

Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations

ANNE G. O'CARROLL, JOHN R. EYRE, AND ROGER W. SAUNDERS

Met Office, Fitzroy Road, Exeter, Devon, United Kingdom

(Manuscript received 13 November 2006, in final form 26 June 2007)

ABSTRACT

Using collocations of three different observation types of sea surface temperatures (SSTs) gives enough information to enable the standard deviation of error on each observation type to be derived. SSTs derived from the Advanced Along-Track Scanning Radiometer (AATSR) and Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E) instruments are used, along with SST observations from buoys. Various assumptions are made within the error theory, including that the errors are not correlated, which should be the case for three independent data sources. An attempt is made to show that this assumption is valid and that the covariances between the different observations because of representativity error are negligible. Overall, the spatially averaged nighttime AATSR dual-view three-channel bulk SST observations for 2003 are shown to have a very small standard deviation of error of 0.16 K, whereas the buoy SSTs have an error of 0.23 K and the AMSR-E SST observations have an error of 0.42 K.

1. Introduction

Sea surface temperatures (SSTs) derived from satellite-borne instruments have some advantages over traditional in situ SST measurements from buoys and ships. First, they provide a global coverage that is important in regions with sparse in situ observations. Second, some satellite instruments can give well-calibrated, accurate skin SST observations with global and temporal consistency that is not possible from in situ measurements. Satellite datasets are also beginning to span long enough periods to start being able to detect long-term changes in SST. However, it is important to understand the error characteristics of these data.

This study investigates the errors in SST observations from three different sources: infrared SSTs from the Advanced Along-Track Scanning Radiometer (AATSR), microwave SST observations from the Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E), and in situ SST observations from drifting and moored buoys. A three-way analysis is performed in which all three SST types

are collocated using data during all of 2003. Other examples of error derivation using three-way collocation include Stoffelen (1998) and Blackmore et al. (2007). These three observation types are complementary, with each having its own strengths and weaknesses.

In situ SST measurements are affected by the varying depth of measurement according to buoy type. Also, the lack of maintenance of in situ instruments, which mainly affects drifting buoys, contributes to variations in the accuracy of in situ SST observations. Additionally, there are large geographic regions not covered by the buoy network (Emery et al. 2001). Errors related to satellite-derived SST observations include cloud contamination, aerosol, and inadequacies of the retrieval process. Microwave SST retrievals can be obtained in areas that are cloudy, though not those that are precipitating, which provide a better coverage than the infrared sensors, although at generally lower spatial resolution.

An additional issue, applicable to all three observation types, is to consider the depth at which the observation is actually representative. For in situ observations, we generally assume the depth to be around 1 m, although this will vary. Infrared satellite observations retrieve the radiative skin SST, which is only 1 μm thick, whereas microwave satellite observations retrieve an SST a couple of millimeters below the skin

Corresponding author address: Anne G. O'Carroll, Met Office, Fitzroy Road, Exeter, Devon EX1 3PB, United Kingdom.
E-mail: anne.ocarroll@metoffice.gov.uk

layer of the ocean (Donlon et al. 2004). In the analyses, we consider these differences and correct for them where possible.

The three SST observation types are collocated, and statistics of the differences between the pairs of observation types at a time are computed. The standard deviations are used within a statistical method to calculate the error of each observation type, assuming uncorrelated errors. A number of experiments are performed using slightly different collocation criteria to assess whether the results are robust and, therefore, whether the various assumptions are justified.

2. Description of sea surface temperature datasets

a. Advanced Along-Track Scanning Radiometer

AATSR was launched upon the sun-synchronous *Envisat* in March 2002. AATSR has three infrared channels centered in the atmospheric windows at 3.7, 10.8, and 12 μm , plus channels in the visible–near-infrared part of the spectrum used for cloud detection. The instrument has an inclined conical scan that enables it to make observations of the surface from two different angles—nadir and forward views ($\sim 55^\circ$) from zenith—within a few minutes of each other to allow for effective correction of atmospheric absorption and aerosol. The instrument is designed to produce *dual-view* SST retrievals to better than 0.3-K accuracy in the derived SST and to give a long-term stability of better than 0.1 K decade⁻¹ (Llewellyn-Jones et al. 2001), which is required of observing systems for climate change detection purposes (Allen et al. 1994). The AATSR instrument continues the data collection of high-quality dual-view SSTs begun by Along-Track Scanning Radiometer (ATSR) ATSR-1 and ATSR-2 upon the launch of the first and second *European Remote Sensing Satellites* (*ERS-1* and *ERS-2*, respectively) in 1991 and 1995, respectively.

The measurements are most closely related to the radiative skin temperature (skin SST), which is usually cooler than the subskin by more than 0.1 K (e.g., Hasse 1963; Saunders 1967; Robinson et al. 1984; Murray et al. 2000; Donlon et al. 2002; Horrocks et al. 2003). The skin SSTs are retrieved using physically based algorithms derived from radiative transfer models (e.g., Závody et al. 1995; Merchant et al. 1999). Two types of dual-view skin SST retrieval are possible: those using all three infrared channels (D3) only during the night because of the solar contribution to the 3.7- μm channel during the day and retrievals using just the 10.8- and 12- μm channels (D2) during both day and night. The AATSR data used for this study are supplied as an averaged clear-sky radiance product to a spatial resolution of 10 arc min,

although 1-km resolution data are available from AATSR. In this study, D3 retrievals have been used [derived using the European Space Agency (ESA) “prelaunch” operational coefficients], which are expected to be the most accurate retrievals because of the availability of the extra shortwave infrared channel. The data supplied from ESA were already cloud screened using the operational system based on Závody et al. (2000).

