased to certain geographic locations. The effect is much less pronounced with a two-hour matchup criterion (Fig. 1).

The geographical variation in the numbers of matchups between August 2002 and December 2007 is shown in Fig. 1. The AATSR-ship pairs are concentrated in the North Pacific and North Atlantic. There are far fewer ships and, consequently, far fewer matches in the southern hemisphere. There are systematic observation shadows, most notably in the central Pacific where the AATSR overpass time does not fall within 2 h of a typical ship observation time. The matches with drifting buoy observations are more evenly spread and the highest concentrations are in the sub-tropical regions of the Atlantic. The smallest numbers of AATSR-drifting buoy pairs are found in the eastern Pacific, which is characterised by persistent cloud cover and divergent surface currents, and in the region between the Pacific and Indian Ocean where the complex topography means that there are few drifting buoys.

The distance between the in situ and AATSR observations was measured from the centre of the AATSR 10 arc min cell. The distribution of AATSR-in situ pairs peaks below a separation distance of 10 km for all observation types (Fig. 2). Fig. 2 also shows the distribution of SST differences for each AATSR-in situ pair, again split according to in situ observation type. The drifting buoys have the narrowest distribution, with a standard deviation of only 0.64 K. The distribution of AATSR-ship pairs is much broader, with a standard deviation of 1.31 K. The distributions are non-normal with a positive kurtosis or excess. Large differences are more likely than they would be if the differences were normally distributed. The implications of non-normality are explored in Section 5.

The mean AATSR-drifting buoy difference is 0.1 K. The mean AATSR-ship difference is -0.08 K. The implied bias between ships and drifting buoys at the over-pass time is therefore 0.18 K, which is above the range of values obtained by Reynolds, Gentemann, and Corlett (2010) - 0.13 to 0.14 K - but similar to that of Emery et al. (2001), 0.19 K. Ship observations are typically warmer than drifting buoy observations.

The variance of these distributions will have components due to the measurement errors of the in situ data and the measurement error of the AATSR data. Although the distributions of the pair separations are peaked at low values, significant numbers of pairs have separations exceeding 20 km. Consequently, there will be a component of the variance that arises from the fact that the paired observations were not made at exactly the same place. To ascertain

Table 1
Latitude dependent corrections applied to ATSR-1 D2 retrievals.

Latitude (°N)	Correction (K)
-75	0.15
-70	0.07
-65	0.06
-60	0.03
-55	0.00
-50	-0.02
-45	-0.04
-40	-0.07
-35	-0.10
-30	-0.12
-25	-0.09
-20	-0.03
-15	0.03
-10	0.06
-5	0.05
0	0.05
5	0.04
10	0.04
15	0.02
20	-0.04
25	-0.08
30	-0.10
35	-0.08
40	-0.06
45	-0.03
50	0.00
55	0.02
60	0.03
65	0.05
70	0.08
75	0.10

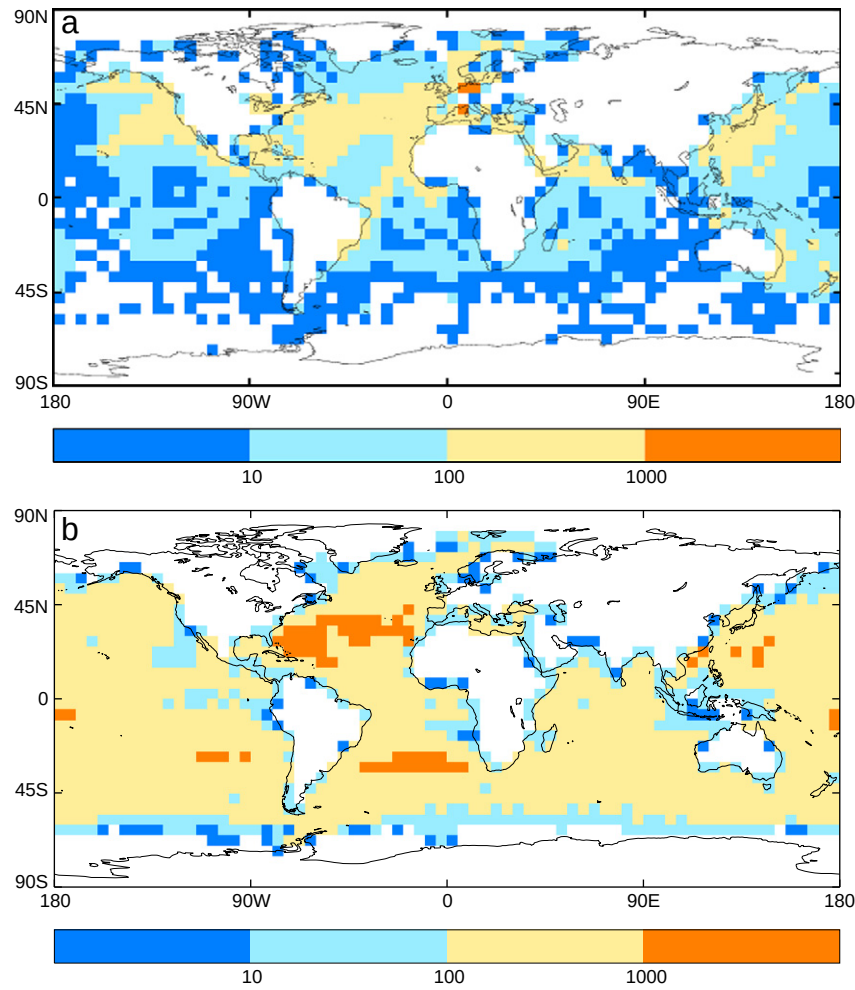


Fig. 1. Number of ATSR-in situ pairs in the analysis for each 5° latitude $\times$ 5° longitude box for ships (top) and drifting buoys (bottom) between August 2002 and December 2007.

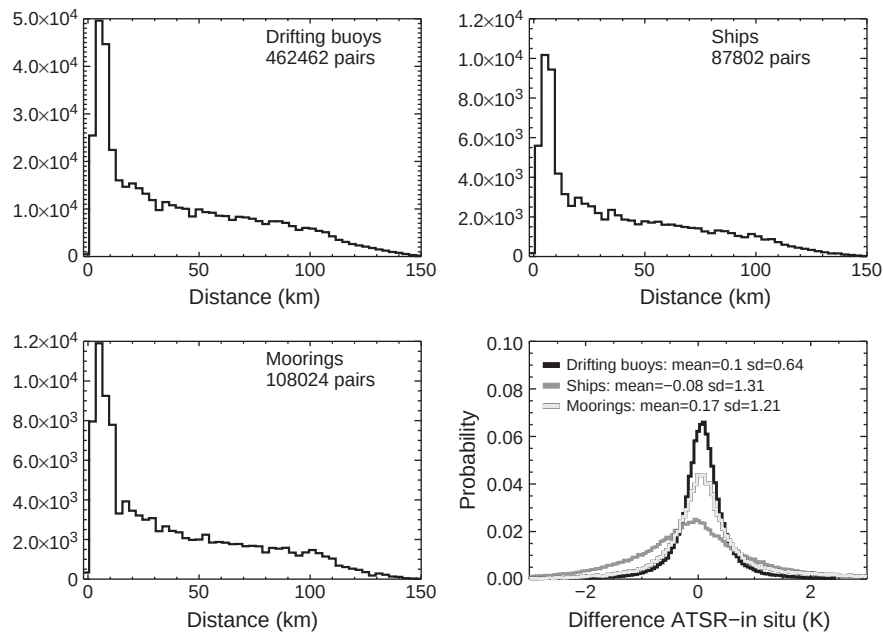


Fig. 2. Numbers of paired observations as a function of separation distance for drifting buoys (top left), ships (top right), moorings (bottom left) between August 2002 and December 2007. The distributions of the SST difference between the ATSR-in situ observation pairs are shown bottom right for drifting buoys (black), ships (dark grey) and moorings (light grey).

the size of the effect, the variance of the observation pairs was plotted as a function of separation. The relationship is approximately linear (Fig. 3) and extrapolating the linear relationship to the y-intercept yields the variance at zero separation. In extrapolating to zero separation it is assumed that the linear relationship holds at small distances. Kent, Challenor, and Taylor (1999) showed that the linear relationship holds at separations below 300 km and above 50 km. The present analysis shows that the relationship remains approximately linear down to separations of less than 10 km.

