Determination of AATSR Biases Using the OSTIA SST Analysis System and a Matchup Database

J. D. STARK, C. DONLON, AND A. O'CARROLL

Met Office, Exeter, United Kingdom

G. CORLETT

Space Research Centre, University of Leicester, Leicester, United Kingdom

(Manuscript received 5 March 2007, in final form 16 January 2008)

ABSTRACT

Sea surface temperature (SST) analyses are produced on a daily basis at the Met Office using the Operational SST and Sea Ice Analysis (OSTIA) system. OSTIA uses satellite SST data, provided by international agencies via the Global Ocean Data Assimilation Experiment (GODAE) High-Resolution SST Pilot Project (GHRSST-PP) regional/global task sharing (R/GTS) framework, which includes an estimate of bias error (available online at http://www.ghrsst-pp.org). The OSTIA system produces a foundation SST estimate (SST_{fnd}), which is the SST that is free of diurnal variability, at a resolution of 1/20° (~6 km). Global coverage outputs are provided each day in GHRSST-PP L4 netCDF format. The verification and intercomparison of the OSTIA analysis, with observations and analyses, has revealed a cold bias of approximately 0.1 K in the OSTIA outputs. Because OSTIA uses the operational 1-km *Envisat* Advanced Along-Track Scanning Radiometer (AATSR) ATS_NR_2P data [via the GHRSST-PP/European Space Agency (ESA) Medspiration Project, available online at http://www.medspiration.org] as a reference dataset for bias adjustment of other satellite data, the AATSR data were identified as the likely cause of the observed bias. To test this, a series of experiments were carried out in June 2006 using the Medspiration AATSR observations in which the Single Sensor Error Statistics (SSES) bias estimate was assigned fixed magnitudes of 0.0, 0.05, 0.15, and 0.2 K. The authors find that the AATSR data have approximately zero bias relative to in situ buoys. Because AATSR measures the SST skin temperature (SST_{skin}) and was given a mean global SST_{skin} deviation of −0.17 K (based on in situ radiometer data), this result suggests that ATS_NR_2P SST_{skin} data have a warm bias of 0.17 K. Using a matchup database of near-contemporaneous 10 arc min AATSR and in situ data, the authors find that the AATSR SST_{skin} dual- and triple-window retrievals have a warm bias of 0.14 and 0.17 K, respectively, between August 2002 and July 2006. The results of the experiments confirm that the current Medspiration SSES bias correction provided with the Medspiration AATSR L2P observations is poorly specified. The database was not configured to test the relationship between the cloud proximity confidence value and the AATSR bias error. Based on the matchup database and reanalysis results, the authors suggest that Medspiration be modified to use an SSES bias estimate of 0.17 K for all category 2–6 proximity confidence values for the current AATSR dual-view SST ATS_NR_2P products to provide a correct SST_{skin} estimate. In response to the results presented in this study, operational changes have been made to the Medspiration processing, which improve the bias estimates provided in the AATSR data. The authors suggest that a concerted effort be invested to develop the most appropriate SSES for the AATSR class of sensors that have specific characteristics that must be included in the SSES estimation scheme. The main elements of such a scheme are presented in this paper.

1. Introduction

Satellite sea surface temperature (SST) measurements are used in many applications because they provide a synoptic view of the dynamic thermal character of the ocean's surface. SST measurements are fundamentally important to agencies and institutions tasked with the study of operational weather and ocean forecasting, climate variability, and military operations. Accurate maps of SST are arguably one of the best climate indicators in their own right. Statistical seasonal forecasts are based on predictors derived from tropical At-

Corresponding author address: John Stark, Met Office, FitzRoy Road, Exeter, Devon EX1 3PB, United Kingdom.
E-mail: john.stark@metoffice.gov.uk

DOI: 10.1175/2008JTECHO560.1

© 2008 American Meteorological Society

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

lantic and tropical Pacific SST indices, and an accurate time series of SST is required to initialize dynamical seasonal forecasts using coupled ocean–atmosphere seasonal prediction models. Numerical weather prediction (NWP) systems require SST as a bottom boundary condition, which can, in certain cases, have an important impact on the forecasts, for example, fog formation, cyclogenesis, hurricane intensity, and storm track and air–sea flux calculations. Modern ocean forecasting systems provide a full four-dimensional description of the ocean at various vertical and horizontal spatial resolutions. SST data are used as a boundary condition and in data assimilation schemes within these model systems. For these real-time applications, SST data must be accurate (e.g., mean bias less than 0.1 K and root-mean-square error less than 0.6 K) and available in a timely manner (e.g., less than 6 h from measurement) from operationally robust systems.