For climate purposes, we require the bulk temperature, the temperature of the ocean at around 1 m in depth, which provides a more comparable measurement when looking at other climate datasets, such as the Hadley Centre Sea Ice and Sea Surface Temperature Analysis (HadISST) (Rayner et al. 2003), than skin SST and is more representative of the overall heat capacity of the upper layers of the ocean (Harris et al. 1998). The AATSR skin SSTs used in this study have been processed to a pseudo-bulk SST using the Fairall model (Fairall et al. 1996). The Fairall model was an extension of work by Saunders (1967) to treat both shear and convectively driven turbulence together through their contribution to the near-surface turbulent kinetic energy dissipation rate, which directly affects the cool skin-layer thickness. Further information on applying a skin effect model to (A)ATSR skin SSTs can be found in Horrocks et al. (2003). Figure 1 shows an example of nighttime AATSR bulk (D3) SSTs for July 2003.

b. In situ observations

In this study, in situ SST observations have been obtained by extracting global moored and drifting buoy SST measurements obtained via the Global Telecommunications System. The buoy SSTs are collocated to the AATSR 10 arc min cell within a 3-h time window to provide an AATSR–buoy matchup dataset. Figure 2 shows a time series of weekly mean AATSR SST–buoy SST differences for 2003. The AATSR skin and bulk SSTs used within this plot are a combination of day and night D2 and D3 retrievals and are not bias corrected. The time series shows that the weekly mean differences are well within 0.5 K of each other. An obvious anomaly, which occurs in November 2003 when the AATSR SSTs are much cooler than the buoy SST, is due to two matchups where the AATSR SSTs are around 1 K cooler than the buoy SSTs. These anomalies could be investigated in more detail using 1-km AATSR data, but this is beyond the scope of this paper. The time series shows that the AATSR bulk SSTs are warmer than the AATSR skin SSTs as expected. However, the AATSR skin SSTs are closer to the buoy SSTs than the AATSR bulk SSTs because of a 0.2-K warm

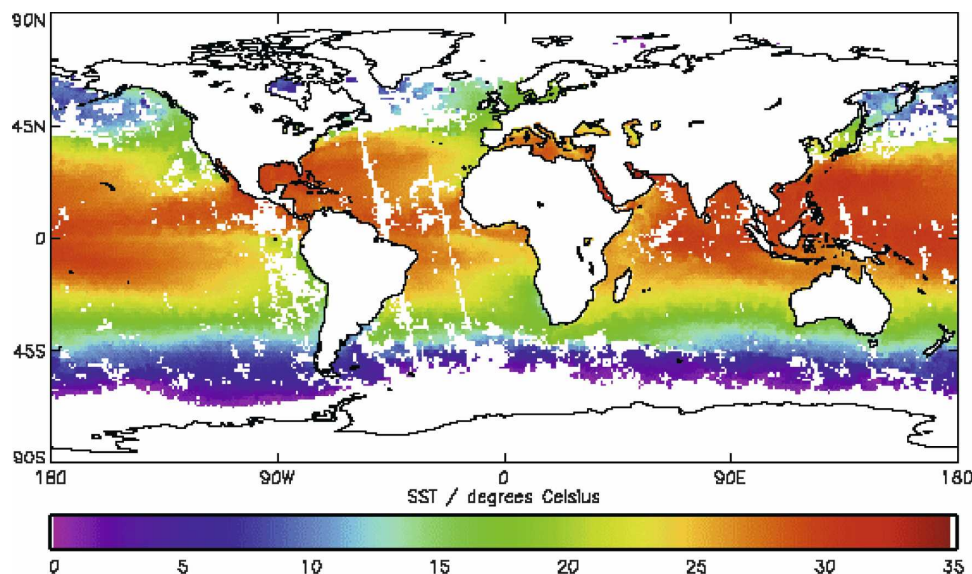


FIG. 1. Monthly mean map of nighttime AATSR bulk (D3) SSTs ($^{\circ}\text{C}$) for July 2003.

bias in the AATSR skin SSTs. This warm bias, which is especially significant in the AATSR D3 retrievals, has been observed by analyzing the AATSR–buoy matchup dataset. The matchup time window is 1 h, and the buoy “bulk” SSTs have been converted to a buoy “skin” SST using the Fairall model during nighttime only to evaluate the biases on the D2 and D3 retrievals. The D3 biases are found to be $\sim +0.21$ K with respect

to in situ observations, whereas the D2 biases are found to be $\sim +0.05$ K. More AATSR validation results against in situ SSTs and other SST datasets can be found in O’Carroll et al. (2006).

c. Advanced Microwave Scanning Radiometer

The AMSR-E instrument is onboard the sun-synchronous *Aqua* satellite that was launched in May

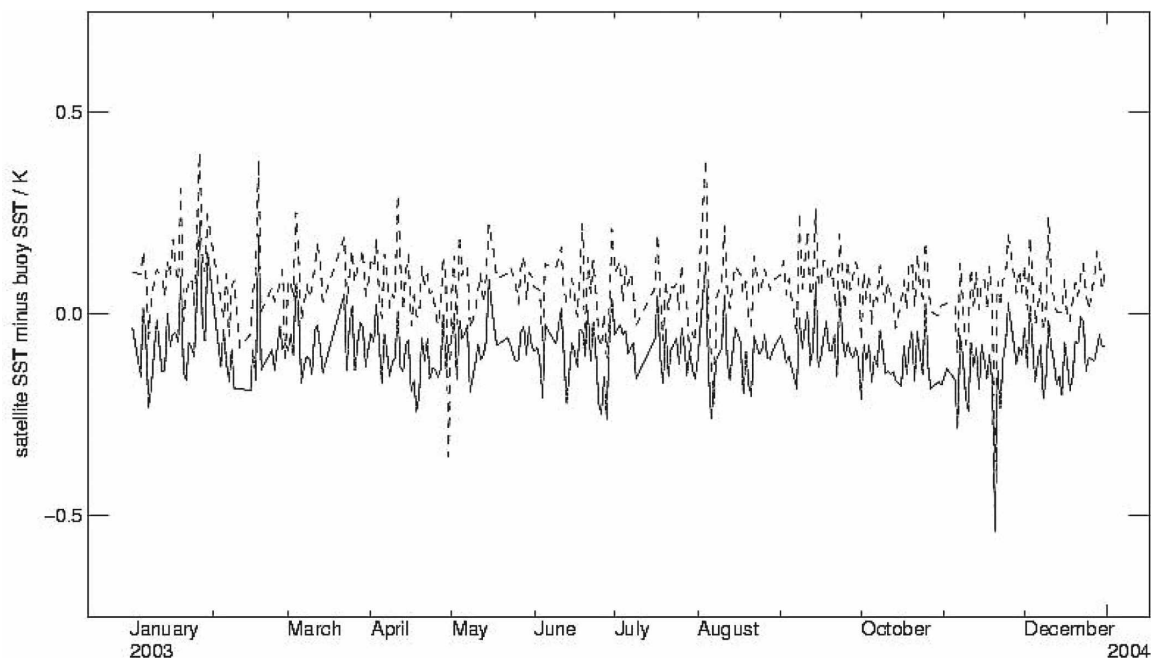


FIG. 2. Time series of statistics of weekly mean global AATSR – buoy SST differences for January–December 2003 with dual-view skin SST (solid line) and dual-view bulk SST (dashed line).

