Remote Sensing of Environment 116 (2012) 62–78

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects

Owen Embury [a], Christopher J. Merchant [a,*], Gary K. Corlett [b]

[a] School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom
[b] Department of Physics and Astronomy, University of Leicester, United Kingdom

ARTICLE INFO

Article history:
Received 19 February 2010
Received in revised form 9 February 2011
Accepted 22 February 2011
Available online 20 August 2011

Keywords:
ATSR
Sea surface temperature
SST
Validation

ABSTRACT

An initial validation of the Along Track Scanning Radiometer (ATSR) Reprocessing for Climate (ARC) retrievals of sea surface temperature (SST) is presented. ATSR-2 and Advanced ATSR (AATSR) SST estimates are compared to drifting buoy and moored buoy observations over the period 1995 to 2008. The primary ATSR estimates are of skin SST, whereas buoys measure SST below the surface. Adjustment is therefore made for the skin effect, for diurnal stratification and for differences in buoy–satellite observation time. With such adjustments, satellite-in situ differences are consistent between day and night within ~0.01 K. Satellite-in situ differences are correlated with differences in observation time, because of the diurnal warming and cooling of the ocean. The data are used to verify the average behaviour of physical and empirical models of the warming/cooling rates.

Systematic differences between adjusted AATSR and in-situ SSTs against latitude, total column water vapour (TCWV), and wind speed are less than 0.1 K, for all except the most extreme cases (TCWV <5 kg m^{-2}, TCWV >60 kg m^{-2}). For all types of retrieval except the nadir-only two-channel (N2), regional biases are less than 0.1 K for 80% of the ocean. Global comparison against drifting buoys shows night time dual-view two-channel (D2) SSTs are warm by 0.06 ± 0.23 K and dual-view three-channel (D3) SSTs are warm by 0.06 ± 0.21 K (daytime D2: 0.07 ± 0.23 K). Nadir-only results are N2: 0.03 ± 0.33 K and N3: 0.03 ± 0.19 K showing the improved inter-algorithm consistency to ~0.02 K. This represents a marked improvement from the existing operational retrieval algorithms for which inter-algorithm inconsistency is >0.5 K. Comparison against tropical moored buoys, which are more accurate than drifting buoys, gives lower error estimates (N3: 0.02 ± 0.13 K, D2: 0.03 ± 0.18 K). Comparable results are obtained for ATSR-2, except that the ATSR-2 SSTs are around 0.1 K warm compared to AATSR.

1. Introduction

A new sea surface temperature (SST) retrieval has been developed for the Along Track Scanning Radiometer (ATSR) instruments, with a view to developing an independent SST time-series for climate applications (Merchant et al., 2008a). Accurate retrieval of SST was the primary motivation for the ATSR missions (for a review and references, see Llewellyn-Jones and Remedios (this issue)). The instruments are well calibrated, and use along-track scanning to give both a nadir view of the ocean surface and a forward view at a satellite zenith angle of around 55°. The dual view capability is particularly important in ensuring continuity of accuracy in anomalous conditions, such as when stratospheric volcanic aerosol is present. ATSR SST retrieval coefficients have developed significantly (Merchant & Harris, 1999) since the work of Závody et al. (1995). The results presented here arise from the project ATSR Reprocessing for Climate (ARC) (Merchant et al., 2008a). Companion papers in this issue have detailed the ARC's radiative transfer simulations (Embury et al., this issue) and improved design of retrieval coefficients (Embury & Merchant, this issue), which includes a refined treatment of the ATSR viewing geometry and the use of coefficients banded by total column water vapour to reduce biases with respect to atmospheric state (Barton, 1998; Merchant et al., 2006). As a point of comparison, the ARC results are at points compared with results from techniques used in creating the ATSR SSTs in the ATSR multi-mission archive version 2.0.

This paper presents an initial comparison of the new satellite-derived SST estimates obtained for ATSR-2 and Advanced ATSR (AATSR) with independent measurements.

There are two novel aspects to the work with wider ramifications, as follows.

Firstly, in Section 2, we discuss how the purpose for which a validation is undertaken affects the quality control to be applied and

* Corresponding author. Tel.: +44 131 650 1000; fax: +44 131 650 5780.
E-mail addresses: owen.embury@ed.ac.uk (O. Embury), c.merchant@ed.ac.uk (C.J. Merchant), gkc1@leicester.ac.uk (G.K. Corlett).
URL: http://www.geos.ed.ac.uk/~chris/ (C.J. Merchant).

0034-4257 © 2011 Elsevier Inc. Open access under CC BY license.
doi:10.1016/j.rse.2011.02.028

the statistics to be reported. Here, we apply minimal quality control of in situ observations and report robust statistics, for reasons given.

Secondly, as described in Section 3, the satellite-derived estimates validated here are of two types. The primary observations are of skin SSTs (estimates from the satellite observations of the infrared radiometric temperature of the sea surface at the time of observation). In a further step, we derive time-adjusted depth SSTs (estimates adjusted for the ocean thermal skin effect, near-surface stratification and differences in observation time). The latter are intended to be more directly comparable to the in situ data used for validation. It is shown how adjustment for differences in time of observation can be modelled from a sufficiently large data set itself, and that this reduces systematic effects in the residuals between satellite and in situ SSTs (Section 4).

Nevertheless, the main purpose of the paper is the initial validation of new SSTs obtained within the ATSR Reprocessing for Climate project (ARC) (Merchant et al., 2008a). The validation is "initial" in the sense that all results presented here are derived before homogenisation (blending) of the multi-satellite SST time series. Ideally, the SSTs obtained from ATSR-2 and AATSR would be so consistent that no homogenisation would be necessary. Practically, the aim of the careful work done on retrieval development is to minimise any homogenisation required, since this increases the confidence we can have in the stability of the homogenised record that will ultimately be created. We find here that the global and regional retrieval biases in the new satellite SSTs are low, mostly <0.1 K. The discrepancy between ATSR-2 and AATSR is also found to be <0.1 K. Although small, this discrepancy is not negligible. Homogenisation will therefore be a crucial step in developing a record appropriate for climate, and will be reported in a subsequent article.

2. Approach to validation

2.1. Statistical approach

It is important to be clear about the purpose of the SST validation reported here. Our objective is to assess the accuracy of the new SST retrieval scheme defined in Embury and Merchant (this issue). This is a narrower focus than validation of a new SST product, since the quality of an SST product is determined by the whole satellite-processing system, which includes determination of which pixels are valid for SST retrieval (by cloud and sea ice detection, checks on validity of radiances, etc.). Cloud detection failures in particular are known to introduce occasional large errors in retrieved SST that significantly increase the standard deviation of satellite SSTs in comparisons with in situ SSTs. In this paper, we make the assumption that the incidence of cloud detection failure is low (this is shown to be reasonable in Section 2).

Outliers can also arise from in situ errors. Operational systems for flagging erroneous in situ data typically offer a basic level of quality control (QC), and it is usual for additional QC to be applied (e.g., O'Carroll et al., 2008). However, any QC to remove doubtful data inevitably involves a significant element of judgement: the harder one looks, the more cases one can find justifications for rejecting. There is therefore an arbitrary element that makes reported statistics difficult to compare between validation studies: to some degree, the validation statistics validate the judgements made regarding QC and not just the SST retrieval method. In this work, the in situ observations are subject to only minimal additional QC (e.g., in situ location is confirmed to be ocean, temperature is not below freezing).