For ships, the variance at zero separation was $1.10 \pm 0.01 \text{ K}^2$. For drifting buoys the variance was $0.06 \pm 0.001 \text{ K}^2$, for moored buoys $0.07 \pm 0.02 \text{ K}^2$, and for the equatorial moored buoys, $0.04 \pm 0.003 \text{ K}^2$. The uncertainties are those from the regression, but additional uncertainties are discussed at the end of the section. The gradient of the linear relationship represents the small scale variability in the sea-surface temperature. It is higher for the moored buoys because many are located in, or near, areas with strong SST gradients such as the Gulf Stream, Kuroshio and coastal waters. In more thermally homogeneous areas, such as the tropics, the gradient is much smaller, as can be seen in the case of the equatorial moorings.

The variance at zero separation contains a component from measurement errors in the AATSR retrievals. O'Carroll et al. (2008) used three-way matchups – near-simultaneous SST measurements from the AATSR, the Advanced Microwave Scanning Radiometer, and drifting buoys – to estimate the measurement errors for each of the three platforms. The AATSR uncertainty was estimated to be around 0.16 K.

The paired observation method employed here can be used to estimate the AATSR uncertainty as well. The earth's surface was divided into 1° latitude $\times$ 1° longitude areas. Pairs of AATSR observations taken on the same day that fell within a single 1° area were chosen at random. The distance between them and the difference between the SSTs were recorded. No more than ten pairs were taken from each 1° area each day to avoid biasing the final result towards better-sampled areas.

Fig. 4 shows the total number of pairs taken in each 1° area, along with the variance of the SST differences as a function of the separation distance. The estimated variance at zero separation, 0.04 K^2 , is twice the AATSR error variance. It corresponds to an AATSR measurement uncertainty of 0.14 K. The estimate is lower than that of O'Carroll et al.

(2008), although it is within the range of estimates that they obtained by varying their experimental parameters (0.12–0.16 K). A number of factors could explain the discrepancy. The error variances of near-coincident AATSR retrievals are likely to be correlated over the small distances considered here. Correlation would reduce the variance in the difference relative to the case where the retrievals were uncorrelated. Geographical variation of the error in the AATSR retrievals would mean that the differing locations of the matchups in the two analyses could affect the results.

Assuming the AATSR measurement uncertainty is 0.16 K, as it is in O'Carroll et al. (2008), the standard deviation for ship observations is 1.04 K, which is consistent with the values calculated by Kent and Challenor (2006), $1.2 \pm 0.4 \text{ K}$ or $1.3 \pm 0.3 \text{ K}$, but lower than those in: Kent et al. (1999), $1.5 \pm 0.1 \text{ K}$; Kent and Berry (2005), $1.3 \pm 0.1 \text{ K}$ and $1.2 \pm 0.1 \text{ K}$; and Reynolds, Rayner, Smith et al. (2002), 1.3 K. The standard deviation for drifting buoy observations, 0.20 K, is: lower than the estimate in Reynolds et al. (2002) of 0.5 K; higher than in Emery et al. (2001), 0.15 K; and roughly the same as the 0.23 K reported by O'Carroll et al. (2008). The moored buoys have a standard deviation of 0.21 K and the equatorial moorings 0.12 K. These values do not change significantly if the AATSR measurement uncertainty is assumed to be 0.14 K.

There is a potential complication, in that the AATSR product is an area average rather than a point observation so the comparison with the in situ data is not an exact like-for-like comparison. In order to investigate the likely effect of this problem a toy example was devised. An array containing random numbers with unit variance and covariance equal to $\exp(-\text{dist}/200)$ was created where *dist* is simply the distance between the points in the array. These represent the SST in the example. Pseudo-in situ measurements were created by picking pairs of points, adding specified noise (0.20 units), calculating the difference and building up the variance vs distance relationships as above. AATSR-equivalent observations were created by averaging together 10 neighbouring points in the array and adding specified noise (0.14 units). AATSR-in situ pairs were calculated by measuring the distance from the AATSR cell centre to the in situ location. Finally, pairs of AATSR-equivalent measurements were compared. The picking of points was repeated 1000 times for 1000 different pseudo-SST fields. The AATSR and in situ errors were calculated as they were above by comparing the y-intercepts. The estimated AATSR error was

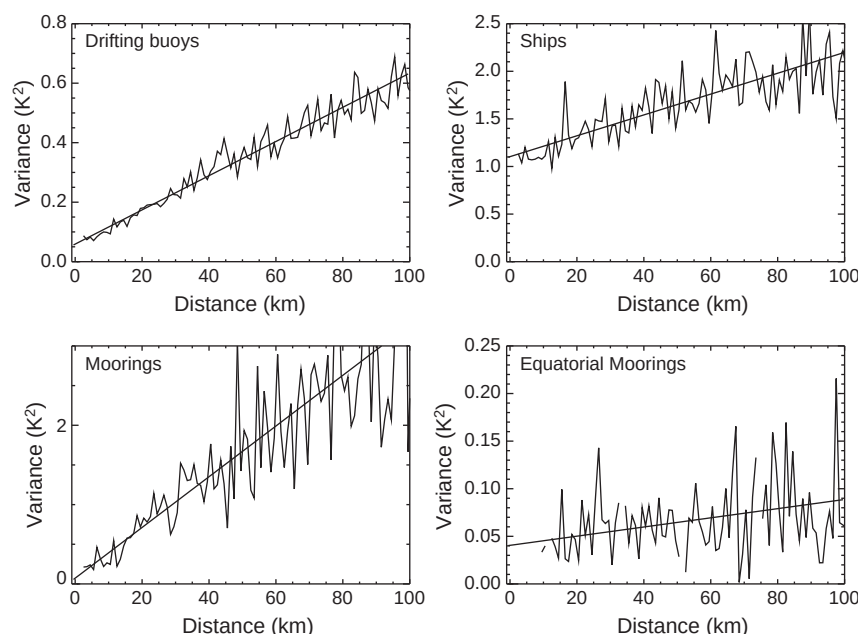


Fig. 3. Variance of AATSR-in situ differences as a function of distance for drifting buoys (top left), ships (top right), all moorings (bottom left) and equatorial moorings (bottom right) between August 2002 and December 2007. The straight line is the linear least squares fit.

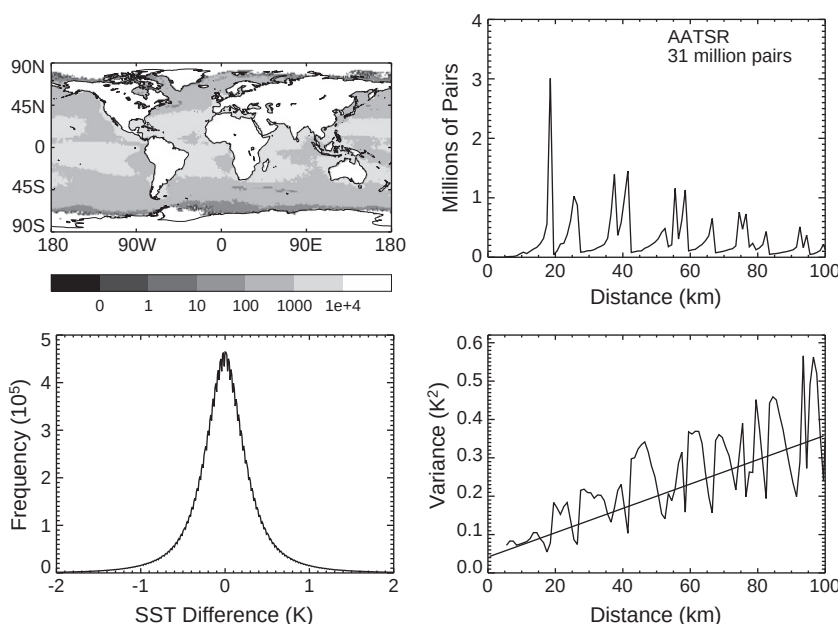


Fig. 4. (top left) Number of ATSR-ATSR pairs per 1° area. (top right) Number of ATSR-ATSR pairs as a function of separation. (bottom left) Distribution of the SST difference between the ATSR-ATSR observation pairs between August 2002 and December 2007. (bottom right) Variance of ATSR-ATSR differences as a function of distance.

underestimated at 0.13 units and the estimated in situ error was correctly estimated at 0.20 units. The size of the over and under estimate depends on the relative length scales of the AATSR footprint and the SST variability. In this example the ratio was 1:20, based on an AATSR footprint of (at most) 18 km and an SST length scale of 360 km. Increasing the ratio to 1:7 leads to an underestimate of the ATSR error (0.12) but only a slight overestimate (0.203) of the in situ error.