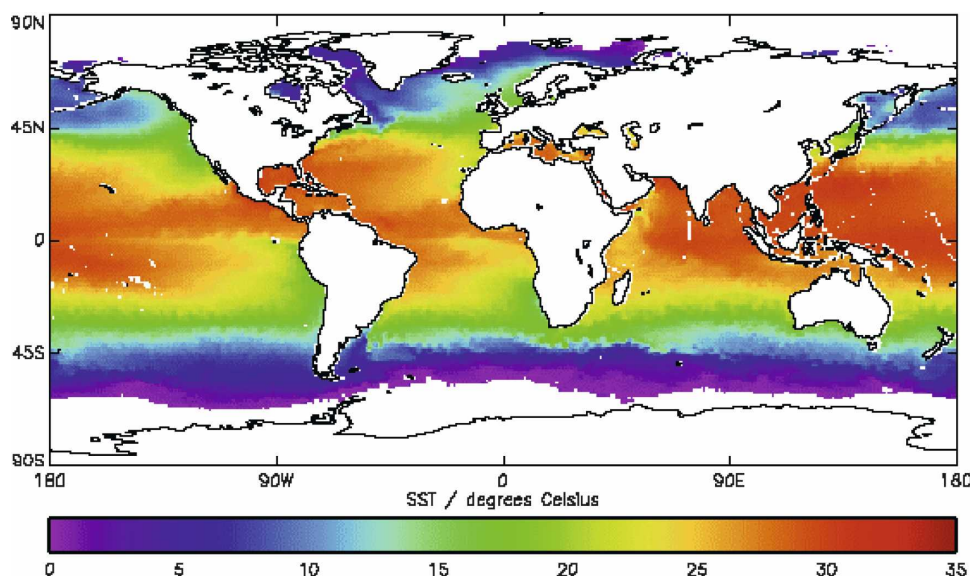


FIG. 3. Monthly mean map of AMSR-E SSTs ($^{\circ}\text{C}$) for July 2003 averaged to 1° spatial resolution.

2002. The instrument was provided to the National Aeronautics and Space Administration (NASA) by the National Space Development Agency of Japan (NASDA). Global and daily AMSR-E version 4 products of SST, wind speed, atmospheric water vapor, cloud water, and rain rate were obtained at 0.25° spatial resolution. Note that since the release of AMSR-E version 5 products, a subset of data were reprocessed. For experiment 1, the standard deviations were found to be within 0.03 K and the overall derived standard deviation of error was within 0.02 K of the version 4 products. The ability of AMSR-E to view the surface through nonprecipitating clouds is a major advantage over traditional cloud-free infrared measurements of SST. Details of the AMSR SST algorithm are given in Wentz and Meissner (2000). The global SSTs have been retrieved as daily averaged files from the Remote Sensing Systems (RSS) Web site (online at <http://www.ssmi.com>). The microwave-derived SSTs are representative of a depth of a few millimeters. RSS has applied a method for correcting a flaw in the AMSR-E calibration (more details from Wentz et al. 2003). It should be noted that RSS has applied bias corrections to the AMSR-E SSTs (additional information available online at <http://www.ssmi.com>), but the resultant SSTs remain representative of the same depth. Figure 3 shows an example of AMSR-E SSTs for July 2003.

3. Method of comparison

The AATSR bulk SSTs for 2003 used in this study have been previously validated against buoy SSTs

(Watts et al. 2004). These results show that the mean AATSR D3 minus buoy SST is 0.22 K with a standard deviation of less than 0.3 K. Additionally, the results show that there is a difference of 0.14 K between D2 and D3 retrievals at night where the D3 SSTs are warmer. Other validation exercises of AATSR SSTs have been performed using ship radiometer data (summarized in Corlett et al. 2006), which confirm the warm bias in the AATSR SSTs.

A matchup database between AATSR SSTs and buoy SSTs was produced containing collocated observations. The matchup criteria were that the collocated observations occur within 3 h of each other and that the buoy SSTs were located within the AATSR 10 arc min grid. Throughout 2003, more than 16 000 day and nighttime matchups between AATSR SSTs and buoy SSTs were obtained. A map of their distribution is shown in Fig. 4.

These collocated observations were then compared to daily AMSR-E SSTs at 0.25° spatial resolution. The AATSR SST locations were then matched to the nearest AMSR-E SST cell within which the AATSR–buoy matchup would be located.

Before statistics on the differences were calculated between these three-way daily matchups, various quality-control processes were performed on the observations. Only nighttime AATSR observations were used to reduce the effect of diurnal warming on the observations. In addition, a thermocline flag contained within the AATSR–buoy matchup database was interrogated to see if a diurnal thermocline, created in scenarios of high insolation and low wind speed, was likely.