In response to the increasing demand for accurate high-resolution SST (Smith 2000), the Met Office has recently developed a new Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Stark et al. 2007) system that provides global products at 1/20° (~6 km) resolution each day. OSTIA has been designed to provide an estimate of the foundation SST (SST_{fnd}; see Donlon et al. 2002), which is free of diurnal variability, and to include both cool-skin and warm-layer diurnal effects. The system uses a multiscale optimal interpolation (OI) scheme to combine infrared and microwave satellite, and in situ SST measurements (Martin et al. 2007). The full range of data sources is outlined in Stark et al. (2007). The scheme is based on the analysis correction method introduced by Lorenc et al. (1991), which provides an efficient means of calculating the OI solution using an iterative procedure. The scheme requires a priori estimates of the error covariances in the observations and background field. These error covariances are estimated using several years of the Met Office Forecasting Ocean Assimilation Model (FOAM) SST data. OSTIA makes extensive use of L2P data products, provided within the framework of the Global Ocean Data Assimilation Experiment (GODAE) High-Resolution SST Pilot Project (GHRSST-PP; available online at http://www.ghrsst-pp.org). [OSTIA SST_{fnd} products can be accessed online (http://ghrsst-pp.metoffice.com/pages/latest_analysis/ostia.html).] Diagnostic data and SST anomaly data can also be accessed at this site.

2. Advanced Along-Track Scanning Radiometer data used by OSTIA

The OSTIA system uses Advanced Along-Track Scanning Radiometer (AATSR) data supplied by Medspiration as a reference dataset to correct biases in the other satellite data it uses. Medspiration supplies the AATSR data as an SST product with ancillary metadata. However, it is useful to outline how these data are formed from the infrared radiance measurements because it will be important in the discussion later. For daytime observations, a two-channel SST algorithm (D2; at 11 and 12 μm) is used; at night, the 3.7-μm channel may also be used, giving a three-channel (D3) retrieval. The AATSR instrument measures the radiance in two view positions: nadir and a forward-viewing angle. The two views and two or three channels are combined by the SST algorithm to give the dual-view D2 and dual-view D3 SSTs. In some circumstances, the nadir-only measurement may be used, giving nadir-only two-channel (N2) and nadir-only three-channel (N3) SSTs. The Medspiration product uses only the dual-view (D2 and D3) measurements.

The AATSR retrieves a measurement of SST_{skin}, which is on average ~0.17 K cooler than the SST_{fnd} because of a cool-skin temperature deviation [based on extensive studies using in situ radiometer data; Donlon et al. (2002)]. Because OSTIA computes an estimate for the SST_{fnd}, but AATSR provides an estimate of the SST skin, AATSR data must be adjusted to compensate for a mean SST_{skin} temperature deviation. To adjust the AATSR to represent SST_{fnd} we use

$$SST_{fnd} = SST_{AATSR} - SSES_{bias} + \Delta T_{skin}, \qquad (1)$$

where SST_{fnd} is the estimate of foundation SST used for the OI analysis bias correction, SST_{AATSR} is the SST derived from the AATSR without adjustment, $SSES_{bias}$ is the AATSR bias estimate, and ΔT_{skin} is the cool-skin deviation. However, at the time this work was conducted, no compensation for the cool-skin effect using Eq. (1) was implemented in either the operational system or the reanalysis system, that is, $\Delta T_{skin} = 0$. Here, ΔT_{skin} and the $SSES_{bias}$ area were treated together as a single quantity ΔT_{bias}, where $\Delta T_{bias} = SSES_{bias} - \Delta T_{skin}$ is the total offset applied to the AATSR data.

Figure 1 shows results from the verification studies using in situ observation datasets [quality-controlled ship and buoy data from global task sharing (GTS) feeds] that have revealed a global cold bias of approximately 0.1 K in the OSTIA SST_{fnd} relative to the in situ data. Compared to the Met Office's Hadley Centre Sea Ice and Sea Surface Temperature (HadISST; Rayner et al. 2003) SST monthly products, Table 1 shows that OSTIA SST_{fnd} has a global cold bias of 0.23 K. Note that the HadISST dataset represents a blend of SST observations within the upper layer of the ocean rather than a depth-specific SST estimate. However, the OSTIA SST_{fnd} bias appears to have little spatial structure outside

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

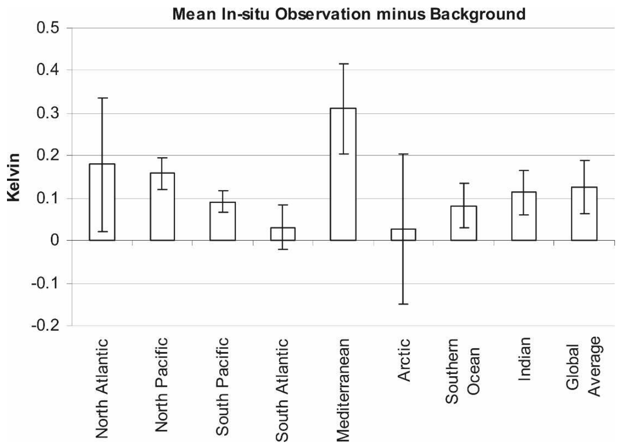

FIG. 1. Verification of OSTIA using in situ observations compared to the previous days' SST from OSTIA, showing the mean value for several subareas for May 2006. The error bars indicate the uncertainty in the mean, based on the std dev of the daily values. Note that the number of observations varies substantially between regions with very limited coverage in the Mediterranean Sea and the Arctic Ocean.

of the most dynamic ocean areas, which is shown in Fig. 2. Significant differences at high latitudes are likely to be associated with the different treatment of sea ice within the HadISST and OSTIA systems, and contribute little to the area-weighted mean statistic reported here.