The error estimates calculated in this section are a baseline for comparison with earlier estimates. They assume that the errors on each measurement are unrelated, an assumption that is questioned in the next section.

4. Correlation of errors

Kent and Berry (2008) proposed an alternative error model for in situ observations. They decomposed the error into two components. The first is a constant offset that is fixed for a given platform (ship, buoy or mooring) and is the same for all observations made by that platform. Kent and Berry (2008) called it the inter-platform error; the same convention is adopted here. The second is an uncorrelated random error term that is different from one observation to the next. The variance of the second error term is fixed for each ship and Kent and Berry (2008) called it the intra-platform error.

The inter-platform error can arise for a number of reasons. Persistent biases between different methods of measurement have been noted since the early 20th century. Some ships' crews measure the temperature of water scooped up in specially designed buckets, others read the temperature of the water taken in to cool the ship's engines. Each method leaves its own characteristic biases in the data (Kent & Kaplan, 2006). But there are also variations peculiar to a particular ship. James and Fox (1972) identified engine inlet biases that depended, amongst other things, on the size of the ship, the depth of the water inlet and the distance travelled by the water between inlet and thermometer. Brasnett (2008) identified the inter-platform errors in ship data. Instead of including the differences in an error model, the observations from each ship were compared to a background analysis and the differences were used to calculate a ship-by-ship bias correction.

One difficulty with estimating inter-platform errors consistently is that the errors need to be calculated relative to a reference field.

Ideally, the reference field would closely approximate the true SST. Existing comparison standards such as simultaneous oceanographic profiles (Folland, Reynolds, Gordon et al., 1993) are too few to make reliable estimates, so it has been necessary to make practical compromises. Kent and Berry (2008) exploited model analysis output from the Met Office Numerical Weather Prediction system as a reference SST, but were unable to separate cleanly the model errors from the measurement errors. The Brasnett (2008) analysis compared SSTs to an interpolated analysis, essentially comparing the SST to an average of all other observations including the observations to be corrected. The method assumed there was no geographical correlation between errors from different ships. This might not be true because ship biases and errors depend on the country that recruited the ship and ships recruited by a given country tend to have a limited geographical range.

SST retrievals from the AATSR instrument are a good candidate for a reference data set. They have low measurement errors and they are characterised by temporal stability with little geographical variation in bias. In the following, it is assumed that, aside from a random measurement error, the AATSR retrievals are perfect. i.e. they have no bias relative to the true SST.

Individual in situ platforms were identified using their call sign, an alphanumeric indicator unique to that ship or buoy. The mean and variance of the SST differences between the AATSR-in situ pairs were calculated for platforms with 25 or more paired observations where the separation was less than or equal to 20 km. 25 observations were required to ensure that the spread in the inter-platform error estimates was not artificially inflated by random intra-platform errors. The mean difference is an estimate of the inter-platform error and the variance of the differences is an estimate of the sum of the intra-platform error variance and the AATSR error variance. Subtracting the AATSR error variance at zero separation, 0.16 K, gives an estimate of the intra-platform error. This will contain a small contribution from SST variability at distances less than 18 km, so the resulting intra-platform errors will be slightly overestimated. However this will mostly affect the intra-platform errors which are the least critical for the observing system assessment performed in Section 6 as they diminish rapidly when observations are averaged together. Distributions of the inter- and intra-platform errors are shown in Figs. 5 and 6.

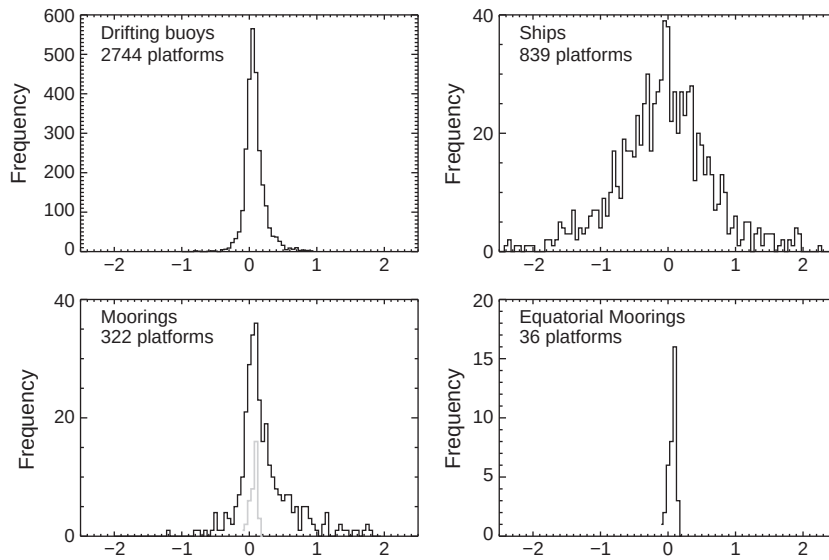


Fig. 5. Distributions of inter-platform error for drifting buoys (top left), ships (top right), moorings (bottom left, equatorial moorings are shown in grey) and equatorial moorings (bottom right) between August 2002 and December 2007. Only platforms with more than 25 AATSR-in situ pairs are shown.

To check the validity of the assumption that persistent AATSR biases were unimportant, the distributions were recalculated after the AATSR data had been ‘corrected’ by removing a geographically-varying average AATSR-buoy difference. The mean and standard deviations of the distributions changed by less than 0.02 K in each case with some increasing and others decreasing.

The distribution of inter-platform errors was widest for the ship observations and narrowest for observations from drifting buoys and equatorial moorings. Out of 839 ships, 10 had biases exceeding 2 K and 122 had biases exceeding 1 K. By contrast, 97% of drifting buoys had biases smaller than 0.5 K, and fewer than 2% exceeded 1 K. The intra-platform errors show a similar behaviour. The average intra-platform error for drifting buoys was 0.18 K whereas the average for ships was 0.63 K. Ship observations have higher uncertainties on the average than drifting buoy observations. It is not true, however, that all ship observations are unreliable. Around 29% of ships had inter-platform errors of less than 0.3 K and intra-platform errors of less than 0.5 K and around 8% of ships had both inter- and intra-platform errors of less than

0.2 K. The converse is also true: although most drifting buoys had low errors, around 1% exhibited gross errors in excess of 1 K.

The spread in inter-platform errors is large for all moorings and the intra-platform errors exceed 1 K in some instances. However, the equatorial moorings in the TAO, PIRATA and Indian Ocean arrays have a much smaller range of inter- and intra-platform errors.