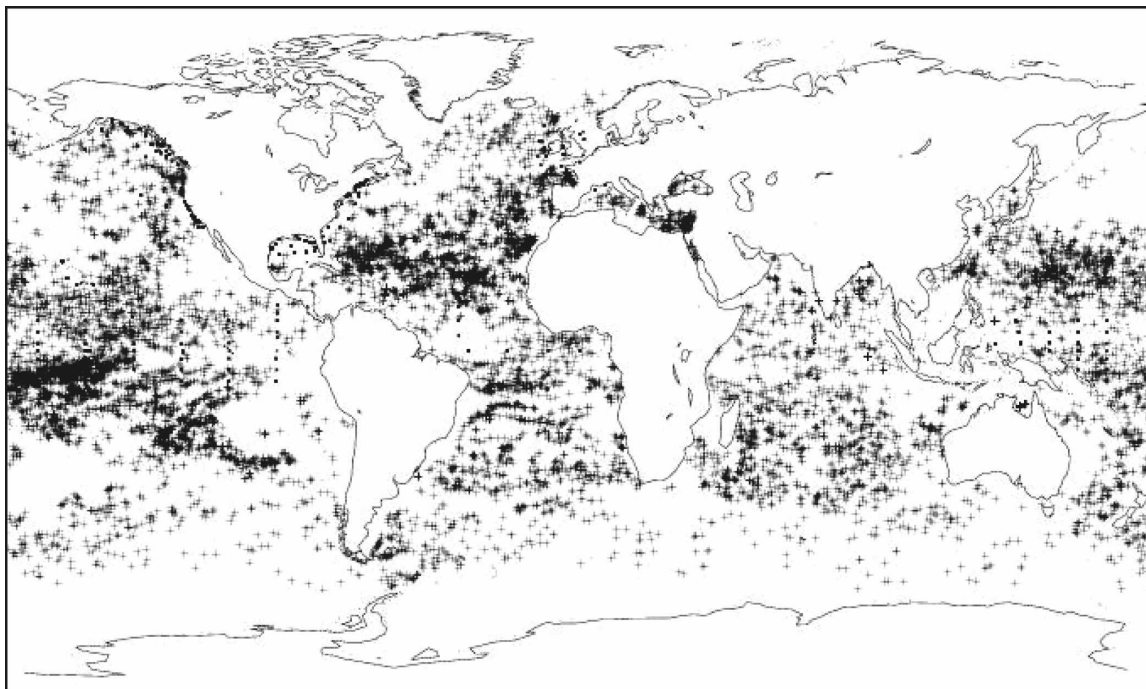


FIG. 4. Distribution of AATSR–buoy matchups throughout 2003.

This thermocline flag was created in the initial processing of the AATSR observations at the Met Office within which a diurnal thermocline model is run based on Kantha and Clayson (1994), although this procedure is only relevant during the day.

A buoy quality-control flag is contained in the AATSR–buoy matchup database, which indicates whether the buoy SST observation is deemed to be of high quality. The flag is set within the processing of the AATSR–buoy matchup database. Several processes affect whether the flag is set. First, a blacklist of known problem buoys is referenced. The list also contains details of inland buoys and lakes that are not used in the analysis. Second, during the processing of the AATSR–buoy matchups, a week's worth of buoy SSTs is compared to the Met Office NWP global SST analysis and values of observation minus analysis are computed. Buoys that showed a mean weekly bias of more than 1.2 K or a standard deviation of greater than 0.6 K compared to the NWP SST were not used within the analysis. More information on the buoy quality-control procedure is given in O'Carroll et al. (2006).

The AATSR skin SSTs contained within the matchups are bias corrected before use. The bias correction is calculated by observing the difference between nighttime AATSR skin SSTs and the buoy SSTs, which have been converted to a buoy skin SST using the Fairall model (Horrocks et al. 2003). The time difference be-

tween the selected collocated observations is less than 1 h to increase the confidence of the quality of the matchup. The derived bias correction to be applied to the D3 AATSR skin SSTs is -0.21 K. The bulk SSTs are then computed from the bias-corrected skin SSTs.

A statistical analysis is performed on these quality-controlled AATSR–buoy–AMSR matchups. The differences between the collocated AATSR and buoy SSTs are computed and the mean difference and standard deviation of all the differences are assessed using a three-sigma standard deviation test to remove outliers. The same method is applied to derive the mean difference and standard deviation between the buoy and AMSR-E SSTs, and also the AATSR and AMSR-E SSTs.

Six different experiments were performed with slightly different areas and AATSR–buoy matchup criteria. The use of buoy types is defined in Table 1. When dividing the matchups spatially, the following two different global regions have been used: region 1 (0° – 90° N, 0° – 180° W) and region 2 (90° S– 0° , 0° – 180°). By selecting these different experiment types, we can observe whether the error analysis is consistent using these different regions and matchup criteria. In addition, the method of collocation was slightly altered for experiments 7 and 8, with the rest of the criteria for the two additional experiments remaining the same as experiment 1. For the main experiments (1–6), the

TABLE 1. Definitions of subsets of the AATSR–buoy–AMSR-E matchups from which statistics are calculated.

Expt	Details of matchup scheme
1	Global, 3-h AATSR–buoy matchup cutoff, moored and drifting buoys
2	Global, 3-h AATSR–buoy matchup cutoff, moored buoys only
3	Global, 3-h AATSR–buoy matchup cutoff, drifting buoys only
4	Global, 1-h AATSR–buoy matchup cutoff, moored and drifting buoys
5	Region 1, 3-h AATSR–buoy matchup cutoff, moored and drifting buoys
6	Region 2, 3-h AATSR–buoy matchup cutoff, moored and drifting buoys
7	Same as 1, except AMSR-E SSTs interpolated to AATSR location
8	Same as 1, except AMSR-E SSTs interpolated to buoy location

AMSR-E observations are collocated to the nearest $\frac{1}{6}^\circ$ grid box of the AATSR–buoy matchup. However, for experiments 7 and 8, four AMSR-E observations are taken surrounding the AATSR matchup cells and interpolated—first to the AATSR observation location (experiment 7) and second to the buoy observation location (experiment 8).

4. Theory of error analysis

In the appendix, we derive a set of simultaneous equations for estimating the error variances σ_i^2 for observation type i (where $i = 1, 2$, or 3) for an ensemble of collocations of observation triplets:

$$\begin{aligned}
 \sigma_1^2 &= \frac{1}{2}(V_{12} + V_{31} - V_{23}) \\
 &\quad + (r_{12}\sigma_1\sigma_2 + r_{31}\sigma_3\sigma_1 - r_{23}\sigma_2\sigma_3), \\
 \sigma_2^2 &= \frac{1}{2}(V_{23} + V_{12} - V_{31}) \\
 &\quad + (r_{23}\sigma_2\sigma_3 + r_{12}\sigma_1\sigma_2 - r_{31}\sigma_3\sigma_1), \quad \text{and} \\
 \sigma_3^2 &= \frac{1}{2}(V_{31} + V_{23} - V_{12}) \\
 &\quad + (r_{31}\sigma_3\sigma_1 + r_{23}\sigma_2\sigma_3 - r_{12}\sigma_1\sigma_2), \quad (1)
 \end{aligned}$$

where V_{ij} is the variance of the difference between observation types i and j , and r_{ij} is the correlation of error between observation types i and j .

If the errors in the three observation types are un-

correlated, then $r_{ij} = 0$ for all $i \neq j$, and so Eq. (1) becomes

$$\begin{aligned}
 \sigma_1 &= \left[\frac{1}{2}(V_{12} + V_{31} - V_{23}) \right]^{1/2}, \\
 \sigma_2 &= \left[\frac{1}{2}(V_{23} + V_{12} - V_{31}) \right]^{1/2}, \quad \text{and} \\
 \sigma_3 &= \left[\frac{1}{2}(V_{31} + V_{23} - V_{12}) \right]^{1/2}. \quad (2)
 \end{aligned}$$

For Eq. (2) to be a reasonable approximation to Eq. (1), we require that the covariances of error between the observation types are small relative to the error variances. The validity of this approximation is not obvious, for the reasons discussed in the appendix. We will proceed in section 5 by tentatively making this assumption, and we will discuss its implications in section 6.