As our purpose is to assess the new SST retrieval scheme, we are not concerned with characterizing the distribution of outliers that arise for the reasons above. Therefore we use outlier tolerant statistics, also known as robust statistics (Huber, 1981), to assess the retrieval results. No further attempt is made to catch and eliminate failures of cloud detection or outliers in the in situ data during the statistical analysis.

To describe the central value of a distribution we use the median. For a Gaussian distribution or any symmetric distribution, this would equal the mean. To describe the spread of a distribution, we use a "robust standard deviation" or RSD. The RSD used is 1.48 times the median absolute deviation from the median (MAD). For a Gaussian distribution, the RSD and conventional standard deviation are equal. There are many other ways to reach comparable statistics, such as finding the best fit Gaussian, scaling the inter-quartile range, or applying a 3-sigma filter and recalculating conventional statistics on a second pass. However, the median and RSD are easy to calculate and have a simple analogy to their conventional counterparts. Both measures are highly robust meaning that a significant minority of data points can be replaced with outliers of infinity without the calculated statistics also going to infinity. In the particular case of SST validation, these robust statistics are not overly sensitive to cloud detection or gross in situ errors, and therefore isolate the performance of the SST retrieval itself.

It could also be argued that the retrieval itself is a source of outliers arising from extreme atmospheric conditions (e.g., (Minnett, 1986)). This is true, but we cannot identify in a validation data set which outliers do arise in this way, rather than from cloud detection or in situ gross errors. The best information we have is from radiative transfer simulations for a comprehensive range of atmospheric states (Embury et al., this issue). Fig. 1 shows the distribution of simulated retrieval error for dual-view 11 and 12 μm retrieval, and shows that while the distribution is slightly skewed there are relatively few outliers and 96.5% of data points lie within 3-RSDs of the median. The outliers are mostly associated with high amounts of atmospheric water vapour. Table 1 shows various statistical measures of the simulated retrieval errors. The statistics are given for all four types of retrieval as discussed in Embury and Merchant, (this issue): N2 and N3, these being single-view retrievals using 11 and 12 μm and 3.7, 11 and 12 μm channel sets respectively; and D2 and D3, the corresponding dual-view combinations. In all cases the mean and median agree to better than 0.01 K, and only the nadir-only two-channel retrieval (N2) has markedly different SD and RSD.

2.2. Match-up data

2.2.1. In situ data

For the purpose of this initial validation, we created a match-up database (MDB) that contains both in situ SST measurements and the

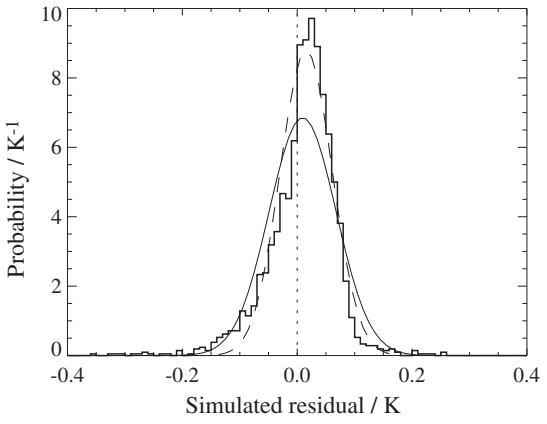

Fig. 1. Probability distribution of simulated D2 retrieval error. Thin solid line shows Gaussian distribution using standard statistics, the standard deviation (SD) being 0.058 K; thin dashed line shows Gaussian distribution using robust statistics, the robust SD being 0.045 K. There are 2100 simulated cases, as described in Embury and Merchant (this issue), and those shown are for the AATSR instrument. Results for other ATSRs are similar.

Table 1
Summary statistics of simulated retrieval error in the absence of instrumental noise for retrievals using four channel combinations, N2, N3, D2 and D3 (see main text for definitions). "Min" and "Max" are the minimum and maximum simulated retrieval error; "LQR" and "UQR" are the lower and upper quartiles; "RSD" is the robust standard deviation (as defined in main text); "SD" is the standard deviation. All quantities are in Kelvin.

	Min	LQR	Median	RSD	Mean	SD	UQR	Max
N2	−1.127	−0.099	0.015	0.157	0.010	0.284	0.116	1.462
N3	−0.239	−0.049	−0.005	0.067	0.003	0.080	0.044	0.584
D2	−0.359	−0.018	0.016	0.045	0.009	0.058	0.045	0.256
D3	−0.146	−0.024	0.002	0.039	0.003	0.039	0.028	0.187

collocated ATSR observations. The in situ SST observations are provided by the Met Office Hadley Centre in the form of ASCII files using one decimal place accuracy, i.e. in situ latitude and longitude are reported to 0.1° precision and SST measurements to 0.1 K, and are comprised of data from the International Comprehensive Ocean–atmosphere Dataset (ICOADS, (Worley et al., 2005)) from the start of the ATSR-1 mission to the end of 1997, and National Centers for Environmental Prediction (NCEP) near-real-time (NRT) marine reports (see http://icoads.noaa.gov/nrt.html) from the start of 1998 to present. The majority of data are received through the Global Telecommunication System (GTS) which includes data from three main sources: drifting buoys, moored buoys, and ships. Additional QC information using the method of Rayner et al. (2006) is provided with the in situ data, from which we exclude obviously bad data before matching with the satellite data. Checks included in the QC flags are: invalid or missing data; blacklisted data sources; inland and lake buoys; temperatures below −1.8 °C; temperatures >8 °C different from climatology; and a buddy check (in which nearby buoys are compared).

Each of the three in situ data types has particular characteristics. Drifting buoys have the most complete spatial coverage, although it is far from uniform, and they measure the sea temperature at around 20 cm depth. The uncertainty in drifting buoy observations was estimated to be 0.2 K by O'Carroll et al. (2008), and part of the uncertainty arises from reporting of drifting buoy temperatures with only 0.1 K resolution (i.e., rounded to one decimal place). During the AATSR mission drifting buoys formed the majority of the in situ dataset. Moored buoys can be roughly separated into two distinct types, the Global Tropical Moored Buoy Array (GTMBA) and coastal moorings. The GTMBA includes data from the TAO/TRITON (Tropical Atmosphere Ocean/Triangle Trans Ocean Buoy Network; (McPhaden, 1995)), PIRATA (Prediction and Research Moored Array in the Atlantic; (Bourlès et al., 2008)), and RAMA (Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction (McPhaden et al., 2009)) arrays, and all use ATLAS (Autonomous Temperature Line Acquisition System) moorings. These provide the most accurate measurements in the ATSR MDB: the instrument accuracy is better than the 0.1 K resolution reported in GTS records, but the number of match ups is low, representing <3% of the MDB. ATLAS moorings measure SSTs at 1 m depth. Outside of the GTMBA, most other moored buoys are located in North American and European coastal waters. While coastal moored buoys do record more information than drifters, their designs are more varied hence temperature measurements are from a wider range of depths. Also, the coastal location is particularly challenging for SST retrieval from space, and so the coastal moored buoys are excluded from this initial ARC SST validation. Finally, there are measurements from Voluntary Observing Ships (VOS). These represented the majority of the in situ record at the start of the ATSR1 mission (1991); however, the share was rapidly reduced both due to the reduction in total number of VOS reports and the increase in buoy data. VOS measurements use a range of different sensors—hull contact, engine intake etc.—resulting in significantly poorer precision than buoy data (e.g. (Emery et al., 2001)), and therefore are also excluded from the initial ARC SST validation.