The likely cause of OSTIA SST_{fnd} bias is the use of *Envisat* AATSR dual-view SST_{skin} data as a reference dataset by OSTIA. The OSTIA system ingests *Envisat* AATSR data in near–real time on a daily basis via the GHRSST-PP European Regional Data Assembly Centre (RDAC), implemented by the European Space Agency (ESA) Medspiration project (available online at http://www.medspiration.org) in GHRST-PP L2P format. The OSTIA preprocessor system first applies a GHRSST-PP sensor-specific Single Sensor Error Statistics (SSES) error estimate bias supplied with the data to AATSR ATS_NR_2P 1-km observations (the SSES bias is subtracted from the AATSR data on a pixel-by-pixel basis to compensate for known persistent measurement biases). The SSES-corrected ATS_NR_2P data are then used to adjust each GHRSST-PP L2P satellite data type used within the OSTIA analysis [which currently includes SST estimates from Advanced Very High Resolution Radiometer (AVHRR); Meteosat Second Generation (MSG); Spinning Enhanced Visible and Infrared Imager (SEVIRI); Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E); and Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI)]. This is done by deriving near-contemporaneous matchups in time and space between the reference set of in situ and AATSR data and each satellite data type in turn. Consequently, AATSR data have a significant impact on the bias error of the OSTIA analysis output.

TABLE 1. Area-weighted monthly mean OSTIA minus HadISST for selected months in 2006.

Month (2006)	Mean	RMS
April	−0.24 K	0.56 K
May	−0.23 K	0.55 K
June	−0.24 K	0.58 K

This paper describes a series of experiments conducted using the OSTIA analysis system to reverse engineer the most appropriate SSES to be applied to Medspiration's L2P ATS_NR_2P data and a more complete analysis of a matchup database (MDB) with average 10 arc min area SST AATSR records. The purpose is not to provide a definitive set of values for AATSR SSES but instead to highlight the need for urgent action to properly define the most appropriate SSES for AATSR in a real-time operational manner. Section 3 describes the experiments that were conducted using the OSTIA system to investigate AATSR SSES. Section 4 presents our discussion and describes the sensor-specific issues that should be considered when deriving SSES for AATSR data. In section 5 other issues are raised regarding a new format product. Finally, in section 6, we present our conclusions and recommendations.

3. Investigation into AATSR biases

The Medspiration L2P data are supplied with a cloud proximity confidence flag, in which each pixel has a value between 0 and 6. Lower numbers indicate close proximity to cloud and, therefore, potential cloud contamination from thin or diffuse undetected cloud (Donlon et al. 2006). In the AATSR L2P data supplied by Medspiration, the SSES bias is set to 0.19 K for all points with a cloud proximity confidence of 4, 5, and 6 (good data), and to 0.33 K when the proximity confidence is 3 or 2 (suspected cloud contamination). However, the vast majority (approximately 90%) of AATSR data points with a valid SST are flagged with values of 2 or 3, giving mean SSES biases of approximately 0.3 for all data with a quality better than or equal to 2 (OSTIA uses AATSR data when the proximity confidence flag is greater than 1). The distance to cloud proximity confidence thresholds were updated by Medspiration on 9 August 2006 (based on discussions at

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

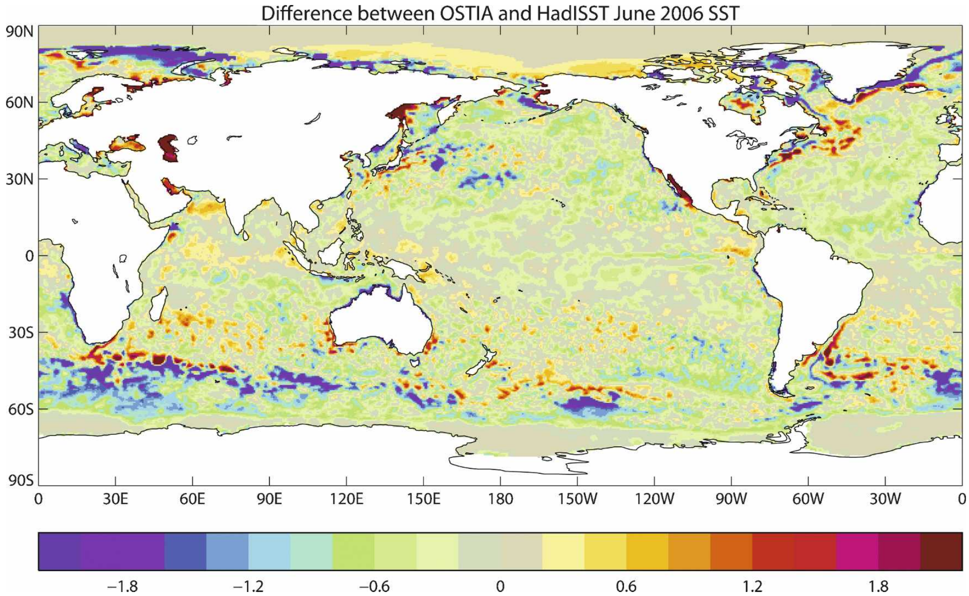

FIG. 2. SST difference (K) between monthly mean OSTIA and HadISST analysis for June 2006.

the third Medspiration collocation meeting in June 2006 in Frascati, Italy), but the SSES bias was not changed. A summary overview of AATSR data suggests that following this change, a similar fraction of AATSR L2P data will remain flagged a confidence 1 or 2 (low).