5. Error analysis

The statistics of differences between AATSR, buoy, and AMSR-E SSTs are presented in Table 2 and for the experiments 1 to 8 as described in Table 1. The results show that the mean AATSR bulk (D3) SSTs minus buoy SSTs are close to zero for all experiments. This is expected for the mean difference for this study because the AATSR SSTs have been bias corrected with respect to buoy SSTs. The standard deviation of their differences is also very low at 0.27–0.30 K.

The AMSR-E SSTs are 0.02 K cooler than the AATSR bulk (D3) SSTs and the buoy SSTs for experiment 1, and similar results are obtained for the remaining experiments, except for experiment 5. AMSR-E measures a subsurface SST, so it is expected that AMSR-E SSTs should be cooler than the bulk SST measurements. Both measurement types have been bias corrected, leading to a near-zero difference.

By inserting the standard deviations presented in Table 2 into Eqs. (2), we can calculate the standard deviation of error on each observation type for each different experiment. These errors are presented in Table 3. For experiment 1, AATSR bulk (D3) SSTs have errors of ~ 0.16 K, the buoy SSTs have errors of ~ 0.23 K, and the AMSR-E SSTs have errors of ~ 0.42 K. All of the experiments have errors that are broadly consistent with each other, suggesting that the assumption that all of the errors are uncorrelated is quite good. The smallest error on the AATSR SSTs is estimated for experiment 2 at 0.12 K and increases to its maximum of 0.16 K for experiments 1, 6, and 8.

We have neglected terms involving the correlation of

TABLE 2. Statistics on mean differences and std dev between AATSR, buoy, and AMSR-E SSTs for 2003.

Expt	Number of matchups	[AATSR bulk (D3) – AMSR-E] SST (K)		(Buoy – AMSR-E) SST (K)		[AATSR bulk (D3) – buoy] SST (K)	
		Mean diff	Std dev	Mean diff	Std dev	Mean diff	Std dev
1	2970	0.02	0.45	0.02	0.48	0.00	0.28
2	228	0.02	0.53	0.02	0.57	0.00	0.27
3	2844	0.03	0.44	0.02	0.48	0.01	0.28
4	2135	0.02	0.47	0.02	0.50	0.00	0.27
5	1001	0.03	0.45	0.06	0.51	−0.03	0.30
6	664	0.00	0.42	0.00	0.50	0.00	0.27
7	2706	0.02	0.45	0.02	0.48	0.00	0.27
8	2600	0.03	0.45	0.02	0.48	0.00	0.28

errors between different observations in Eq. (2), which is discussed further in the appendix. We need to assess whether this has an impact on the sizes of derived errors. For example, if we define the truth on the scale of the buoy observations, then there is no representativity error for buoys and so $r_{12} = r_{23} = 0$, in which the subscripts refer to 1 = AATSR, 2 = buoys, and 3 = AMSR-E. While examining Eq. (1) and assuming that $r_{12} = r_{23} = 0$, we can argue that r_{31} must be positive. Therefore, the σ_1^2 will be underestimated but σ_3^2 will be underestimated by the same amount. To further explain the positive nature of r_{31} , consider that if the average SST in the area around the buoy differs from the SST at the buoy, then AATSR and AMSR-E, which both sense some sort of areal average, will both sense this. Therefore, their differences from the “truth” (when defined as buoy location value) will tend to be positively correlated. In addition, the same effect will tend to occur when we consider the difference in depth of the AATSR and AMSR-E observations compared to buoy SSTs and that it may be even larger.

This analysis shows that we need to be confident that the assumption whereby the potential covariances of error between the observations is negligible for the results to stand. To test this assumption, we can analyze the error of representativeness by changing the method of collocation between the observation pairings. The collocation has been done by the two similar methods explained in section 3. Compiling the same difference statistics as for the other experiments gives an error of 0.15 K on the AATSR observations for experiment 7 and 0.16 K for experiment 8, which are close to the error of 0.16 K for experiment 1. Additionally, the derived errors of buoy SSTs and AMSR-E SSTs for experiments 7 and 8 support those derived from experiment 1. These results indicate that the error of representativity is small, and it is a reasonable assumption to ignore the covariances of errors between the observations when deriving Eq. (2).

Global mean difference maps of nighttime AATSR bulk D3 SSTs minus nighttime AMSR-E SSTs have been studied to look at the global variation in differences. Figure 5 shows a global map for July 2003. For most of the globe, the differences are within ± 0.5 K. However, at around 45°N , AATSR is cooler than AMSR-E by up to 2 K and at around 45°S , AATSR is warmer than AMSR-E by up to 2 K. The observed pattern is seen for other months throughout the year and is currently unexplained. However, the pattern is thought to be an artifact of the AMSR-E SSTs because AATSR comparisons against other SST datasets (O’Carroll et al. 2006; Watts et al. 2004) do not show this pattern.

The largest differences at around $\pm 45^\circ$ are also in the regions of the highest global wind speeds, most significantly in the southern oceans (see Fig. 6). Perhaps sea state is affecting the AMSR-E SST retrievals. Figure 7 shows the global AATSR minus AMSR-E difference versus binned wind speed (from the Met Office NWP model) for July 2003. The graph shows that at lower wind speeds, a larger difference between AATSR and AMSR-E SSTs occurs, with AATSR being the warmer. This relationship does not relate to the problem of larger differences between AATSR and AMSR-E at

TABLE 3. Derived standard deviation of error on each observation type for 2003 according to experiment type.