2.2.2. Satellite data

Matching the in situ data with the satellite imagery is done as follows. Firstly the in situ data are filtered to select only reports within 3 h of a given ATSR data-file (corresponding to a single orbit); these are then collocated to the nearest ATSR pixel (whether clear or cloudy) and all matches with separation >1 km (i.e. outside the ATSR swath) are rejected. Secondly, matches to duplicate satellite data are detected and removed. These are cases where a single buoy has reported multiple observations, within the ±3 hour window, that are matched to the same ATSR observation. The match with the smallest time separation is retained and the remaining are ignored. Matches where a single buoy observation is matched against more than one ATSR orbit can occur at higher latitudes where the same location is observed on consecutive overpasses. Such matches are considered valid and are retained.

While matching against the pixel level product in this way should ensure that all satellite-in situ distances are less than 1 km, in practise it is limited by the geolocation accuracy of the two data types. As the in situ locations are reported to only 0.1° resolution, spatial separations of up to ~7 km may exist. ATSR SSTs are extracted for a 5×5 pixel window surrounding the nominal in situ position and the average SST is calculated using all the clear-sky pixels. This pixel window reduces the random SST uncertainty from sensor noise by a factor of up to $5=\sqrt{25}$.

Additionally, various numeric weather prediction (NWP) data used in the ARC processing system, such as surface wind speed and total column water vapour (TCWV), are extracted along with the satellite retrievals.

There are two consequences from the approach of matching to full-resolution data, rather than spatially averaged SST products as done in some previous AATSR validation studies (e.g., drifter validation in Corlett et al. (2006) and O'Carroll et al. (2006). The first is that this allows the satellite-in situ spatial and temporal separation to be more accurately estimated and minimised. The second relates to the fact that the operational ATSR spatially averaged products are generated by first averaging the brightness temperatures and retrieving the SST from those averages. The dual-view nature of the ATSR instruments means that different pixels can be cloudy in the two views. This can introduce unpredictable errors into the dual-view retrievals as the nadir and forward views are no longer guaranteed to correspond to the same source locations. We calculate SSTs from individual pixel brightness temperatures, ensuring the locations match as far as is possible between nadir and forward views.

The cloud screening algorithm used in the ARC processing system is based on the probabilistic Bayesian method described in Merchant et al. (2005), updated for day-time observations using methods in Mackie et al. (2010), and further developed for simultaneous exploitation of the two views of ATSR. A future article will give a full description and report on validation of the implementation for ARC of the Bayesian cloud detection algorithm; here we simply comment that, overall, it seems more effective than the threshold-based "SADIST" method (Závody et al., 2000) used with operational ATSR data. Table 2 compares the SD and RSD of satellite-in situ differences for the two cloud masks. The standard deviations show a substantial decrease when using the Bayesian cloud detection, equivalent to removing independent noise of magnitude >0.3 K. Even the outlier-tolerant RSDs show modest improvement (equivalent to removing ≃0. 12 K of independent noise), suggesting that the improvement in cloud detection is not limited to outliers. For a dual view SST, the global retrieval bias is not greatly affected by the cloud detection used, probably because cloud contamination can cause warm or cold biases depending on which view is affected by residual cloud. The ratio of the SD to RSD gives a measure of how much a

Table 2
Impact of different cloud detection on validation results, for AATSR between July 2002 and December 2007. Statistics are shown in Kelvin for residuals between satellite and in situ observations when using cloud-flags from the operational imagery ("SADIST") and from ARC (Bayesian) cloud screening algorithms. SD represents the standard deviation of satellite-in situ differences. RSD is an outlier-tolerant robust estimate of standard deviation. The satellite SSTs are dual-view two-channel skin SSTs (D2 SSTs) based on the same ARC algorithm throughout, to isolate the effects of cloud screening alone. D2 SSTs are not the most accurate of the available types of retrieval, but are available day and night, allowing the impact of the different day and night cloud detection schemes to be compared.

	Drifters			GTMBA		
	N	SD	RSD	N	SD	RSD
Day						
SADIST	89,831	0.565	0.236	5256	0.370	0.220
Bayesian	120,916	0.458	0.235	6837	0.324	0.218
Night						
SADIST	100,826	0.602	0.257	3571	0.400	0.220
Bayesian	96,857	0.452	0.227	3081	0.242	0.174

Table 3
Drifting buoy matches for ATSR-2 and AATSR. "Clear-sky" indicates matches with at least one clear pixel in the 5×5 pixel array around the reported buoy location.

	ATSR-2	AATSR
Start date	1995-06-01	2002-07-24
End date	2003-06-22	2007-12-24
Total drifters	397,474	922,464
Clear-sky drifters	83,007	218,826
Clear-sky drifters/year	9896	41,945
Total GTMB	24,557	31,398
Clear-sky GTMB	7772	9918
Clear-sky GTMB/year	892	1872

particular SD is affected by outliers (with a ratio of 1 implying no outliers). The ratios are closer to unity for the Bayesian cloud detection than for SADIST. Even for Bayesian detection, however, the ratio is well above 1, reflecting the fact that cloud detection is not the only contribution to outliers. The tropical moored buoys matched with Bayesian-screened SSTs have SD equal to 0.24 K, with the RSD equal to 0.17 K, reflecting both the improved cloud detection and higher quality of the in situ data compared to drifters.

2.2.3. Match-up database summary

To assess the new SST retrieval scheme used within ARC, we require that the comparison data are accurate, consistent, and cover the widest possible range of observing conditions. Drifting buoys provide the widest coverage of all the in situ types, with reasonable consistency and accuracy, and are used throughout this paper, except where otherwise stated. The GTMBA data are more accurate and consistent, but the low spatial coverage and number of matches make them unsuitable for fully characterising global SST. Nevertheless, because of their good accuracy they are also useful for comparison against the ATSR data. All comparisons from this point on are based on ARC cloud screening, irrespective of whether the SST retrieval uses operational or ARC coefficients. Because operational SSTs were derived only for image pixels flagged clear by the operational cloud flagging, SSTs using operational coefficients have been derived by re-applying the operational coefficients to imagery cloud-screened by ARC methods.

Table 3 summarises the total number of matches found. Approximately 75% of all matches found corresponded to completely cloudy scenes—i.e. all 25 pixels in the 5×5 window were classified as cloud by the Bayesian cloud detection algorithm—and are not used for retrieval validation. Fig. 2 shows the geographical distribution of the drifting buoy matches retained for validation for ATSR-2 and AATSR.