The Medspiration AATSR L2P SSES data used in this study are shown in Table 2. The proximity confidence thresholds (but not the bias or standard deviation error estimate) were updated by Medspiration on 5 June 2006 [as requested and agreed to at the third Medspiration collocation meeting, sponsored by the Consiglio Nazionale delle Ricerche (CNR), in Frascati, Italy, in June 2006], but no significant change in the daily mean output of the OSTIA system was observed.

It is known that the GHRSST-PP SSES specification for AATSR data is based on an analysis of a matchup database containing in situ observations collocated in space and time with average 10 arc min area SST AATSR records. Recent work suggests that the SSESs derived from the 10 arc min AATSR data are significantly different from an SSES computed using 1-km ATS_NR_2P data. This was considered to be the best available database for SSES specification at the time. Furthermore, the methodology used to compute SSES was based on the characteristics of the AVHRR sensor, and the analysis was very basic. Consequently, the resulting AATSR SSES analysis did not include issues specific to the AATSR, such as problem SSTs resulting from parallax effects around the leading and trailing edges of cloud in the flight direction, D2–N2 differences (required for aerosol flagging), and the impact of using coarse-resolution average fields instead of 1-km data to derive the SSES. These effects must be considered properly when specifying AATSR SSES. This highlights a need for a concerted effort to properly define the AATSR SSES as soon as possible.

a. *Reanalysis experiments*

To investigate the specification of Medspiration L2P AATSR SSES bias and their impact on the OSTIA SST_{fnd} output, a single-month reanalysis using the OSTIA system was performed. June 2006 was chosen as the reanalysis period because it was the most recent month for which HadISST data were available as a comparison dataset. This period uses the AATSR SST retrieval coefficients, which became operational on 7 December 2005. The AATSR SSES bias estimates provided by Medspiration were ignored, and four separate reanalysis runs were performed using altered AATSR

TABLE 2. Values of the cloud proximity confidence flag and associated bias and std dev SSES for Medspiration AATSR L2P datasets used in this study.

Cloud proximity confidence	Bias (K)	Std dev (K)
2 (bad)	0.33	0.6
3 (suspect)	0.33	0.3
4 (acceptable)	0.19	0.3
5 (excellent)	0.19	0.3
6 (suspect, cold skin, upwelling, riverine input, etc.)	0.19	0.3

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

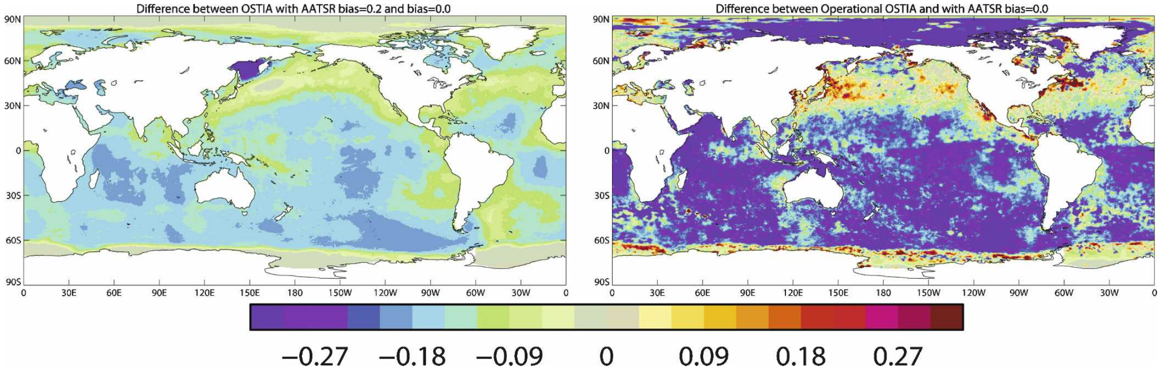

FIG. 3. Mean June SST difference between OSTIA runs. (left) The difference between using AATSR biases of 0.2 and 0.0 K. The area-weighted global mean is −0.14 ± 0.05 K. (right) The difference between the operational OSTIA and 0.0-K AATSR bias setting. The area-weighted global mean is −0.20 ± 0.13 K.

SSES bias estimates. Four reanalyses were performed using constant values of ΔT_{bias} in each, 0, 0.05, 0.15, and 0.2K, for all AATSR data regardless of the proximity confidence value (i.e., with a proximity confidence greater than or equal to 2) and applied to all AATSR data used by the OSTIA system in each reanalysis experiment. These values were chosen to span the likely range of ΔT_{bias}. Each reanalysis run was identical to the operational system, except for the following notable differences:

- To reduce processing time, the sea ice analysis was not performed, leaving only a relaxation to climatological SSTs at high latitudes. The operational model includes modifications to the SST under ice, taken from the European Organisation for the Exploitation of Meteorological Satellites' (EUMETSAT) Ocean and Sea Ice Satellite Application Facility (OSI-SAF) sea ice analysis, which was not included in the reruns.
- More data were available to the reanalysis because some data had not arrived in time for the operational system (highlighting the need for a sustained reanalysis program).
- The AATSR biases were altered as described above.