For gridded SST products, it is necessary to convert uncertainties on single observations into uncertainties on a grid-box average. For a population of m ships and drifters with inter-platform errors σ_{bi} and intra-platform errors σ_{gi} the grid-box average uncertainty σ_{tot} can be written as:

$$\sigma_{tot}^2 = N^{-2} \sum_{i=1,m} n_i \sigma_{gi}^2 + N^{-2} \sum_{i=1,m} n_i^2 \sigma_{bi}^2 + \text{sampling error} \quad (1)$$

Where N is the total number of observations and n_i is the number of observations from platform i . The final term is a sampling error term that is not considered here. An assessment of sampling errors

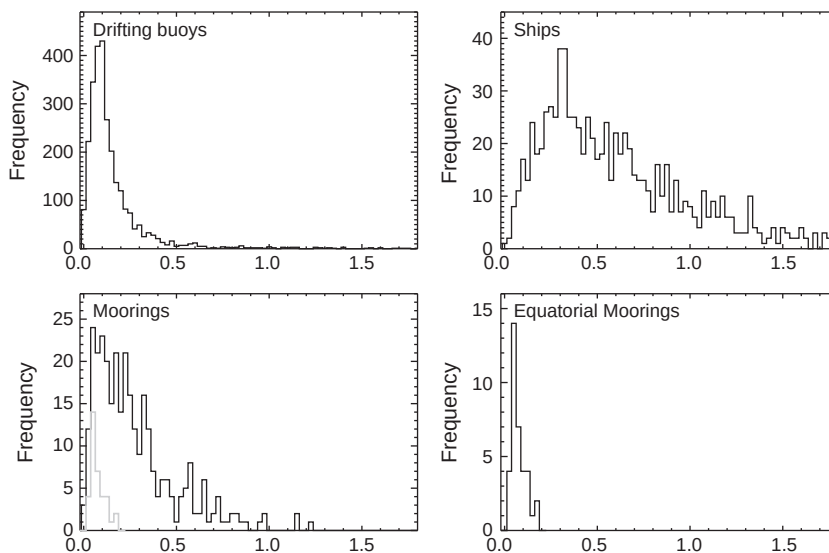


Fig. 6. Distributions of intra-platform error for drifting buoys (top left), ships (top right), moorings (bottom left, equatorial moorings are shown in grey) and equatorial moorings (bottom right) between August 2002 and December 2007. Only platforms with more than 25 AATSR-in situ pairs are shown.

can be found in Morrissey and Greene (2009), Kent and Berry (2008) and Kennedy et al. (2011a). To better understand Eq. 1 it is instructive to consider the case where each of the m platforms takes the same number of observations and each platform has the same inter- and intra-platform error characteristics. The above formula then reduces to:

$$\sigma_{\text{tot}}^2 = \sigma_g^2 / N + \sigma_b^2 / m + \text{sampling error} \quad (2)$$

The equi-partition of observations between the platforms minimises the grid-box average uncertainty for a given number of observations N and platforms m . It also shows that it is only possible to reduce grid-box average uncertainties to a minimum by increasing the number of observations (N) and the number of different platforms making them (m). Fig. 7 shows a schematic representation of the variance for the case where σ_g and σ_b are each 0.71 K.

For many platforms the values for the inter- and intra-platform errors will be unknown, but the platform type – ship, drifter, mooring – will be. With such a wide range of inter- and intra-platform errors (Figs. 5 and 6) it is not straightforward to choose values for σ_g and σ_b that are representative for the different platform types. The choice is complicated by the fact that the distributions are non-normal and that there is some correlation between the values of the inter- and intra-platform errors: a ship that has a large inter-platform error is more likely to have a large intra-platform error.

To work out representative values for each platform type, pseudo grid-box averages were created. The sampling error was assumed to be zero as was the true grid-box average SST. A number of ships, m , and a number of observations per ship, N/m , were chosen and m sample values of the inter-platform and intra-platform errors were drawn from the distributions shown in Figs. 5 and 6. The N/m pseudo-observations for Ship 1 were randomly generated from a normal distribution with mean equal to the inter-platform error and standard deviation equal to the intra-platform error. The pseudo-observations for Ship 2 were calculated in the same way and so on for each of the m ships. The average of all the pseudo observations, the number of ships and the total number of observations were recorded and the process was repeated 10,000 times for values of m and N/m between 1 and 50.

Figs. 8 and 9 show the variance of the 10,000 pseudo grid-box averages as a function of the number of observations and the number of platforms for ships and drifting buoys respectively. The variance for a single observation is the sum of the inter- and intra-platform error variances. Values of σ_g and σ_b were estimated using multiple regression to fit Eq. 2 to the variances. The sampling error was set to zero. The grey lines show the best-fit calculated according to Eq. 2. The representative values of the inter-platform and intra-platform errors for ships were 0.71 K and 0.74 K respectively. For buoys the inter-platform error was 0.29 K and the intra-platform error was 0.26 K.

The sum of these values should approximately equal the error variances at zero separation deduced in the previous section. The ship data are consistent, but the drifting buoy data are not. The sum of the inter- and intra-platform variances exceeds the variance at zero separation calculated from the individual observations. One possible reason is that buoys often exhibit larger biases towards the end of their lifetime. The biases last only a short time so they make only a small contribution to the distributions in Figs. 2 and 3, which are based on individual observations. However, the biases would have a larger effect on the statistics for a single buoy leading to an increase in the estimated uncertainty derived from the distributions shown in Figs. 5 and 6.

5. Non-normality of the error distributions

For many applications of SST data, it is assumed that the measurement errors are approximately normal. It is clear from the AATSR-in situ pairings that the error structure of in situ observations is non-normal, with a positive kurtosis characterised by a high, narrow peak and broad tails. Although the errors on individual observations are not normal, when large numbers of observations are aggregated into gridded products, the central limit theorem states that the distribution of the gridded averages will approach that of a normal distribution.

Pseudo-gridded averages were generated by drawing samples from the distribution of AATSR-ship differences shown in Fig. 2 and averaging them together. A 25% trimmed mean was used to calculate a gridded average from the samples. 100,000 examples of gridded

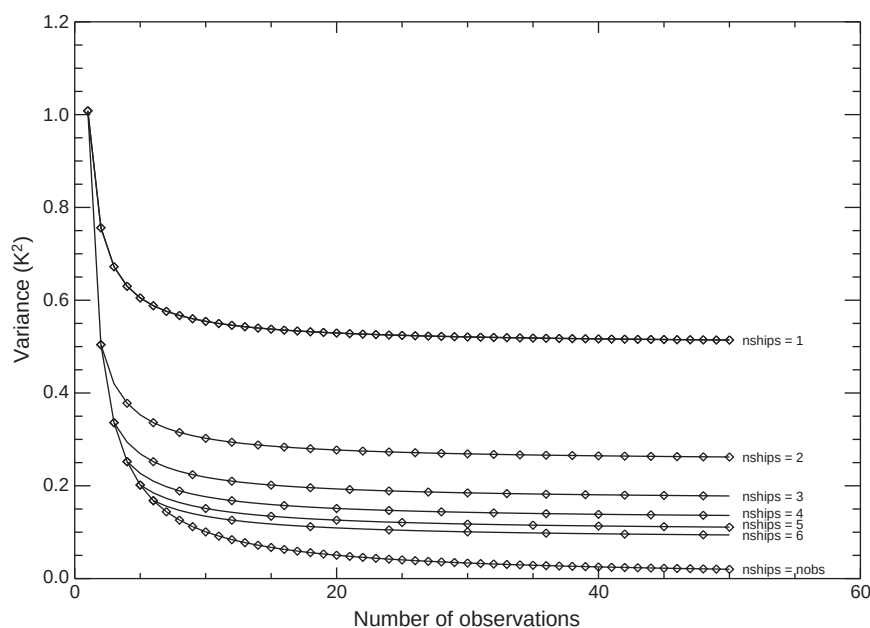


Fig. 7. Schematic showing the grid-box average error variance as a function of the number of observations and number of ships in the case where each ship takes the same number of observations.