3. Satellite–buoy differences

3.1. Skin SST and skin effect

Radiometers operating at infrared wavelengths, such as the ATSR instruments, are sensitive to radiation emitted from the layer between the air–sea interface and about 20 μm below the air–sea interface, depending on wavelength. In situ measurements of SST, however, are usually made at some distance below the surface, the exact depth dependent on the observing platform used. During night there is a temperature difference of approximately 0.2 K between the two measurements. This is a "skin effect", reflecting the large thermal gradient required to maintain a flux of heat through the near-surface layer in which molecular rather than turbulent heat transfer processes dominate (e.g., (Saunders, 1967)). The skin effect exists also during the day, but may be modified by absorption within the ocean skin layer of near-infrared solar insolation. Under conditions of sufficiently low wind-driven mixing, solar heating will also cause thermal stratification of the upper ocean (discussed in the next section) further altering the difference between what is observed by satellite and measured by the in situ instrument. The Met Office has provided estimates of the skin effect for the ARC project using the Fairall et al., (1996) method forced by ECMWF-Interim data (Simmons et al., 2007). This accounts for both wind speed and net heat flux (full details of the implementation used are given in Horrocks et al.(2003a)). The

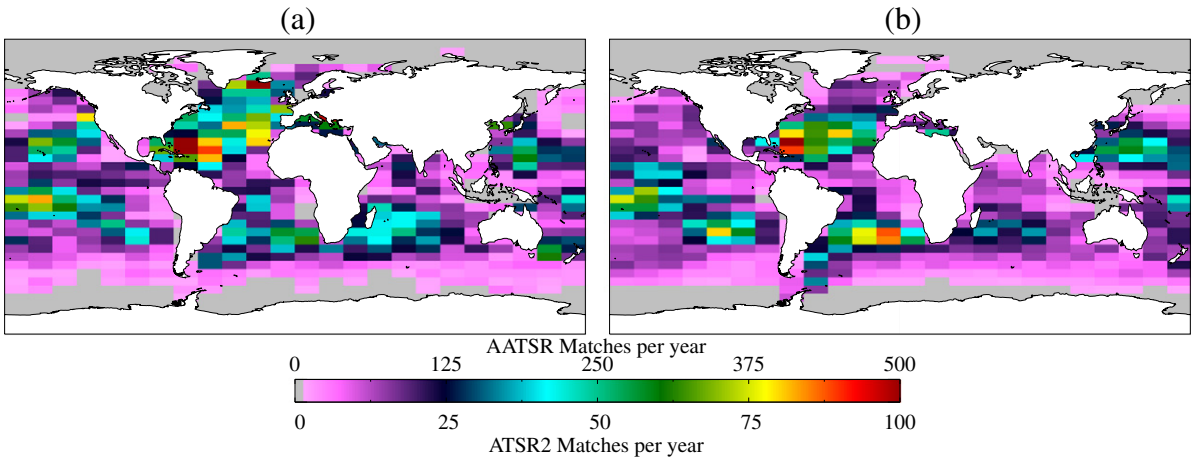

Fig. 2. Global plots of clear-sky matches per year with drifting buoys for (a) ATSR-2 and (b) AATSR, per cell of 15° in longitude by 5° in latitude. The rate of matches can be compared in terms of matches per unit area by scaling the plotted value at a given location by the arc–cosine of its latitude.

Fairall model includes two tuneable parameters (λ_0 and A), which have been set to $\lambda_0 = 4.5$ and $A = 0.2$ for this project, based on a previous study as discussed later in this section.

Fig. 3 shows the difference between the AATSR retrieved SST_{skin} and SST_{depth} as measured by drifting buoys along with the predicted skin effect from the Fairall model output. For comparison we also show the empirical parameterization of night-time skin effect in terms of wind speed (u) from Donlon et al. (2002), namely:

$$SST_{subskin} - SST_{skin} = 0.14 + 0.30\exp(-u/3.7) \quad (1)$$

In the nighttime case, other than a warm offset of ~0.05 K, the observed satellite-drifter differences follow the expected variation with wind speed predicted by the Donlon curve. The offset may represent an overall skin SST retrieval bias, a bias in the Donlon parameterisation, or a combination of these factors. The variations with wind speed of the Donlon estimate and the night time data are similar, in that there is little variation above 6 or 7 m s^{-1}.

Applying the Fairall method at night yields the shaded (\\\) band. The band shows the median plus and minus one RSD of the result of the Fairall calculation; there is a range for a given wind speed because the Fairall method accounts for the heat flux variability through the skin layer. The Fairall skin-effect matches the observations at higher wind speeds (>10 m s^{-1}), cools with respect to the observed difference for winds in the range ~10 down to ~5 m s^{-1}, and is colder than but parallel to the observed difference at lower wind speeds.

During day-time, the absorption of insolation by the ocean skin leads to satellite-drifter differences which are almost independent of wind speed at ~0.1 K, except for a rise at wind speeds <3 m s^{-1}. The observed difference is a combination of the day-time skin effect and, particularly at wind speeds <3 m s^{-1}, thermal stratification (Murray et al., 2000) between the surface and the drifting buoy observation depth. The shaded (//) band is the day-time skin effect calculated using the Fairall method, which shows much less variation with wind speed than the nighttime data.

For the next step in comparison, we formulate an $SST_{subskin}$ estimate by using the Fairall model to adjust the SST_{skin} retrieval. Note that we could apply the adjustment to the drifting buoy SSTs to form a skin SST estimate to compare with the satellite SST_{skin}. The residual differences would be identical. The convention we adopt here means that all estimation steps (SST retrieval and adjustments) are kept on one side of the comparison and are validated against unadjusted in situ measurements.

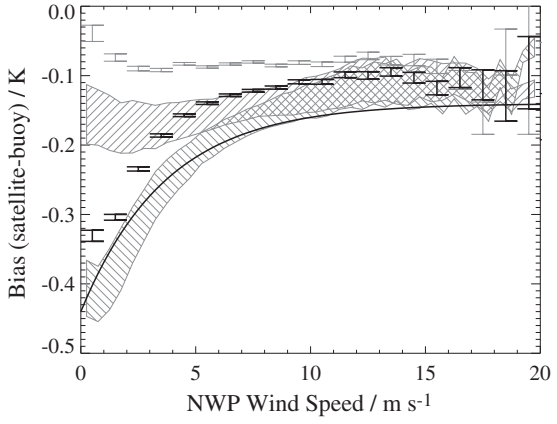

Fig. 3. AATSR retrieved SST_{skin} minus drifting buoy SST as a function of wind speed. Error bars show the range of median difference plus and minus one standard error for (black) night-time and (grey) day-time matches. Solid line shows empirical parameterization of skin effect from Donlon et al.(2002). The shaded areas show the estimate using the Fairall et al.(1996) method, the vertical range being the median ± RSD for each bin of wind speed: (\\\) night-time cases; (//) day time cases. Unlike the Donlon method, which is a function only of wind speed, Fairall's method attempts to account for heat flux through the skin layer and solar absorption within the skin.

Fig. 4 shows the difference between the estimated $SST_{subskin}$ (satellite skin SST adjusted using the Fairall method) and drifter SSTs as a function of windspeed. The $SST_{subskin}$-drifter difference is <0.03 K for higher wind speeds (>10 m s^{-1}), with a warmer bias of up to 0.1 K for lower wind speeds. The day and night $SST_{subskin}$–drifter differences are now very consistent—for all wind speeds above ~3 m s^{-1} the day and night differences agree to ~0.02 K. A plausible interpretation is that the divergence of the day and night differences for winds <3 m s^{-1} is the consequence of thermal stratification within the ~20 cm of the surface (between the drifter thermistor depth and the surface). There is good consistency of the day and night curves at higher wind speeds (3 to 15 m s^{-1}) despite quite distinct day and night behaviour over this range in Fig. 3. This suggests that the Fairall formulation is an effective tool for estimating the different wind-dependence of the skin effect in day and night. The residual variation of bias with wind speed above 3 m s^{-1} could arise from: (i) an inadequacy of the physics in Fairall's formulation that is in common between day and night; (ii) an inadequacy in the value of the λ_0 and A parameters used here; (iii) biases at low wind speeds in the NWP fluxes used to drive the model (e.g., Horrocks et al., 2003a); (iv) a wind-related error in the skin SST retrieval; or (v) an error in the skin SST retrieval that depends on a factor correlated with wind speed (a confounding factor). A wind-related retrieval error would arise, for example, via an inadequate simulation of the dependencies of emissivity and reflected radiance (most likely in the forward view) when defining retrieval coefficients. Potential confounding factors are: inability of the retrieval to cope wholly with marine aerosol (whose concentration is correlated to wind speed); or systematic retrieval error related to atmospheric humidity, since column water vapour is correlated to wind speed globally.