We note that the OSTIA bias-correction scheme makes use of both in situ data as well as AATSR where it is available and, at the time this work was conducted, no correction for skin temperature deviation was made to the AATSR data in both the operational system and reanalysis experiments.

To provide an independent analysis in which to compare our results, we also use a reference matchup database of near-contemporaneous AATSR and in situ buoy observations (O'Carroll et al. 2008). This allows us to further investigate AATSR bias values and the validity of the associated SSES, and to consider differences between the D2 and D3 SST retrieval algorithms made by the AATSR. This cannot be done using the OSTIA system because the GHRSST-PP format AATSR data do not include any information to determine the retrieval algorithm used by the AATSR for a given SST measurement.

b. *Results*

Figure 3 shows the differences between OSTIA SST_{fnd} reanalysis runs using a constant AATSR ΔT_{bias} setting of 0.0 K and those that use a constant ΔT_{bias} of 0.2 K, that is, the "extreme" AATSR bias settings in our experiments.

Because of the differences in sea ice treatment, the differences between the operational and reanalysis runs at high latitudes are disregarded. Differences were greatest in regions where there are less in situ data (e.g., the central Pacific and Southern Ocean). Differences are smallest in well-sampled areas with more shipping, such as the North Atlantic and North Pacific. In areas with only sparsely sampled in situ data, such as the South Pacific, the full impact of the change in bias can be seen; this is where the AATSR data make the maximum impact because they are the dominant source of bias correction. These plots clearly show the impact of AATSR on the OSTIA system analysis and underline the need to provide accurate SSES for the AATSR if it is to act as a reference satellite sensor.

To determine the most appropriate SSES bias correction to use with the AATSR L2P dataset, we compared each of the four OSTIA SST_{fnd} outputs, which have different bias estimates with assimilated in situ observations. We note that the in situ data do not constitute an independent dataset in this context, but they provide a measure of bias-correction impact that allows us to determine the most appropriate bias value for

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

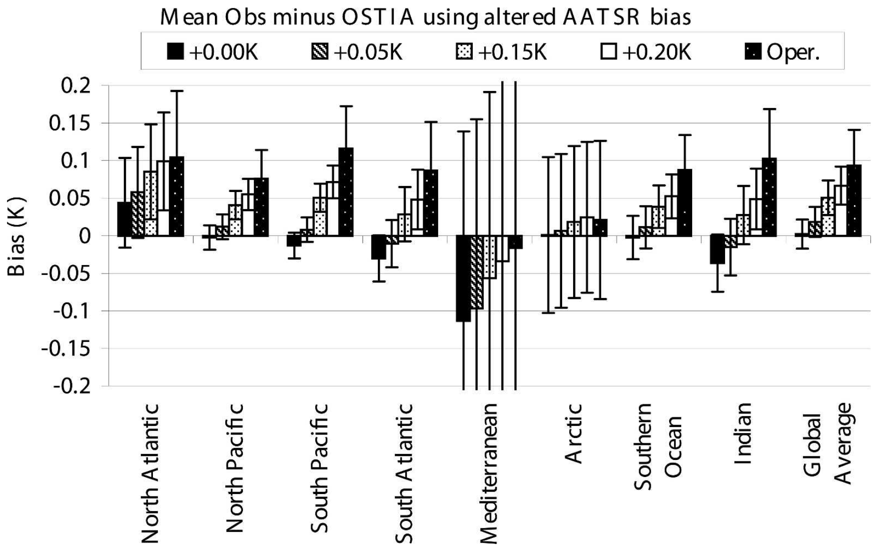

FIG. 4. Mean values of in situ SST minus OSTIA for several regions in June 2006, the runs using altered L2P AATSR bias estimates (0.0, 0.05, 0.15, and 0.2 K), and the operational OSTIA (Oper). The error bars indicate the uncertainty, using ±1 std dev in the daily means.

AATSR SSES. Figure 4 shows the resulting biases that were integrated over the regions and are defined in Table 3. In all four reanalysis cases, the modified bias values improved the accuracy of the OSTIA SST_{fnd} compared to the operational system in all the areas studied, except the Mediterranean Sea. In this area and the Arctic Ocean, the low number of observations and large analysis errors mean that the results are not robust. The best bias-correction impact results using the OSTIA system were obtained when ΔT_{bias} was set to zero. Figure 4 shows that the global bias in this case is 0.002 ± 0.02 K. The global bias was computed using all available observations during the validation period, including those outside the subregions as defined in Table 3. The observations were all equally weighted in the calculation.

As an additional check, the monthly mean difference between the run with the ΔT_{bias} set to zero and the corresponding HadISST data was computed, which is shown in Fig. 5. The mean bias was 0.04 with an RMS of 0.56 K, which is a substantial improvement, assuming that HadISST is correct, on the operational OSTIA outputs.