Expt	Derived std dev of error for each observation type (K)		
	AATSR bulk (D3) SST	Buoy SST	AMSR-E SST
1	0.16	0.23	0.42
2	0.12	0.24	0.51
3	0.14	0.24	0.42
4	0.15	0.23	0.45
5	0.13	0.27	0.43
6	0.16	0.22	0.45
7	0.15	0.22	0.42
8	0.16	0.23	0.42

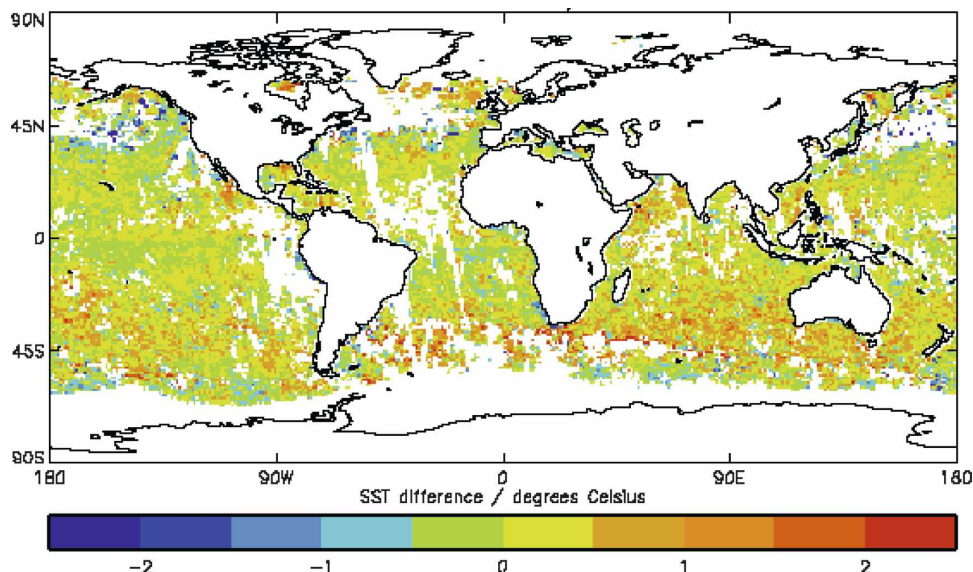


FIG. 5. Nighttime bias-corrected AATSR bulk D3 SST minus AMSR-E SST ($^{\circ}\text{C}$) for July 2003.

higher latitudes. Warmer AATSR SSTs at lower wind speeds could be due to diurnal warming effects of the sea surface. However, because only nighttime data are used in this analysis, and AATSR bulk SSTs are discarded if they are thought to be affected by diurnal warming effects according to Kantha and Clayson (1994), other effects must be present. Figure 7 also shows the results for the tropics ($\pm 30^{\circ}$ latitude). The global mean difference between nighttime bias-corrected AATSR bulk D3 SSTs and nighttime AMSR-E SSTs for July 2003 is 0.03 K, with a standard

deviation of 0.47 K. Because AMSR-E is measuring a subskin SST, that is, their SSTs will be cooler than the AATSR bulk SSTs during nighttime, the global results presented in Fig. 7 are consistent with the results of skin–bulk SST difference versus wind speed in Murray et al. (2000).

6. Conclusions

Overall, the error on nighttime AATSR bulk (D3) SST observations for 2003 was evaluated at 0.16 K,

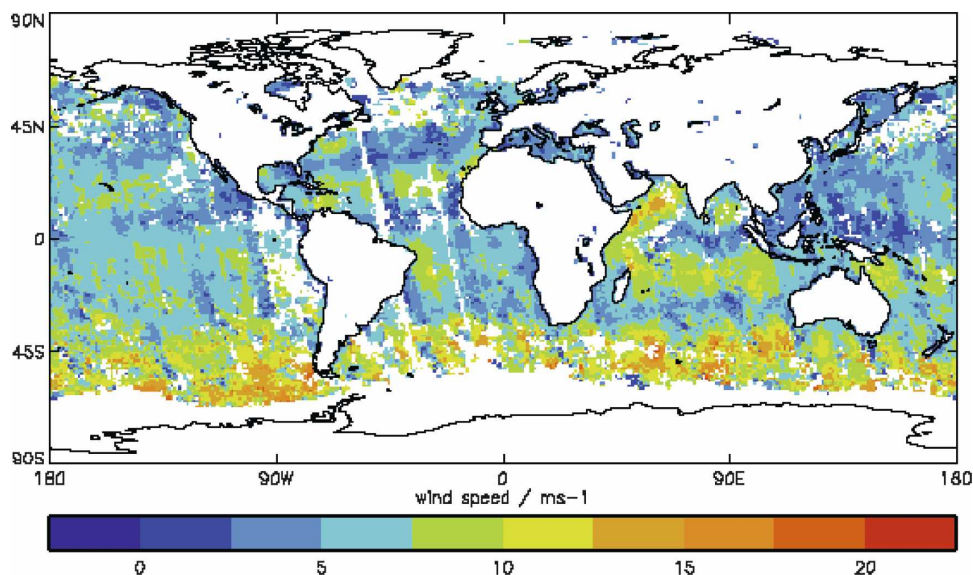


FIG. 6. Global wind speed (m s^{-1}) for July 2003 from Met Office NWP analysis averaged at 1° spatial resolution.

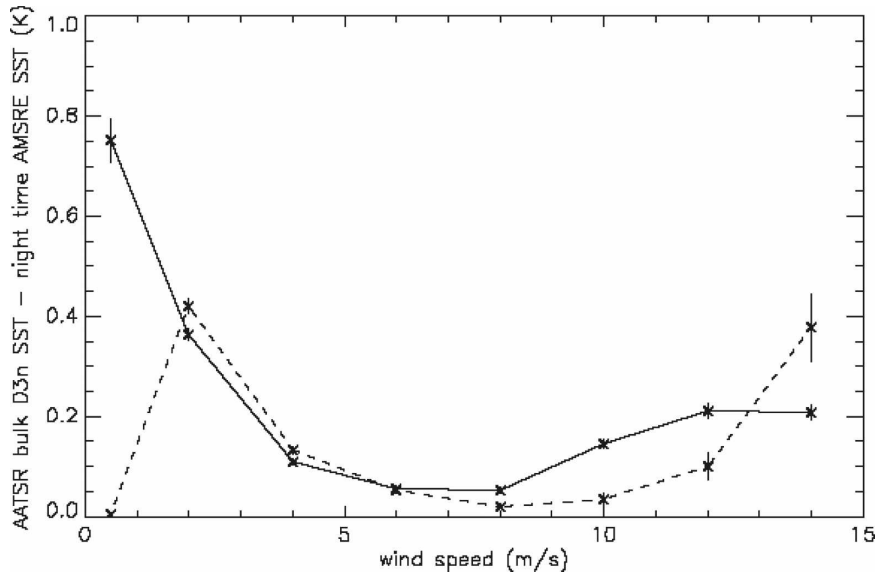


FIG. 7. Global (solid line) and tropical ($\pm 30^\circ$ latitude, dashed line) mean nighttime AATSR bulk D3 SST minus AMSR-E SST ($^\circ\text{C}$) vs binned wind speed (m s^{-1}) July 2003.

whereas the error on the collocated buoy SSTs was 0.23 and 0.42 K for AMSR-E SSTs. Varying the collocation criteria while analyzing the observation errors produce similar values of error throughout for each observation type, giving confidence both in the results and that the assumption that errors are not correlated is valid (specifically for representativity errors). There is good reason to assume that the complete (A)ATSR time series has a similar accuracy, with the exception of the ATSR-1 data after the loss of the $3.7\text{-}\mu\text{m}$ channel.