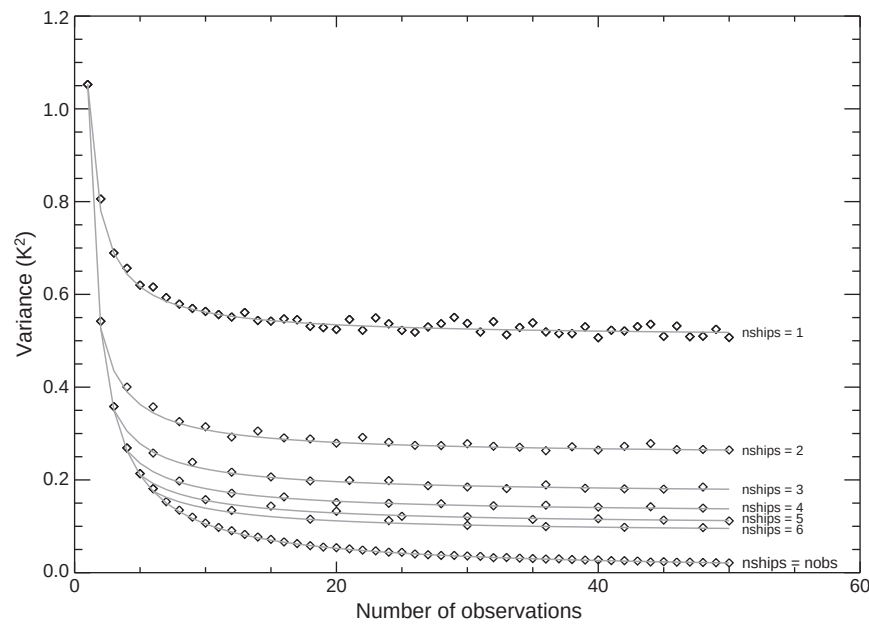


Fig. 8. The diamonds show the variance of pseudo grid-box averages as a function of the number of observations and the number of ships generated from 10,000 estimates. The grey lines show theoretical estimates of the error structure that have been fit to the data.

averages containing 1 observation were created, 100,000 containing 2 observations and so on up to 50 observations. The variance and kurtosis of those estimates are shown in Fig. 10.

For small numbers of observations the observed kurtosis implies that the distributions are non-normal. A true normal distribution will have a kurtosis of zero. The kurtosis decreases until about 20 observations after which little improvement is seen. Estimates of kurtosis, however, are strongly affected by small perturbations from perfect normality and such a deviation from zero is not significant. This suggests that the distribution of grid-box averages comprised of fewer than 20 independent observations might exhibit significant deviations from the normal distribution. If the analysis is sensitive to such outliers, it might be necessary to exclude such grid-boxes from consideration or take steps to minimise their effect.

6. Using the error estimates to assess the in situ network

One of the GOOS requirements for in situ sea-surface temperature measurements states that the sampling and accuracy should be sufficient to remove bias from satellite products. Here, the target accuracy is 0.1 K in a 500 km square on weekly time scales. It is not clear exactly what 'target accuracy' means. It is often taken to mean that the standard deviation of the expected errors is smaller than 0.1 K. Another way of looking at it is to calculate the probability that the error in a grid square exceeds 0.1 K, which is the method followed here.

In situ SST data for 1–7 June 2006 were extracted from the GTS archive and quality-controlled. The number of unique platforms making observations in each 5° grid box (~500 km at the equator) was recorded, along with the number of observations made by each

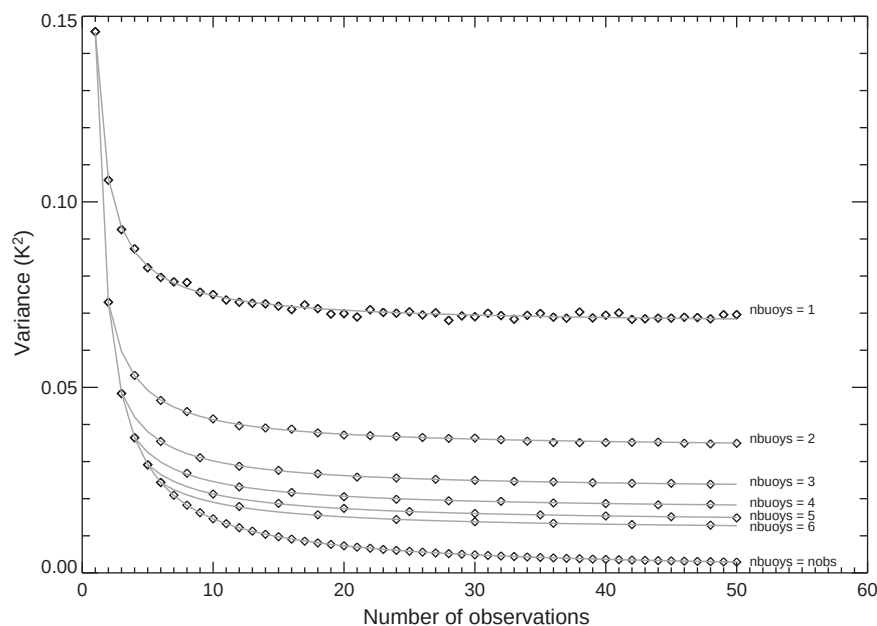


Fig. 9. The diamonds show the variance of pseudo grid-box averages as a function of the number of observations and the number of buoys generated from 1,000,000 estimates. The grey lines show theoretical estimates of the error structure that have been fit to the data.

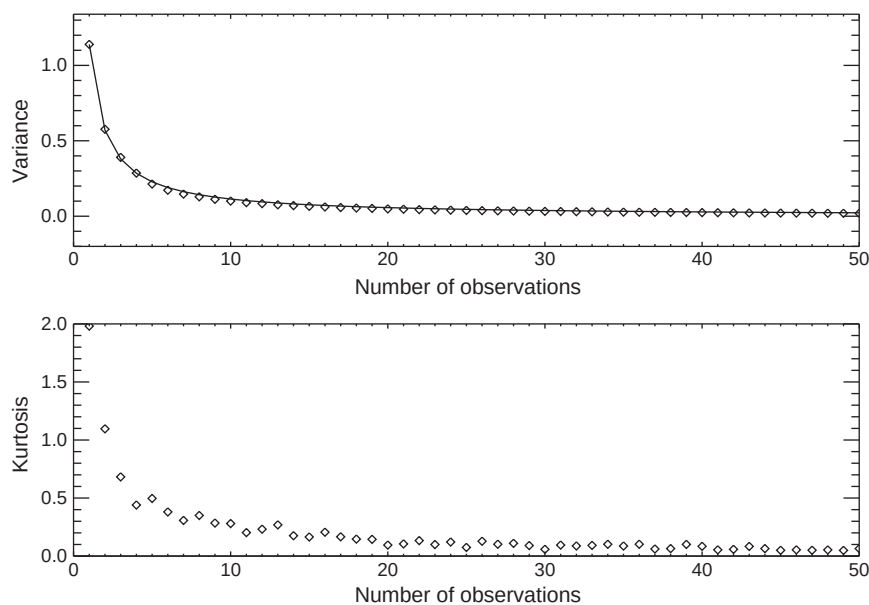


Fig. 10. Variance and Kurtosis of gridded averages as a function of number of observations.

unique platform in each grid box and the type of platform. Three pseudo-data sets, using error models of different complexity, were created using this information. Sampling error is not considered, but we note that sampling error can be an important component of the total grid-square uncertainty and that the uncertainties calculated here will therefore underestimate the true error.

In the first of the three data sets, the full error model and all information in the distributions was used. Estimates of the inter- and intra-platform errors for each unique platform were drawn from the appropriate distributions in Figs. 5 and 6. Random pseudo-observations were generated for each unique platform with mean equal to the inter-platform error and standard deviation equal to the intra-platform error. The number of observations generated for each unique platform was equal to the number made by the platform between 1 and 7 June. The process was repeated for all unique platforms and the pseudo-observations were averaged together onto a 5° grid. The process was repeated 1000 times to give a representative range of realistic measurement errors.

In the second, the error model described in Eq. 1 was used. It assumes that there are correlations between the data, but that the distributions of inter- and intra-platform error are normal. Estimates of the generic inter- and intra-platform errors for each platform were taken from Section 4. For ships these were 0.71 K and 0.74 K and for drifters, 0.29 K and 0.26 K. For each unique platform an inter-platform error was drawn from a normal distribution with standard deviation equal to the generic inter-platform error. Random pseudo-observations were then generated and averaged as in the first case.

In the third, the errors were assumed to be uncorrelated between observations. For each platform the generic intra-platform error and the generic inter-platform error were summed in quadrature. Random pseudo-observations were generated for each unique platform with mean equal to zero and standard deviation equal to the summed error. The pseudo-observations were then averaged as in the first case.