Further work to establish bias errors for AATSR SSTs has been performed using an AATSR–buoy matchup database, which is continuously being compiled at the Met Office. The AATSR data used within this database are the ATS_MET_2P product (also known as the "meteo product"), which contains skin SSTs, and brightness temperatures at 10 arc min spatial resolution, generated in near–real time. This is significantly different in character to the 1-km AATSR data used by the OSTIA system, but they should provide a better estimate of SST because of better cloud clearing and aggregation of data. Nighttime-only AATSR and buoy-observed SST are matched into pairs by choosing buoy observations that are located within the 10 arc min resolution grid box of the AATSR observation (or cell). The time difference between the two data types must be within ±3 h. In the event that two buoy SSTs are matched up to the same AATSR observation, the buoy observation closest in time to the AATSR observation is chosen. Here, D2, D3, N2, and N3 skin SSTs are retrieved separately from the AATSR brightness temperatures within the meteo product at the Met Office. In addition to the raw observations, an estimation

TABLE 3. Regions used for validation of the OSTIA reanalysis bias expt. The mean daily number of in situ observations is shown in the table. Note the relatively small number of available observations in the Mediterranean Sea and Arctic Ocean.

Region	Mean daily number of observations
North Atlantic	8363
North Pacific	8208
South Pacific	5781
South Atlantic	3084
Mediterranean Sea	504
Arctic	223
Southern Ocean	5921
Indian	1542
Global average	33 755

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

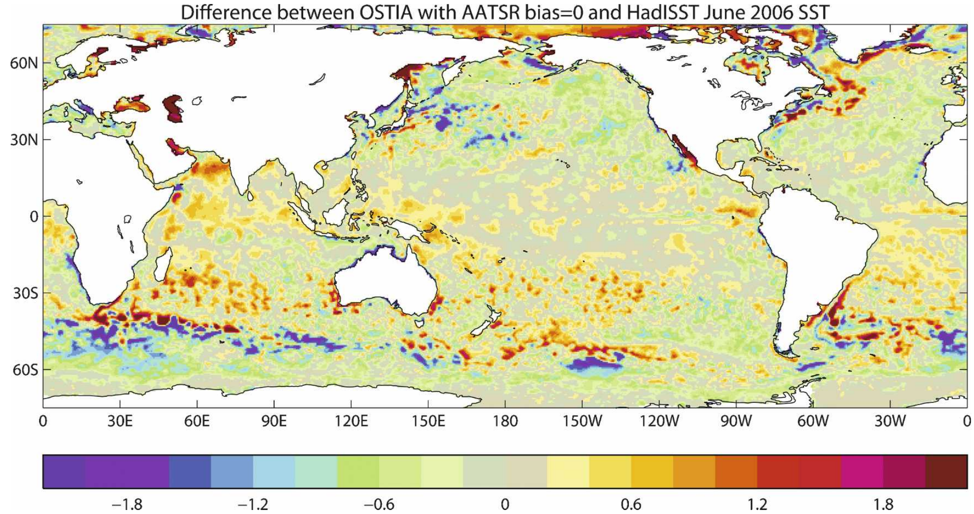

FIG. 5. Difference between mean June 2006 SST for OSTIA and HadISST (OSTIA − HadISST) for the reanalysis experiment in which the AATSR bias was set to 0.0 K. The mean difference for June 2006 is −0.04 K, and the RMS is 0.56.

of the cool-skin temperature deviation using the Saunders model (Saunders 1967) is applied to AATSR SST_{skin} observations, enabling a foundation SST to be derived, which can be usefully compared to collocated buoy SSTs. The AATSR MDB currently contains data from August 2002 to August 2006, and more details and validation results on the meteo product are found in O'Carroll et al. (2008). We are able to derive biases on the AATSR SSTs, separately for D2 and D3 window retrievals, and over a long time period, by analyzing this AATSR–buoy matchup database. This is not possible using the current version of the Medspiration L2P product, which does not distinguish between the D2 and D3 algorithm types. The GHRSST-PP data processing specification, version 1.7, does allow for data providers to include additional sensor-specific datasets within an L2P product, and we suggest that an indication of the type of algorithm be included in AATSR L2P data products.

We use two methods to derive AATSR biases: in our first approach, buoy SSTs are converted to a buoy "skin" SST using the Saunders model, and the resultant buoy skin SSTs are compared to nighttime AATSR D2 and D3 skin SSTs to derive the biases. As an alternative second approach, we can replicate the bias method used within OSTIA by converting the buoy SST to a buoy-derived skin SST by subtracting 0.17 K from the buoy SST while excluding those observations at low wind speeds (Donlon et al. 2002). Using the first method, we obtain biases on the D2 and D3 retrievals of 0.15 and 0.16 K, respectively, over the period from August 2002 to July 2006. Using the second method, we obtain biases on the D2 and D3 retrievals of 0.13 and 0.18 K, respectively, for the same time period. For June 2006, using the second method, the respective biases are 0.18 and 0.19 K. Given the differences between AATSR datasets (1 km versus 10 arc min) and uncertainties associated with the Saunders (1967) parameterization for warm-layer and cool-skin effects, we conclude that the MDB and OSTIA AATSR bias adjustment experiments are in agreement.