Perhaps all of these factors play some role, but it seems likely that parameter error in the implementation of the Fairall model is significant. The implementation of the Fairall model used here was tuned in an earlier study at the Met Office for use with operational AATSR data, so as to minimise the wind-dependence of bias between operational AATSR and in situ buoys. Much of the residual wind-speed dependence seen here could therefore arise as an artefact of the difference between the operational and new ARC retrieval coefficients. A preliminary study suggests that the residual wind-speed dependence is reduced (data not shown) by removing the tuning step in the Met Office's implementation of the Fairall model and reverting to Fairall's original recommended formulation. This will preserve the independence of ARC SSTs from drifting buoy observations (see Embury & Merchant, this issue), and results will be reported in a later paper.

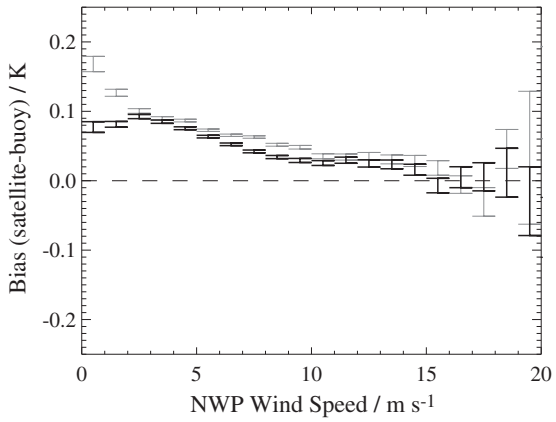

Fig. 4. $SST_{subskin}$ minus drifting buoy SST as a function of wind speed. The $SST_{subskin}$ estimate is the AATSR-retrieved skin SST adjusted for the skin effect using the Fairall model (see main text). Error bars span the range of median difference plus and minus one standard error for (black) night-time and (grey) day-time matches.

Table 4
Statistics of AATSR D2 sub-skin SST estimation for different skin effect corrections. Median of satellite-in situ, with RSD in parentheses, both in Kelvin.

	Day	Day, $u>3$ m s^{-1}	Night
$N_{matches}$	120,916	104,843	96,857
None	-0.084 (0.235)	-0.084 (0.223)	-0.149 (0.227)
Donlon	0.122 (0.240)	0.112 (0.226)	0.061 (0.221)
Fairall	0.070 (0.241)	0.066 (0.229)	0.057 (0.228)

Table 4 shows the global retrieval biases and RSD for no skin correction, with skin correction using the Donlon parameterization, and using the Fairall model. The Donlon formulation is effective at reducing both the night time bias and the RSD. While the Fairall correction does result in a slight positive bias (~0.05 K) it has brought the day and night retrievals into very close agreement. Use of the Fairall method also causes a small increase in RSD; equivalent to an independent noise signal of magnitude 0.04 K (night) or 0.06 K (day).

3.2. Thermal stratification

During the day solar heating of the upper ocean can cause the formation of a near-surface "warm layer" under low wind conditions (e.g., Lukas, 1991). As a result, the sub-skin temperature is warmer than temperatures at lower depths. In cases of sustained wind speed less than ~1 m s^{-1} and high incident solar radiation, the difference can exceed 5 K by early afternoon relative to the SST before dawn, as thermal stratification occurs. Such extreme events are seen in spring and summer in mid and high latitudes (Gentemann et al., 2008; Merchant et al., 2008b). More typically, the diurnal cycle of SST is of order a few tenths of Kelvin.

To model diurnal warming for the ARC project we use the Kantha and Clayson(1994) model implemented at the Met Office, of which full details are given in Horrocks et al. (2003b). This model is a one dimensional turbulence closure model, giving a profile of modelled stratification throughout the model run. More highly parameterized models for diurnal stratification exist (e.g., Gentemann et al., 2009; Zeng & Beljaars, 2005), and these merit evaluation in future work.

Temperature differences are calculated between the sub-skin and depth SST at 0.2, 1.0, and 1.5 m below surface. In 83% of cases these differences are <0.05 K. For conditions conducive to large diurnal warming events, the modelled stratification across the top ~20 cm can reach 0.5 K by 1000 h local time. Fig. 5 shows the difference between the SST at 0.2 m estimated by adjusting AATSR D2 SST_{skin} for both skin effect and diurnal stratification, minus drifting buoy SST. Compared to Fig. 4, there is reduced difference between the day and night-time

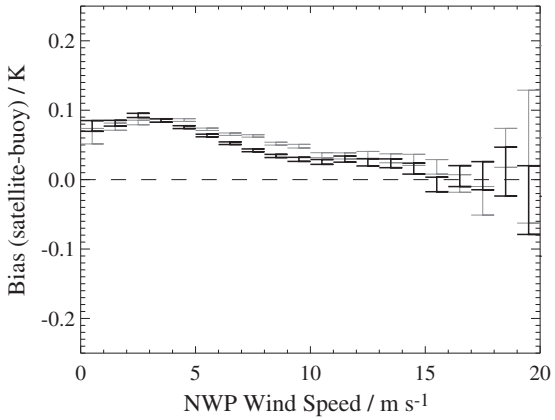

Fig. 5. AATSR estimated $SST_{0.2m}$–drifters as a function of wind speed. Error bars span the range of median difference plus and minus one standard error.

Table 5
Median and (in parentheses) RSD of satellite SST minus drifter SST in kelvin, where the satellite SST is the D2 SST skin retrieval adjusted to sub-skin and depth (20 cm). ΔT is the estimated stratification between the sub-skin and depth.

	All cases	$\Delta T<0.05$ K	$\Delta T<0.05$ K
$N_{matches}$	120,916	116,763	4153
Sub-skin	0.070 (0.241)	0.069 (0.236)	0.150 (0.422)
0.2 m	0.066 (0.241)	0.068 (0.236)	-0.004 (0.441)

dependence of difference at low wind speeds, indicating that the stratification estimate is beneficial.

Table 5 shows how correcting for the diurnal stratification affects the global biases. While the thermal stratification correction has a visible improvement in the wind speed dependent biases shown in Fig. 5, it has negligible effect on the global biases and RSD. This is primarily due to the low incidence (14% below 3 m s^{-1}) of extreme low wind speed conditions in the global dataset. Therefore, in Table 5, we also present statistics for cases of low thermal stratification (<0.05 K) and high thermal stratification (>0.05 K) separately. Day time cases of low modelled stratification have lower RSD than the high stratification cases. Using a harsher threshold of 0.01 K reduces the day time RSD to match the night-time RSD of ~0.23 K (these "extremely" low stratification cases cover 93% of all daytime matches). This is consistent, since stratification at night is usually expected to be negligible. The day-time cases with stratification greater than 0.05 K have warmer sub-skin SST estimates relative to drifters (by ~0.08 K) than the low-stratification cases, suggesting that truly stratified cases are being captured by this criterion. However, the average stratification correction for the high-stratification cases is 0.15 K, which indicates that on average the stratification is over-estimated.