Fig. 11 shows the probability that the error in the gridded pseudo-observations exceeds 0.1 K for the error models and error values deduced in this paper. If an accuracy of 0.1 K is understood to mean that the standard deviation of the errors is 0.1 K and that the errors are normally distributed, the probability that the error exceeds 0.1 K should be one third or less. There are large areas in Fig. 11 where the probability exceeds this threshold even before errors due to under-sampling are considered. Fig. 12 shows the probability that the error

on a grid-square average exceeds 0.1 K for different combinations of numbers of observations and numbers of ships and drifting buoys for the full error model.

The differences between the three error models are interesting. In the simplest model shown in the lower panel, the errors are considered to be uncorrelated. This is a common assumption (see for example Rayner et al., 2006 and Zhang et al., 2009) and, under this assumption, many areas meet the criterion that the probability of the error exceeding 0.1 K is less than one third. Those areas where the probability exceeds one third are areas where there are few observations.

The middle panel of Fig. 11 shows the results for the intermediate-complexity model. In this case, most grid squares exceed a probability of one third. This is unsurprising because the correlation between the errors increases the uncertainty. The inter-platform error of a drifting buoy is 0.29 K, implying that more than 8 drifting buoys would be needed ($0.29/\sqrt{9} < 0.1$) in one grid square to reduce the uncertainty on the average below 0.1 K. It is unlikely that 8 drifting buoys would congregate by chance in one grid square.

The upper panel of Fig. 11 shows the probabilities for the most complex error model. The area where the requirement is met is smaller than in the lower panel because of the correlation between the errors. However, the area is larger than in the middle panel. This is because the distributions from which the inter-platform errors are drawn for the upper panel are more sharply peaked than the normal distributions used in the middle panel. The peaking makes it more likely that a small inter-platform error will be drawn.

The perceived adequacy of the observing network depends on the choice of error model. Furthermore, the parameters of more realistic error models are not easy to estimate and the approach taken was only possible because of the proven stability and low uncertainty that are characteristic of the AATSR data. The results derived here suggest that observing system specifications should be revisited as understanding of the observations improves and that it might be wise to allow some redundancy in observing system design.

7. Long-term trends

Because the historical record of SST consists of in situ measurements that were not made for the purposes of climate monitoring, its long-term stability is questionable. A number of authors (Folland & Parker, 1995, Rayner et al., 2006, Smith & Reynolds, 2002) have

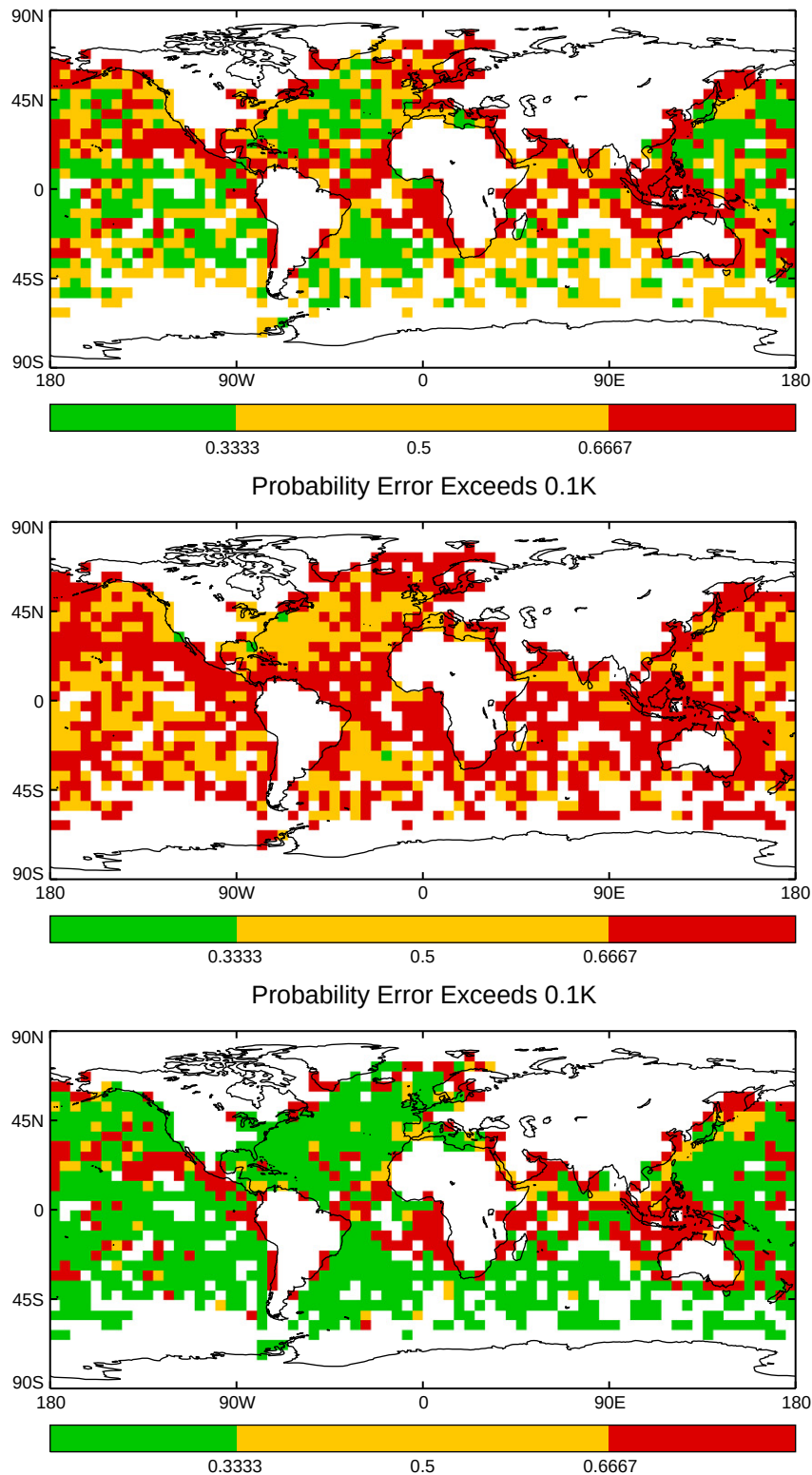


Fig. 11. Probability that measurement error exceeds 0.1 K for a weekly average 5° square for 1 to 7 June 2006. White areas indicate regions where there are no data. For normally distributed errors with standard deviation 0.1 K the probability of exceeding 0.1 K is 0.33. The top panel shows the case where the full error model is used. The middle panel shows the case where the intermediate-complexity model is used. The lower panel shows the case where the errors are assumed to be uncorrelated.

addressed the problems associated with changes in measurement method prior to 1942, but recent research (Thompson, Kennedy, Wallace et al., 2008) shows that there are likely to be uncompensated biases in SST data after this date. The evolution of observation technology has continued to the present day and large numbers of measurements are now made by drifting buoys. The comparisons

earlier in the paper showed that there is a bias between ships and drifting buoys, with ships biased warm relative to drifting buoys by around 0.18 K. Smith, Reynolds, Peterson et al. (2008) remarked that the relative bias combined with an increase in the number of drifting buoy observations would lead to an underestimate of the warming trend in global-average SST.