4. Discussion

We note that the best validation results (Fig. 4) for the OSTIA SST_{fnd} reanalysis runs were obtained when $\Delta T_{bias} = 0$. In this case, the global mean observations minus the OSTIA bias was close to zero. For AATSR to provide a true estimate of the SST_{skin} temperature, this result implies that $\Delta T_{skin} = SSES_{bias}$, and the ATS_NR_2P SST_{skin} data have a warm bias of 0.17 K, which almost exactly compensates for the mean SST cool-skin deviation of 0.17 K. This means that the current generation of AATSR products provide a good representation of the subsurface temperature. This bias can only be determined to an accuracy of approximately 0.03 K because of the observation minus background error estimates and the 0.05-K granularity used to define the bias correction in the reanalysis runs. Our matchup database results, in which in situ observations or the AATSR data are adjusted to compare either SST_{skin} to SST_{skin}, or SST_{1m} to SST_{1m} (where SST_{1m} is the SST at a depth of 1 m), indicate that the AATSR SST_{skin} D2 and D3 retrievals have a warm bias of 0.14 and 0.17 K, respectively (based on the mean

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

from our two analysis methods), over the period from August 2002 to July 2006. This independently confirms our reanalysis results.

For a user to apply GHRSST-PP AATSR observations correctly, an SSES bias adjustment is required to correct for the warm bias in AATSR data. We conclude that the SSES bias value of 0.33 K used for the lower proximity confidence data of Medspiration AATSR L2P data has little justification, and it should be removed and replaced with a single constant (lower) value of 0.17 K for all AATSR observations with a proximity confidence flag greater than 1. When a user applies these revised SSES bias estimates, the current AATSR dual-view SST ATS_NR_2P products will then provide a correct estimate of the SST_{skin} in agreement with the design specification of the AATSR mission.

5. Other issues for a new AATSR L2P GHRSST-PP format product

The current GHRSST-PP SSES scheme applied to AATSR L2P data is based on a series of simple threshold tests. These tests comprise a comparison to climatology and a comparison to the nearest cloudy pixel that assumes any cloud-clearing system applied to the satellite data is less robust in the vicinity of flagged cloudy pixels. These tests were derived from the experience of using AVHRR and MSG SEVIRI data and not AATSR. The heritage of the first test (to climatology) has some merits, but the application of the second test (location to nearest cloud) is not applicable to AATSR for the following two reasons:

1) The AATSR dual-view retrieval algorithm by default has two views for determining the likelihood of the pixel being cloudy. Indeed, during the day, the use of data from the 1.6-μm channel ensures that little (if any) cloudy pixels are flagged as clear sky.
2) The AATSR cloud screening is known to have a high false alarm rate (Merchant et al. 2005), and it, indeed, is more likely to mark clear-sky pixels as being cloudy rather than marking cloudy pixels as being clear sky. Although the results presented in Merchant et al. (2005) apply to ATSR-2, there is little functional difference between the cloud clearing in AATSR and ATSR-2, resulting in similar issues (Merchant et al. 2008).

Therefore, it is essential that any SSES scheme for AATSR builds on the strengths of AATSR.

Merchant et al. (2008) have shown that the AATSR daytime cloud screening is exceptionally good and that very few cloudy pixels are flagged as clear sky, owing to use of the 1.6-μm channel. However, this is not true at night when view difference effects along the edge of clouds mean that undetected cloud in the forward view results in lower-than-expected brightness temperatures and, consequently, warmer SSTs. However, the AATSR nighttime SST retrievals are dominated by the 3.7-μm channel; the forward view 3.7-μm coefficient is negative, so a lower brightness temperature will result in a warmer SST. Importantly, Good et al. (2006, manuscript submitted to *Remote Sens. Environ.*) showed that the dual view of AATSR can be used to identify these incorrectly flagged pixels by applying a threshold test to the difference between the dual-view and nadir-only SST values for each pixel. These findings confirmed the conclusions of Noyes et al. (2006) who, after applying a dual-nadir difference threshold test to split their validation results into two groups, noted a significant difference in bias for each group.

To apply an improved SSES scheme to AATSR that uses knowledge of AATSR, two important changes must be made to the current AATSR L2P data product:

1) The nadir-only SST for each dual-view SST in the L2P file must be included as either the SST or as a view difference. There are pros and cons to including this information in either form that need careful consideration, but the information must be added in one form or the other.
2) It is not currently possible to know if the nighttime L2P SST data used the 3.7-μm channel in the retrieval. This information must be included as a flag in the L2P data file.

With the above information, a much-improved SSES scheme can be developed for AATSR that will provide more meaningful error estimates for users.

We recommend that an extensive database of near-contemporaneous matchups between AATSR 1-km data and in situ observations be compiled and used to develop independent SSES for AATSR. The database should include auxiliary information (e.g., surface wind speed, solar radiation, proximity to dynamic SST regions, etc.) either from direct observation or from operational model outputs and all associated AATSR flags, if possible. However, at a minimum, it should include all SST retrievals, adequate information describing which channels were used in the AATSR SST derivation, nadir-only SST, and dual-nadir SST differences. This database should be maintained as an inte-

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

gral component of the AATSR processing system and analyzed regularly to provide the user community with the most robust SSES for use when applying AATSR SST data.