The RSD of the day-time SST depth minus drifter SST is marginally greater than the RSD of day-time sub-skin SST minus drifter SST, for the more stratified sample. So, it seems that the sub-skin to SST depth adjustment is also adding some noise to these day-time cases. It is challenging to model the magnitude of the thermal stratification as it depends on both the wind speed and solar radiation history since dawn. It may be that the NWP data used to drive the model do not always capture these histories sufficiently accurately. Nonetheless, the stratification model is shown to have some skill at predicting the presence of stratification events and in reducing the bias of SST depth compared to drifting buoy SSTs.

Table 6 shows equivalent statistics for the GTMB array. Here the stratification effect is greater as the ATLAS moorings measure at a greater depth, and they are found in tropical locations where stratification events happen more often because of year-round high insolation and relatively low average wind speeds. The model is now reducing median and RSD satellite-in situ differences when considering both all matches and the low stratification events. For the more stratified cases, the stratification adjustment results in a small increase in RSD, but the significant improvement in median difference (from ~0.25 K to 0.03 K) shows that the stratification model is remarkably effective on average.

3.3. Diurnal cycle and time-adjustment

Sub-daily variability, particularly the diurnal cycle of SSTs, also leads to differences between the satellite SST and in situ measured

Table 6
As Table **5** but for ATLAS moorings in the GTMBA. Measurement depth is 1 m.

	All cases	$\Delta T<0.05$ K	$\Delta T<0.05$ K
$N_{matches}$	6837	5783	1054
Sub-skin	0.072 (0.224)	0.055 (0.203)	0.245 (0.365)
1.0 m	0.049 (0.220)	0.051 (0.201)	0.030 (0.393)

SSTs if the two are not perfectly coincident in time. The maximum time-window allowed in the MDB is 3 h, so up to 3 h of solar heating or night-time cooling can occur between the two observations of the SST. Minnett (1991) showed that spatial separation between satellite and in situ of ~10 km and time separations of ~2 h can introduce RMS errors of ~0.2 K into satellite validation. In this section, we show that, in fact, the errors associated with the time have a significant systematic component that can be compensated for. This is important where the distribution of satellite–buoy time difference is asymmetrical or where the satellite observation time is near the peak or trough of the diurnal SST cycle, because then the time difference effect can introduce an apparent bias in the validation results.

The effect of the time separation on retrieval bias is shown in Fig. 6. There is a cooling trend of 0.015 K/h affecting night-time matches and a warming trend of 0.058 K/h during the day. The Envisat satellite has a fixed 10:00 h Local Equator Crossing Time (LECT), and for most matches we can consider the local satellite observation times to be roughly 10 a.m. and 10 p.m. and therefore interpret Fig. 6 as showing the trends corresponding to two segments of the diurnal cycle. There is an upper limit to the daytime satellite–buoy retrieval bias for time differences >1.5 h, which corresponds to local buoy times of around 8:30 a.m. and earlier, i.e. local times closer to the minimum of the diurnal cycle, before the sea surface begins to stratify significantly due to solar radiation.

The Fairall skin and Kantha–Clayson stratification models, discussed in the previous sections, can also be used to estimate the surface heating/cooling rates (that is, rates of change of surface temperature). This was originally performed in order to homogenise the ARC SST record—the Envisat satellite which carries the AATSR instrument has a 10:00 h LECT while the ERS-1 and ERS-2 satellites carrying the ATSR1 and ATSR-2 instrument have 10:30 h LECTs. This half-hour difference is sufficient to introduce a discontinuity in the day-time SST record of around 0.03 K (half this at night). Therefore, in order to develop a homogeneous record across the three sensors, an adjustment for the diurnal cycle is required. Here we validate the modelled heating and cooling rates against observation-driven estimates.

Fig. 7 shows the mean observed (line) and modelled (band) heating rates at the AATSR observation time (mostly close to 10 a.m. or 10 p.m.) as a function of solar zenith angle (SZA). The observed heating rates are calculated by linear regression of the satellite-buoy retrieval bias against time differences for a series of strata of SZA. The simulated heating rates are represented by the shaded band which gives the median ± RSD of simulated heating rate as a function of SZA for the same set of matches. The night-time rate (SZA>90°) is between –0.01 and –0.02 K h^{-1} (i.e., modest cooling). The day-time

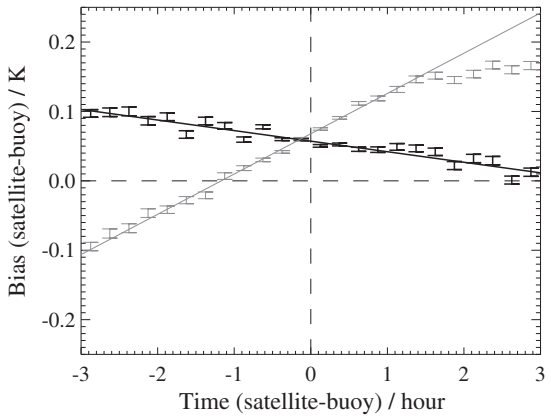

Fig. 6. D2 $SST_{0.2m}$ retrieval bias as a function of satellite–buoy time difference for daytime (grey) and night-time (black) matches. Solid lines show linear best fit to data (only using time differences <1.5 h for daytime matches).

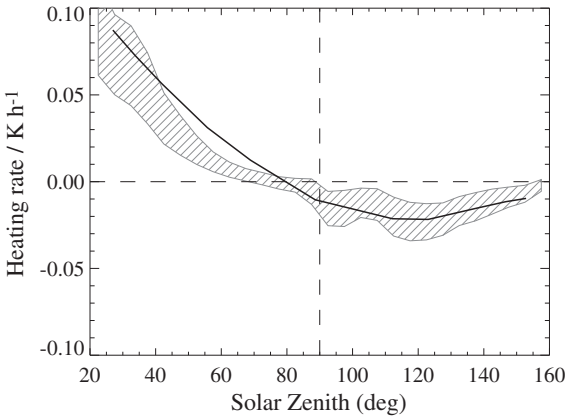

Fig. 7. Rate of change of SST calculated from AATSR D2 $SST_{0.2m}$–drifter SST differences as a function of solar zenith angle. Line: average rate found for different solar zenith angle ranges calculated across the MDB. Shaded band: median *pm* RSD of rate simulated by the skin effect and stratification models for the same matches.

(SZA<90°) warming is strongly dependent on SZA, and is up to 0.1 K h^{-1}, consistent with the ~0.5 K mean amplitude of the daily cycle in the tropics and summer mid-latitudes.

As ocean surface mixing is driven by surface winds, the rate of heating will be strongly dependent on wind speed in addition to solar zenith angle (insolation). This is shown in Fig. 8 where the heating rates are plotted against wind speed for two example ranges of SZA (one day and one night). There is very good agreement between the observed (black bars) and simulated (shaded area) heating rates. Within each range of SZA, the data show a dependence on wind-speed that is well described as an exponential function (solid line). So, an alternative to the model simulated time-corrections is an empirical model, comprising, for each solar zenith angle range, a fit against wind speed of form:

$$\frac{dSST}{dt} = a_0\, exp(a_1 u) + a_2 \tag{2}$$

This empirical fit is less onerous to calculate than the simulated time corrections, and can also be applied to other observations with similar LECTs, including the ATSR-1 and ATSR-2 MDBs (for which we have not modelled the half-hour SST time corrections). Fig. 6 implies that the empirical model can be applied for the in situ observations up to 2 h at least from the 10:30 h overpass time of ATSRs 1 and 2.