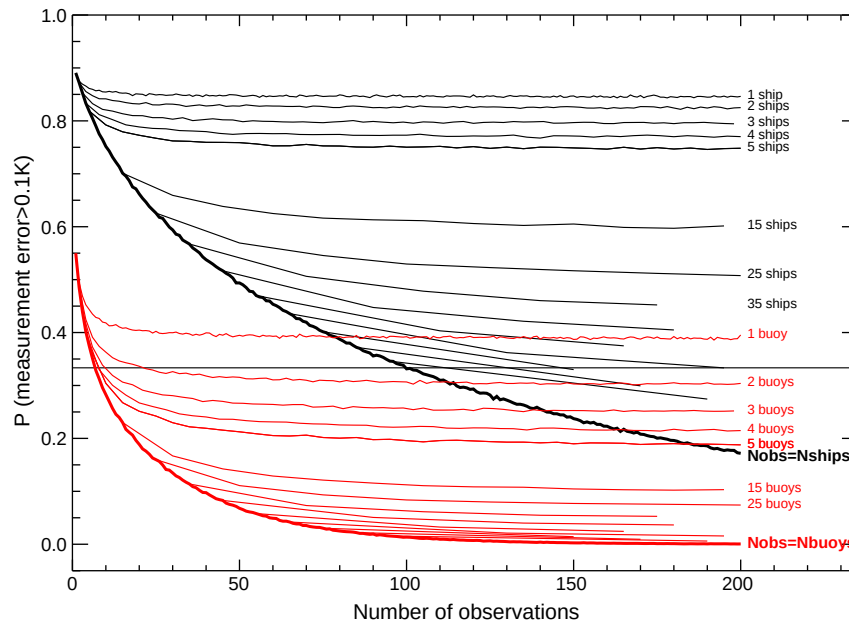


Fig. 12. Probability that the measurement error exceeds 0.1 K as a function of number of observations (x-axis), number of platforms and platform type in the case where each platform takes the same number of observations. The black lines show the results for ships. The red lines show the results for drifting buoys. The heavy black and heavy red lines that delineate the lower bounds of each distribution are the special cases where the number of observations equals the number of ships, or drifters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The ATSR series provides an estimate of long-term SST change that now spans almost 20 years. Because of the high quality of the measurements and the special attention that has been given to their stability, the ATSR retrievals do not need to be calibrated against buoy measurements. Consequently, they can be considered to be independent from the in situ record. An independent record is invaluable for corroborating the trends seen in the in situ data.

Fig. 13 shows two estimates of the monthly global average SST anomaly covering the ATSR period from 1991 to 2007. One estimate

is from the combined ATSR record, the other from HadSST2 (Rayner et al., 2006). Observations from both data sets were processed in the same way to give a 5° latitude $\times$ 5° longitude monthly gridded product and the anomalies are expressed relative to the HadSST2 climatology. Short gaps still exist in the ATSR record, but these should be reduced by future reprocessing of the data. The sudden drop in temperatures in 2007 is associated with the strong La Niña that developed during the latter half of the year. The trend in the ATSR data is 0.25 K/decade, the trend in the uncorrected HadSST2 data is

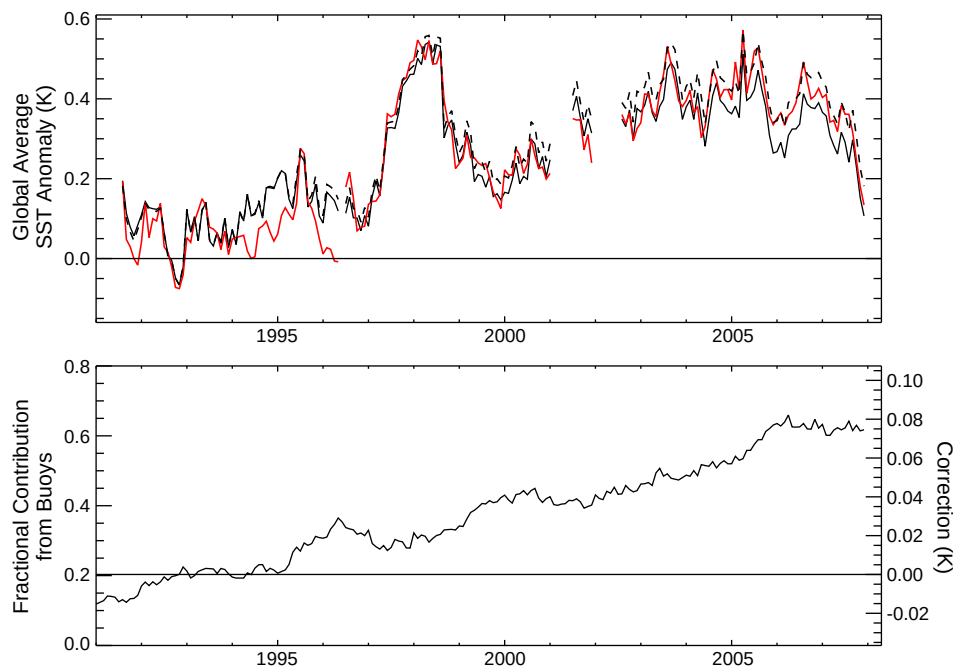


Fig. 13. (top) Co-located monthly, 5° area global average SST anomalies August 1991 to December 2007 (relative to 1961–1990) for the ATSR instruments (red) and HadSST2 (black, Rayner et al., 2006). D3 retrievals were used where they were available, but between 1992 and 1996 D2 retrievals from the ATSR-1 instrument were used. The dashed black line is an estimate of the HadSST2 time series corrected for the bias between drifting buoys and ships. (bottom) Fractional contribution of drifting buoy observations to the global average calculated by taking the area weighted average of the fraction of drifting buoy observations in each grid box. The correction is shown on the right-hand axis.

0.19 K/decade. Trends were calculated using the Ordinary Least Squares method (Davis, 1973).

Between 1992 and 1996, the ATSR series is cooler than the in situ series. During this period, the 3.7 μm channel on the ATSR-1 instrument failed so only D2 retrievals were available. The ATSR1 instrument also suffered problems with its cooling system that might have affected retrievals. There appears to be an overall warm bias in the ATSR series, which tracks the HadSST2 series most closely in the early 1990s when many observations came from ships and therefore are likely to have exhibited a warm bias.

There is a clear trend to higher temperatures in both series. There is also a general warming of the ATSR data relative to HadSST2. This could be symptomatic of the increase in the number of drifting buoy observations that went into making HadSST2 (Fig. 13, lower panel). The bias is most obvious after 2004. Many additional drifting buoys were launched during the first months of 2005 leading to a doubling of the monthly total of drifting buoy observations in a six-month period.

An adjustment for this effect has been applied to the in situ SST series shown in Fig. 13. A global-average adjustment was calculated by multiplying the fractional contribution of drifting buoy observations by the average difference between drifting buoy and ship observations, -0.18 K (see above). Because the absolute biases of the different observation types are unknown, the zero-point for the adjustment to the anomalies is arbitrary. The trend in the adjusted series allows for changes in the observing array and is therefore expected to be closer to the true trend, which is of primary interest for studies of long-term climate change. The adjustment was set to average zero over the period 1991–1995 so that the lines overlaid one another and the improvement in the trend can be most clearly seen. The resulting series, shown as a dashed line, more closely tracks the long-term changes in the ATSR data. The adjustments range from -0.02 K in 1991, to around 0.08 K in 2007. The trend in the adjusted HadSST2 data is 0.24 K/decade , which is much closer to the trend seen in the ATSR data (0.25 K/decade), than was the trend in the unadjusted data (0.19 K/decade).

It should be noted that this is not a demonstration of the stability of the ATSR data. Problems, such as the potential warm bias in the retrievals, remain in the data. Many of these will be addressed by the ATSR Reanalysis for Climate project (ARC, Merchant et al., 2008) which will provide the basis for future climate quality analyses of SST. Nevertheless, such comparisons highlight the need for an ongoing re-evaluation of the biases in the global observing system and the value that can be derived from multiple independent estimates of climate variables.

8. Summary

Pairs of AATSR observations were used to estimate the errors in the AATSR 10 arc min SST product. The errors were shown to be around 0.14 K , within the range of errors estimated by O'Carroll et al. (2008) and much smaller than for most in situ sources.

Using the AATSR as a reference data set, the error characteristics of in situ SST measurements were explored by comparing them to almost-coincident AATSR observations. When the error estimates were assumed to be independent and identically distributed the results were comparable to others in the literature. The ship observations exhibited errors of around 1.04 K ; drifting buoy observations had errors of 0.20 K ; Moored buoys, 0.21 K ; and the equatorial moorings of the TAO, PIRATA and Indian Ocean arrays, 0.12 K .

The in situ error model was then extended. In situ errors were described by a constant offset (the inter-platform error) combined with a randomly varying term (the intra-platform error). The inter- and intra-platform error characteristics for the different in situ platforms were estimated from the data. The standard deviations of inter-platform and intra-platform errors of ships were 0.71 K and 0.74 K

respectively. For drifting buoys the inter-platform and intra-platform errors were 0.29 K and 0.26 K respectively.