The error analysis shows how the AATSR SST retrievals are of the highest accuracy, followed by the in situ SST observations and then the AMSR-E SST observations. However, the AMSR-E SST observations do have the advantage of being able to observe the surface through cloud, giving better coverage. Because most centers are aiming toward compiling SST analyses using both infrared and microwave SST as well as in situ sources, it is essential to gain knowledge of the error characteristics of each measurement type.

Acknowledgments. This work was carried out in support of the Climate Prediction Programme, funded by the U.K. Department for Environment, Food and Rural Affairs (DEFRA). AATSR data are provided by the European Space Agency (ESA; online at <http://www.envisat.esa.int>). AMSR-E data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science REASoN DISCOVER Project and the AMSR-E Science Team (data are available online at <http://www.remss.com>).

APPENDIX

Error Analysis for Three-Way Collocation Statistics

a. Basic theory

In this section, we develop the theory relevant to the analysis of statistics of differences between collocated observations of three different types.

The error in observation x_i , of type i , can be expressed as

$$x_i = x_T + b_i + \varepsilon_i, \quad (\text{A1})$$

where x_T is the true value of variable x , b_i is the bias (mean error) in the observation, and ε_i is the random error in the observation (which, by definition, has zero mean but may be non-Gaussian).

We shall return below to the nontrivial question of what we mean by the “true value of the variable.” For the present, we will only assume that we have a consistent definition of the true value for comparison with each observation.

For a set of three collocated observations, of types $i = 1, 2$, and 3 , we can write the following corresponding set of equations for their errors:

$$\begin{aligned} x_1 &= x_T + b_1 + \varepsilon_1, \\ x_2 &= x_T + b_2 + \varepsilon_2, \quad \text{and} \\ x_3 &= x_T + b_3 + \varepsilon_3. \end{aligned} \quad (\text{A2})$$

Now consider each set of three observations as three sets of pairs. The difference between observations i and j is given by

$$x_i - x_j = b_i - b_j + \varepsilon_i - \varepsilon_j. \quad (\text{A3})$$

For an ensemble of such sets of observations, the mean difference between observations of type i and j is

$$b_{ij} = \bar{x}_i - \bar{x}_j = b_i - b_j, \quad (\text{A4})$$

and the variance of the difference between these two observation types is

$$\begin{aligned} V_{ij} &= [(\bar{x}_i - \bar{x}_j) - (\bar{b}_i - \bar{b}_j)]^2 = (\bar{\varepsilon}_i - \bar{\varepsilon}_j)^2 \\ &= \bar{\varepsilon}_i^2 + \bar{\varepsilon}_j^2 - 2\bar{\varepsilon}_i\bar{\varepsilon}_j. \end{aligned} \quad (\text{A5})$$

Therefore,

$$V_{ij} = \sigma_i^2 + \sigma_j^2 - 2r_{ij}\sigma_i\sigma_j, \quad (\text{A6})$$

where σ_i^2 is the variance of the error in observation type i , and r_{ij} is the correlation of error between types i and j .

Stating Eq. (A6) explicitly for the following three sets of observation pairs:

$$\begin{aligned} V_{12} &= \sigma_1^2 + \sigma_2^2 - 2r_{12}\sigma_1\sigma_2, \\ V_{23} &= \sigma_2^2 + \sigma_3^2 - 2r_{23}\sigma_2\sigma_3, \quad \text{and} \\ V_{31} &= \sigma_3^2 + \sigma_1^2 - 2r_{31}\sigma_3\sigma_1. \end{aligned} \quad (\text{A7})$$

The three simultaneous Eqs. (A7) can be solved to give the variance of error in each observation type as follows:

$$\begin{aligned} \sigma_1^2 &= \frac{1}{2}(V_{12} + V_{31} - V_{23}) \\ &\quad + (r_{12}\sigma_1\sigma_2 + r_{31}\sigma_3\sigma_1 - r_{23}\sigma_2\sigma_3), \\ \sigma_2^2 &= \frac{1}{2}(V_{23} + V_{12} - V_{31}) \\ &\quad + (r_{23}\sigma_2\sigma_3 + r_{12}\sigma_1\sigma_2 - r_{31}\sigma_3\sigma_1), \quad \text{and} \\ \sigma_3^2 &= \frac{1}{2}(V_{31} + V_{23} - V_{12}) \\ &\quad + (r_{31}\sigma_3\sigma_1 + r_{23}\sigma_2\sigma_3 - r_{12}\sigma_1\sigma_2). \end{aligned} \quad (\text{A8})$$

If the errors in the three observation types are uncorrelated, then $r_{ij} = 0$ for all $i \neq j$. Therefore, Eq. (A8) becomes

$$\begin{aligned} \sigma_1^2 &= \frac{1}{2}(V_{12} + V_{31} - V_{23}), \\ \sigma_2^2 &= \frac{1}{2}(V_{23} + V_{12} - V_{31}), \quad \text{and} \\ \sigma_3^2 &= \frac{1}{2}(V_{31} + V_{23} - V_{12}). \end{aligned} \quad (\text{A9})$$

This allows us to estimate the error variance in the three different observations types from the observation difference statistics. However, the validity and accuracy of this method depends crucially on the assumption of the independence of errors, which is examined further in the following section.

b. Some comments on the concept of the “true value” of a variable and the correlation of observation error

Equation (A9) is potentially very powerful because it offers a method of separating out the (usually elusive) error variances of different observations types, given statistics of the observation differences, which are easily obtained from samples of real data. However, the equation is also problematic because it suggests that this separation can be done without carefully defining what we mean by error in each observation, that is, what we mean by the true value of the variable. If, as in this paper, we are concerned with observations of a geophysical variable made by three different measurement techniques, then we might expect that our estimate of the error variance for observations made using any of these techniques should be dependent on how we define the true value, for example, whether it is a point value or an areal average. The true value must be of some variable (e.g., the temperature either at some level or averaged over some layer) at some space–time sampling. Therefore, the discussion applies equally to differences in the measurement type as it does to differences in space–time sampling. However, Eq. (A9) suggests that we do not have to define the true value.