These temperature trends can affect both the apparent retrieval bias and noise in validation studies. For instance, if the distributions of satellite-buoy time differences were uniform and the trend applied exactly to all observations, the independent noise introduced would be:

$$\sigma = \frac{dSST}{dt}\frac{\Delta t}{\sqrt{3}} \tag{3}$$

where Δt is the maximum time difference between satellite and buoy. For AATSR vs. drifter SSTs, the expected increase in SD/RSD is less than the above estimate because the time differences are not uniformly distributed. Larger time windows can also lead to bias artefacts if the satellite–buoy time differences are not symmetrically distributed around zero or span a range of local time over which the diurnal cycle is not linear, as is the case for the morning AATSR observations.

Fig. 9 shows the distribution of satellite-buoy time differences in the AATSR drifter and GTMBA datasets. Most drifting buoys are making frequent reports (e.g., hourly) and as a result the majority, ~66%, of matches have time differences less than an hour. However, in the case of the GTMBA data only 23% of matches were within a one-

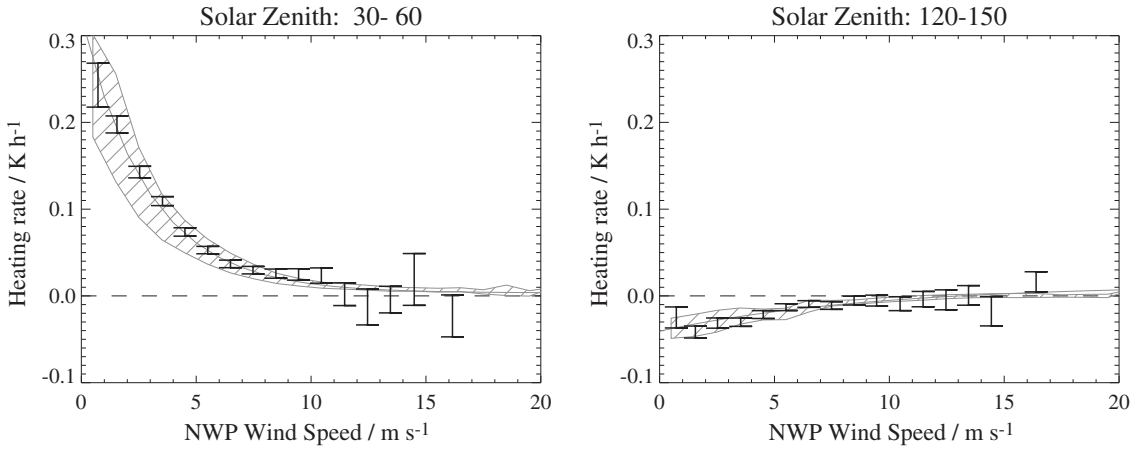

Fig. 8. Surface heating rates as a function of wind speed calculated from observations (black bars) and simulation (shaded grey area). Left panel: subset of day-time matches with solar zenith angles between 30 and 60°. Right panel, night-time matches with solar zenith angles between 120 and 150°.

hour time window, because of how frequently the ATLAS moorings transmit.

Table 7 shows the global statistics of $SST_{0.2m}$–drifter difference for AATSR D2 using different corrections for satellite–buoy time difference: (i) making no correction; (ii) using a single mean heating rate for day and cooling rate for night; (iii) using the SZA-dependent rates shown in Fig. 7; (iv) using the fitted exponential-in-wind-speed model for different SZAs (Eq 2); and (v) using the simulated rates from the Fairall and Kantha–Clayson models. As the AATSR–drifter time differences are already low (see Fig. 9) the effects are relatively modest. Using the full correction reduces the daytime RSD from 0.241 K to 0.230 K (equivalent to removing independent noise of magnitude 0.07 K), which then matches the night-time RSD. A greater improvement in statistics is seen by limiting the data to matches within a one hour time window—day-time RSD reduces to 0.223 K. For comparison, the daytime matches with absolute time differences greater than 1 h have a RSD of 0.286 K without correction for time difference, and 0.258 K with correction.

Table 8 shows the equivalent statistics for the tropical moored buoys. In this case there are far fewer matches and the time differences are much more spread out due to the intermittent data reported though the GTS. Restricting the comparison to the smaller time window causes a significant increase in both RSD and bias due to the reduced sample size. When considering all matches within a three hour window, the time-difference correction reduces day-time RSD from 0.220 K to 0.196 K (equivalent to removal of 0.10 K of independent noise). There is also a positive effect in making day and night observations more consistent on average.

All the tropical moored buoy comparisons show better agreement (lower SD and RSD) with satellite measurement than do the drifting buoy data. In addition, there are clearly fewer outliers affecting the moored buoy comparison as the SDs are only slightly higher than RSDs, while the drifter SDs were approximately double the corresponding RSD. This is despite the tropical regions tending to have lower precision satellite SSTs due to the higher water vapour loading. The lower incidence of outliers seems, therefore, to reflect the higher quality of the tropical moored buoys compared to the drifting buoys in the data sets we use.

There is a trade-off between minimising the satellite-buoy separation, in both time and space, and maximising the number of matches to increase coverage and reduce sampling error. By matching in situ data to the closest ATSR pixel, we apply a spatial separation criteria of nominally 1 km, although it is effectively ~5 km due to the 0.1° resolution of the buoy location reports available to us. Increasing this space-window does not significantly increase the number of matches as the only "extra" ones are outside the satellite swath. When selecting the maximum temporal separation allowed we should consider how the SST can vary between two observations, which is driven by the diurnal cycle (or, sub-daily variability more generally), and the distribution of time separations. For example, a linear temperature trend and a symmetric distribution of time differences will give an unbiased satellite-in situ difference with only modest increase in validation RSD/SD.

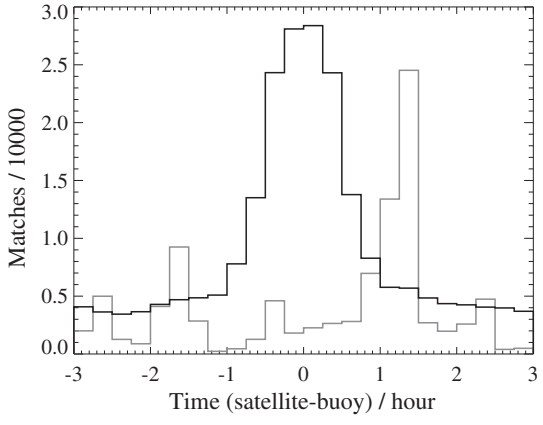

Fig. 9. Distribution of satellite–buoy time differences for drifter (black) and GTMBA (grey) matches. Number of GTMBA matches has been scaled by 10.

Table 7
Median and (in parentheses) RSD, in kelvin, of satellite $SST_{0.2m}$ minus drifter SST, for different methods of satellite–buoy time difference corrections indicated in leftmost column.