6. Conclusions

The Met Office Operational Sea Surface Temperature and Sea Ice analysis (OSTIA) SST_{fnd} outputs are biased cool relative to in situ observations by ~0.1 K. We attribute the bias error to the method we use to adjust satellite data inputs for bias errors that rely on a combination of in situ and *Envisat* AATSR ATS_NR_2P 1-km SST_{skin} data. In this process, we apply Single Sensor Error Statistics (SSES) bias estimates to all AATSR data using SSES provided in each GHRSST-PP L2P AATSR data file, although the AATSR SSES are known to be poorly specified.

We have conducted a series of reanalysis runs using the OSTIA system to determine the most appropriate SSES bias estimate for the AATSR. For each reanalysis run, we specified a fixed bias estimate of 0.0, 0.05, 0.15, and 0.2 K, and applied it to all L2P data having a proximity confidence value above 1. Using in situ observations as a reference data source, we find that the best estimate of SST_{fnd}, using the OSTIA system, is generated when the AATSR SSES bias is set to 0.0 K. Using a matchup database of near-contemporaneous 10 arc min AATSR and in situ data, we find that the AATSR SST_{skin} dual- and triple-window retrievals have a warm bias of 0.14 and 0.17 K, respectively, over the period from August 2002 to July 2006. The results of our experiments confirm that the current Medspiration SSES bias correction, provided with GHRSST-PP AATSR L2P observations, is poorly specified. Our database was not configured to test the relationship between the proximity confidence value and AATSR bias error. An independent comparison between AATSR data and the operational OSTIA product (A. G. O'Carroll 2006, personal communication) has also confirmed that without amending the AATSR SSES, OSTIA is representing a skin, rather than foundation, SST.

From our analysis, the sensor-specific biases used by Medspiration for AATSR, upgraded on 6 June 2006, are in need of revision. We conclude that the SSES bias value of 0.33 K used for the lower proximity confidence data of Medspiration AATSR L2P data has little justification, and it should be removed. Based on our matchup database and reanalysis results, we suggest that Medspiration be modified to use an SSES bias estimate of 0.17 ± 0.03 K for all category 2–6 proximity confidence values for the current AATSR dual-view SST ATS_NR_2P products to provide a correct SST_{skin} estimate. This result highlights a need for the ESA Medspiration/GHRSST-PP to conduct a study to properly define the AATSR SSES. We recommend that an extensive database of near-contemporaneous matchups between AATSR 1-km data and in situ observations be compiled and used to develop independent SSES for AATSR. This database should be maintained as an integral component of the AATSR processing system and analyzed regularly to provide the user community with the most robust SSES for use when applying AATSR SST data. Furthermore, our analysis and results indicate that such a study should be initiated as a matter of urgency to minimize the impact of inaccurate AATSR SST observations on user community applications.

Acknowledgments. The authors thank the U.K. Ministry of Defence for supporting OSTIA through public weather service funding. We also thank two anonymous reviewers for their helpful suggestions, which resulted in improvements in the manuscript.

REFERENCES

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.

——, and Coauthors, 2006: The Global Ocean Data Assimilation Experiment High Resolution Sea Surface Temperature Pilot Project (GHRSST-PP) Data Processing Specification version 1.7 (GDSv1.7). GODAE, 270 pp.

International Global Ocean Data Assimilation Experiment Steering Team, 2001: The Global Ocean Data Assimilation Experiment: Strategic Plan. GODAE Rep. 6, 23 pp.

Lorenc, A. C., R. S. Bell, and B. Macpherson, 1991: The Meteorological Office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **117,** 59–89.

Martin, M., A. Hines, and M. Bell, 2007: Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. *Quart. J. Roy. Meteor. Soc.,* **133,** 981–995.

Merchant, C. J., A. R. Harris, E. Maturi, and S. Maccallum, 2005: Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quart. J. Roy. Meteor. Soc.,* **131,** 2735–2755.

——, and Coauthors, 2008: Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Adv. Space Res.,* **41,** 1–11.

Noyes, E. J., P. J. Minnett, J. J. Remedios, G. K. Corlett, S. A. Good, and D. T. Llewellyn-Jones, 2006: The accuracy of AATSR sea surface temperatures in the Caribbean. *Remote Sens. Environ.,* **101,** 38–51.

O'Carroll, A. G., J. R. Eyre, and R. W. Saunders, 2008: Three-way error analysis between AATSR, AMSR-E, and in situ

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC

sea surface temperature observations. *J. Atmos. Oceanic Technol.,* **25,** 1197–1207.

——, 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., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. 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.

Saunders, P. M., 1967: The temperature at the ocean-air interface. *J. Atmos. Sci.,* **24,** 269–273.

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch, 2007: OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. *Proc., Oceans 2007, Marine Challenges: Coastline to Deep Sea,* Aberdeen, Scotland, IEEE/OES, 1–4.

Unauthenticated | Downloaded 04/18/21 09:10 PM UTC