The distributions of inter- and intra-platform errors are non-normal. It was shown that averages of fewer than 20 independent observations are likely to deviate from normality with large errors occurring more often than might be expected if normality were assumed.

The effect of these considerations was highlighted by showing that the perceived adequacy of the in situ network for the correction of satellite biases depends on the choice of error model. Assuming that the errors are uncorrelated can potentially lead to a false impression of adequacy and an underestimate of the uncertainty.

A comparison of the long-term temperature changes in global SST estimated from the independent in situ and ATSR records was made. It provides further evidence for a previously anticipated change in bias in the in situ data arising from an increase in the number of drifting buoy observations since the 1970s. The results suggest that warming of the global oceans between 1991 and 2007 is underestimated in HadSST2 by around 0.05 K/decade . In HadSST3, a more comprehensive assessment of data biases was carried out, showing that the transition from ships to drifting buoys is only one factor that affects the stability of the modern in situ SST record. (Kennedy et al. 2011b).

Acknowledgements

The authors would like to thank Elizabeth Kent for useful comments on the manuscript. The authors were supported by the Joint DECC and Defra Integrated Climate Programme – DECC/Defra (GA01101).

References

- Barton, I. J. (2007). Comparison of in situ and satellite-derived sea surface temperatures in the Gulf of Carpentaria. *Journal of Atmospheric and Oceanic Technology*, 24, 1773–1784.
- Brasnett, B. (2008). The impact of satellite retrievals in a global sea-surface-temperature analysis. *Quarterly Journal of the Royal Meteorological Society*, 134(636), 1745–1760.
- Chiodi, A. M., & Harrison, D. E. (2006). Summertime subtropical sea surface temperature variability. *GRL*, 33, L08601. doi:10.1029/2005GL024524.
- Corlett, G. K., Barton, I. J., Donlon, C. J., et al. (2006). The accuracy of SST retrievals from AATSR: An initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. *Advances in Space Research*, 37(4), 764–769.
- Davis, J. C. (1973). *Statistics and data analysis in geology* (pp. 192). New York: John Wiley & Sons, Inc.
- Emery, W. J., Baldwin, D. J., Schlusell, P., et al. (2001). Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research—Oceans*, 106(C2), 2387–2405.
- Folland, C. K., & Parker, D. E. (1995). Correction of Instrumental biases in historical sea-surface temperature data. *Quarterly Journal of the Royal Meteorological Society*, 121(522), 319–367.
- Folland, C. K., Parker, D. E., Scaife, A. A., et al. (2006). The 2005/06 winter in Europe and the United Kingdom: Part 2 – Prediction techniques and their assessment against observations. *Weather*, 61(12), 337–346. doi:10.1256/wea.182.06.
- Folland, C. K., Reynolds, R. W., Gordon, M., et al. (1993). A study of six operational sea surface temperature analyses. *Journal of Climate*, 6, 96–113.
- Goddard, L., DeWitt, D. G., & Reynolds, R. W. (2009). Practical implications of uncertainty in observed SSTs. *GRL*, 36, L09710. doi:10.1029/2009GL037703.
- GOOS (1999). Document. "Report global physical ocean observations for GOOS/GCO – An action plan for existing bodies and mechanisms" Paris, 66, http://www.jodc.go.jp/info/ioc_doc/INF/126052eo.pdf 1999, July.
- James, R. W., & Fox, P. T. (1972). Comparative sea surface temperature measurements. *World Meteorological Organization reports on marine science affairs*, rep. 5, 336. (pp. 27): WMO.
- Keenlyside, N. S., Latif, M., Jungclauss, J., et al. (2008). Advancing decadal-scale climate prediction in the North Atlantic sector. *Nature*, 453, 84–88. doi:10.1038/nature06921.
- Kennedy, J. J., Rayner, N. A., Smith, R. O., Saunby, M., & Parker, D. E. (2011a). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. *JGR Atmospheres*, 116. doi:10.1029/2010JD015218.
- Kennedy, J. J., Rayner, N. A., Smith, R. O., Saunby, M., & Parker, D. E. (2011b). Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 2: biases and homogenisation. *JGR Atmospheres*, 116. doi:10.1029/2010JD015220.
- Kent, E. C., & Berry, D. I. (2005). Quantifying random measurement errors in Voluntary Observing Ships' meteorological observations. *International Journal of Climatology*, 25(7), 843–856. doi:10.1002/joc.1167.

- Kent, E. C., & Berry, D. I. (2008). Assessment of the Marine Observing System (ASMOS): Final report. *National Oceanography Centre Southampton research and consultancy report*, 32. (pp. 55). Southampton, UK: National Oceanography Centre Southampton <http://eprints.soton.ac.uk/50260/>
- Kent, E. C., & Challenor, P. G. (2006). Toward estimating climatic trends in SST. Part II: Random errors. *Journal of Atmospheric and Oceanic Technology*, 23(3), 476–486.
- Kent, E. C., Challenor, P. G., & Taylor, P. K. (1999). A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *Journal of Atmospheric and Oceanic Technology*, 16(7), 905–914.
- Kent, E. C., & Kaplan, A. (2006). Toward estimating climatic trends in SST. Part III: Systematic biases. *Journal of Atmospheric and Oceanic Technology*, 23(3), 487–500.
- Llewellyn-Jones, D., Edwards, M. C., Mutlow, C. T., et al. (2001). AATSR: global-change and surface temperature measurements from Envisat. *ESA Bulletin*, 105, 10–21 http://www.esa.int/esapub/bulletin/bullet105/bull105_1.pdf
- Merchant, C. J., Llewellyn-Jones, D., Saunders, R. W., et al. (2008). Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Advances in Space Research, Volume 41*(1), 1–11.
- Morrissey, M. L., & Greene, J. S. (2009). A theoretical framework for the sampling error variance for three-dimensional climate averages of ICOADS monthly ship data. *Theoretical and Applied Climatology*, 96(3–4), 235–248.
- O'Carroll, A. G., Eyre, J. R., & Saunders, R. W. (2008). Three-way error analysis between AATSR, AMSR-E and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, 25(7), 1197–1207.
- O'Carroll, A. G., Watts, J. G., Horrocks, L. A., et al. (2006). Validation of the AATSR Meteo product sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, 23(5), 711–726.
- Rayner, N. A., Brohan, P., Parker, D. E., et al. (2006). Improved analyses of changes and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century: The HadSST2 data set. *Journal of Climate*, 19(3), 446–469.
- Reynolds, R. W., Gentemann, C. L., & Corlett, G. K. (2010). Evaluation of AATSR and TMI Satellite SST Data. *Journal of Climate*, 23(1), 152–165. doi:10.1175/2009JCL132421.
- Reynolds, R. W., Rayner, N. A., Smith, T. M., et al. (2002). An improved *in situ* and satellite SST analysis for climate. *Journal of Climate*, 15(13), 1609–1625.
- Smith, T. M., & Reynolds, R. W. (2002). Bias corrections for historical sea surface temperatures based on marine air temperatures. *Journal of Climate*, 15, 73–87.
- Smith, T. M., Reynolds, R. W., Peterson, T. C., et al. (2008). Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *Journal of Climate*, 21(10), 2283–2296.
- Thompson, D. W. J., Kennedy, J. J., Wallace, J. M., et al. (2008). A large discontinuity in mid 20th century global-mean surface temperatures. *Nature*, 453, 646–649. doi:10.1038/nature06982 2008, May 29.
- Worley, S. J., Woodruff, S. D., Reynolds, R. W., et al. (2005). ICOADS Release 2.1 data and products. *International Journal of Climatology*, 25, 823–842.
- Zhang, H. -M., Reynolds, R. W., Lumpkin, R., et al. (2009). An integrated global observing system for sea surface temperature using satellites and in situ data. *Bulletin of the American Meteorological Society*, 90, 31–38. doi:10.1175/2008BAMS2577.1.
- Zhang, H., Reynolds, R. W., & Smith, T. M. (2006). Adequacy of the in situ observing system in the satellite era for climate SST. *Journal of Atmospheric and Oceanic Technology*, 23, 107–120.