The resolution of this paradox lies in the step from Eqs. (A8) to (A9), where we neglect the potential covariances of error between the observations. For the measurement errors themselves, it is usually reasonable to assume that the errors in measurements made by totally independent techniques will be truly independent. However, this is not the only source of error. We must also consider the “error of representativeness,” which captures the difference between the value of the variable on the space–time scale on which it is actually measured and its value on the space–time scale on which we wish to analyze it. For a single observation, the latter scale can be chosen arbitrarily, but whenever we wish to compare two observations, it must be taken into account.

As we change our definition of the true value of the variable, we change the value of the error variance of the observation. However, we also change the values of the error covariances between observations; thus Eq. (A8) continues to hold whatever our definition of the true value might be.

Therefore, Eq. (A8) will always be valid, but Eq. (A9) will only be a reasonable approximation to it in certain circumstances. Our ability to step between Eqs. (A8) and (A9) assumes that we are using a space–time scale for the “true value”; which, although we may avoid defining it precisely, is such that the covariances of the errors of representativeness on this scale are negligible when compared with the error variances. In this paper, we proceed tentatively on the assumption that these covariances are negligible, but we also make analyses to determine if this assumption is valid.

REFERENCES

- Allen, M. R., C. T. Mutlow, G. M. C. Blumberg, J. R. Christy, R. T. McNider, and D. T. Llewellyn-Jones, 1994: Global change detection. *Nature*, **370**, 24–25.
- Blackmore, T. A., A. G. O'Carroll, R. W. Saunders, and H. H. Aumann, 2007: Numerical weather prediction: A comparison of sea surface temperature from the AATSR and AIRS instruments. Met Office Tech. Rep. 499, 21 pp.
- Corlett, G. K., and Coauthors, 2006: The accuracy of SST retrievals from AATSR: An initial assessment through geophysical validation against in situ radiometers, buoys and other SST datasets. *Adv. Space Res.*, **37**, 764–769.
- Donlon, C. J., P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369.
- , L. Nykjaer, and C. Gentemann, 2004: Using sea surface temperature measurements from microwave and infrared satellite measurements. *Int. J. Remote Sens.*, **25**, 1331–1336.
- 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. *J. Geophys. Res.*, **106** (C2), 2387–2405.
- Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, 1996: Cool-skin and warm-layer effects on sea surface temperature. *J. Geophys. Res.*, **101**, 1295–1308.
- Harris, A. R., C. J. Merchant, B. Candy, C. A. Clayson, M. J. Murray, A. M. Závody, and M. R. Allen, 1998: Satellite retrievals of sea surface temperature for long-term climate monitoring. Preprints, *Ninth Conf. on Satellite Meteorology and Oceanography*, Paris, France, Amer. Meteor. Soc., 162–165.
- Hasse, L., 1963: On the cooling of the sea surface by evaporation and heat exchange. *Tellus*, **15**, 363–366.
- Horrocks, L. A., B. Candy, T. Nightingale, R. W. Saunders, A. G. O'Carroll, and A. R. Harris, 2003: Parameterizations of the ocean skin effect and implications for satellite-based measurement of sea-surface temperature. *J. Geophys. Res.*, **108**, 3096, doi:10.1029/2002JC001503.
- Kantha, L. H., and C. A. Clayson, 1994: An improved mixed layer model for geophysical applications. *J. Geophys. Res.*, **99** (C12), 25 234–25 266.
- Llewellyn-Jones, D., M. C. Edwards, C. T. Mutlow, A. R. Birks, I. J. Barton, and H. Tait, 2001: AATSR: Global-change and surface-temperature measurements from ENVISAT. *ESA Bull.*, **105**, 10–21.
- Merchant, C. J., A. R. Harris, M. J. Murray, and A. M. Závody, 1999: Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modeling and interalgorithm comparison. *J. Geophys. Res.*, **104** (C10), 23 565–23 578.
- Murray, M. J., M. R. Allen, C. J. Merchant, A. R. Harris, and C. J. Donlon, 2000: Direct observations of skin-bulk SST variability. *Geophys. Res. Lett.*, **27**, 1171–1174.
- O'Carroll, A. G., J. G. Watts, L. A. Horrocks, R. W. Saunders, and N. A. Rayner, 2006: Validation of the AATSR Meteo product sea surface temperature. *J. Atmos. Oceanic Technol.*, **23**, 711–726.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Robinson, I. S., N. C. Wells, and H. Charnock, 1984: Review article: The sea surface thermal boundary layer and its relevance to the measurement of sea surface temperature by airborne and spaceborne radiometers. *Int. J. Remote Sens.*, **5**, 19–45.
- Saunders, P. M., 1967: The temperature at the ocean-air interface. *J. Atmos. Sci.*, **24**, 269–273.
- Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, **103** (C4), 7755–7766.
- Watts, J. G., A. G. O'Carroll, and R. W. Saunders, 2004: Validation of the first two years of the AATSR Meteo product sea-surface temperatures at the Met Office. *Proc., 2004 Envisat and ERS Symp.*, Salzburg, Austria, European Space Agency, ESA SP-572, 2A4.
- Wentz, F., and T. Meissner, 2000: Algorithm Theoretical Basis Document (ATBD), version 2. AMSR Ocean Algorithm, RSS Tech. Proposal 121599A-1, 66 pp.
- , C. Gentemann, and P. Ashcroft, 2003: On-orbit calibration of AMSR-E and the retrieval of ocean products. Preprints, *12th Conf. on Satellite Meteorology and Oceanography*, Long Beach, CA, Amer. Meteor. Soc., P5.9. [Available online at <http://ams.confex.com/ams/pdfpapers/56760.pdf>.]
- Závody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones, 1995: A radiative transfer model of sea surface temperature retrieval for the along-track scanning radiometer. *J. Geophys. Res.*, **100**, 937–952.
- , —, and —, 2000: Cloud clearing over the ocean in the processing of data from the along-track scanning radiometer (ATSR). *J. Atmos. Oceanic Technol.*, **17**, 595–615.