	Day		Night	
Time window	±3 h	±1 h	±3 h	±1 h
N. matches	120,916	81,829	96,857	65,963
None	0.066 (0.241)	0.069 (0.223)	0.057 (0.228)	0.056 (0.221)
Global mean rate	0.068 (0.234)	0.068 (0.221)	0.057 (0.228)	0.056 (0.221)
SZA-dependent rate	0.069 (0.232)	0.068 (0.220)	0.057 (0.228)	0.056 (0.220)
Wind and SZA dependent rate	0.068 (0.229)	0.068 (0.219)	0.057 (0.227)	0.057 (0.221)
Modelled	0.068 (0.230)	0.068 (0.220)	0.057 (0.228)	0.056 (0.221)

Table 8
Median and (in parentheses) RSD, in kelvin, of satellite $SST_{1.0m}$ minus GTMBA SST, for different methods of satellite–buoy time difference corrections indicated in leftmost column.

Time window	Day		Night	
	±3 h	±1 h	±3 h	±1 h
N. matches	6837	1449	3081	832
None	0.049 (0.220)	0.105 (0.244)	0.021 (0.181)	0.049 (0.210)
Global mean rate	0.032 (0.197)	0.088 (0.248)	0.030 (0.177)	0.053 (0.208)
SZA-dependent rate	0.029 (0.199)	0.085 (0.248)	0.028 (0.179)	0.052 (0.209)
Wind and SZA dependent rate	0.030 (0.194)	0.084 (0.245)	0.028 (0.180)	0.051 (0.208)
Modelled	0.029 (0.196)	0.086 (0.240)	0.032 (0.179)	0.052 (0.204)

The RMS of the simulated diurnal cycle is 0.03 K h^{-1} at night and 0.10 K h^{-1} during the day, which matches the 0.2 K difference for a two hour time separation identified by Minnett (1991). Simulations of the diurnal heating rates are in good agreement with the observed satellite–drifter differences. In the case of the AATSR–GTMBA matches, where the distribution of time differences is asymmetric with less than 25% of matches within a one hour window, correcting for the diurnal heating effect significantly reduced the satellite-in situ RSD while bringing the day/night differences into closer agreement.

4. Validation

4.1. Global AATSR

In the previous section we use the satellite-in situ differences to explore and quantify the geophysical differences between the types of measurement. We now validate SST-depth estimates derived by adjusting retrieved skin SSTs to SST depth. The adjustments comprise the Fairall model for the skin effects, the Kantha–Clayson model for diurnal stratification, and time-difference effects parameterized in terms of solar zenith angle and wind speed, as described in previous sections. Performance of the new ARC retrievals is assessed partly in comparison with operational coefficients. Identical adjustments to SST depth have been applied, so that the effects of the different skin retrieval schemes can be clearly seen. The operational retrievals have already been shown to have relatively good performance. O'Carroll et al. (2008) performed a three-way uncertainty analysis between operational AATSR, in situ, and AMSR-E SSTs, and found the night time AATSR retrievals to have a lower error (standard deviation ~0.16 K) than in situ buoy measurements (0.23 K, although genuine geophysical variability may also contribute to this value).

As well as basic comparisons, we investigate whether the residuals vary systematically with factors which should affect the satellite retrieval but not the in situ measurements (such as atmospheric water vapour or satellite viewing geometry). Additionally, comparing the residuals from the different channel combinations and both the new ARC and the operational retrieval algorithms are informative about the characteristics of satellite retrieval uncertainties.

Note—we are not using the operational SSTs present in the ATSR multi-mission archive version 2.0, but are implementing the operational retrieval algorithm within the ARC processing chain which ingests Level 1b brightness temperature files. Both sets of SSTs (ARC and "ATS*") are generated from an identical set of input brightness temperatures using identical cloud screening etc., the only difference being the SST retrieval algorithm.

Table 9 shows global statistics comparing satellite-retrieved $SST_{0.2m}$ against drifter SST for both the operational and ARC AATSR retrievals. In all cases the ARC retrieval coefficients result in both a smaller median AATSR–drifter difference and lower RSD. With the operational SST there is a large spread in median AATSR–drifter differences with N2 more than 0.5 K warmer than D2. In practise, the operational nadir-only retrievals are used only in the absence of the D2 (during day) and D3 (at night) retrievals. There is a ~0.2 K difference between the dual-view retrievals, with D3 ~0.1 K warmer than drifters and D2 ~0.1 K cooler. The ARC SSTs are more consistent between types of retrieval: nadir-only N2 and N3 are ~0.03 K warmer than in situ while D2 and D3 are ~0.06 K warmer.

When considering the differences in RSD between the operational and ARC SSTs, the greatest improvement, from 0.45 K to 0.33 K, is seen for the basic N2 SSTs, followed by 0.29 to 0.23 K for the D2. This is due to these retrievals, which use only the 11 and 12 μm channels, being the most sensitive to atmospheric water vapour and therefore benefiting most from the improved ARC algorithm. For the night-only N3 and D3 the decrease in RSD is much lower ~0.01 K. However, the improvement may somewhat be masked because these RSDs are around 0.2 K, and are probably dominated by the random error of the buoy measurements.

Next, consider the differences in RSD between different algorithms. Largest RSD is obtained for N2 SSTs, reflecting the fact that single-view observations at 11 and 12 μm only do not have sufficient information content to deal with the full range of variability of atmospheric absorption. D2 SSTs improve significantly over N2 SSTs in terms of RSD, because the additional view significantly adds to the information content of the observations. The use of the 3.7 μm channel in the N3 SST also adds significant information compared to N2 SST, and N3 inversion does not amplify noise as much as D2 or D3 inversion (the magnitude of N3 coefficients is generally smaller than for the dual-view algorithms). N3 SSTs thus turn out to have the smallest RSD. Note, however, that the D3 SSTs still are less noisy than D2 SSTs. Moreover, the dual-view algorithms generally maintain accuracy better under unusual atmospheric conditions, and only they are highly robust (insensitive) to episodes of elevated stratospheric aerosol.

Table 10 shows global statistics comparing satellite estimated $SST_{1.0m}$ against in situ SSTs from the Global Tropical Moored Buoy

Table 9
Global AATSR estimated $SST_{0.2m}$–drifter in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

	N2	N3	D2	D3
Day (120,916)				
ATS*	0.464 (0.431)		-0.086 (0.290)	
ARC	0.020 (0.317)		0.068 (0.229)	
Night (96,857)				
ATS*	0.448 (0.450)	0.033 (0.200)	-0.108 (0.291)	0.088 (0.223)
ARC	0.031 (0.327)	0.034 (0.192)	0.057 (0.227)	0.055 (0.207)

Table 10
Global AATSR-estimated $SST_{1.0m}$–GTMBA SST in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

	N2	N3	D2	D3
Day (6837)				
ATS*	0.399 (0.389)		-0.155 (0.246)	
ARC	-0.064 (0.369)		0.030 (0.194)	
Night (3081)				
ATS*	0.412 (0.424)	0.055 (0.160)	-0.155 (0.242)	0.073 (0.155)
ARC	-0.043 (0.379)	0.018 (0.128)	0.028 (0.180)	0.036 (0.145)

arrays. Again, ARC SSTs all have lower RSD and more consistent median differences than operational SSTs. There are two important differences compared to Table 9. Firstly, the measurement error associated with the ATLAS moorings is much lower than drifters, and this is reflected in the lower RSDs for N3, D2, and D3 SSTs in Table 10. The ARC N3 SSTs now have a RSD of 0.128 K while the operational N3 SSTs have a RSD of 0.160 K. Secondly, the behaviour of the ARC N2 SSTs appears inconsistent with the previous comparison. While the ARC N2 SST–GTMBA does have lower RSD than the operational N2, it has a higher RSD than in the drifter comparison. This reflects the more

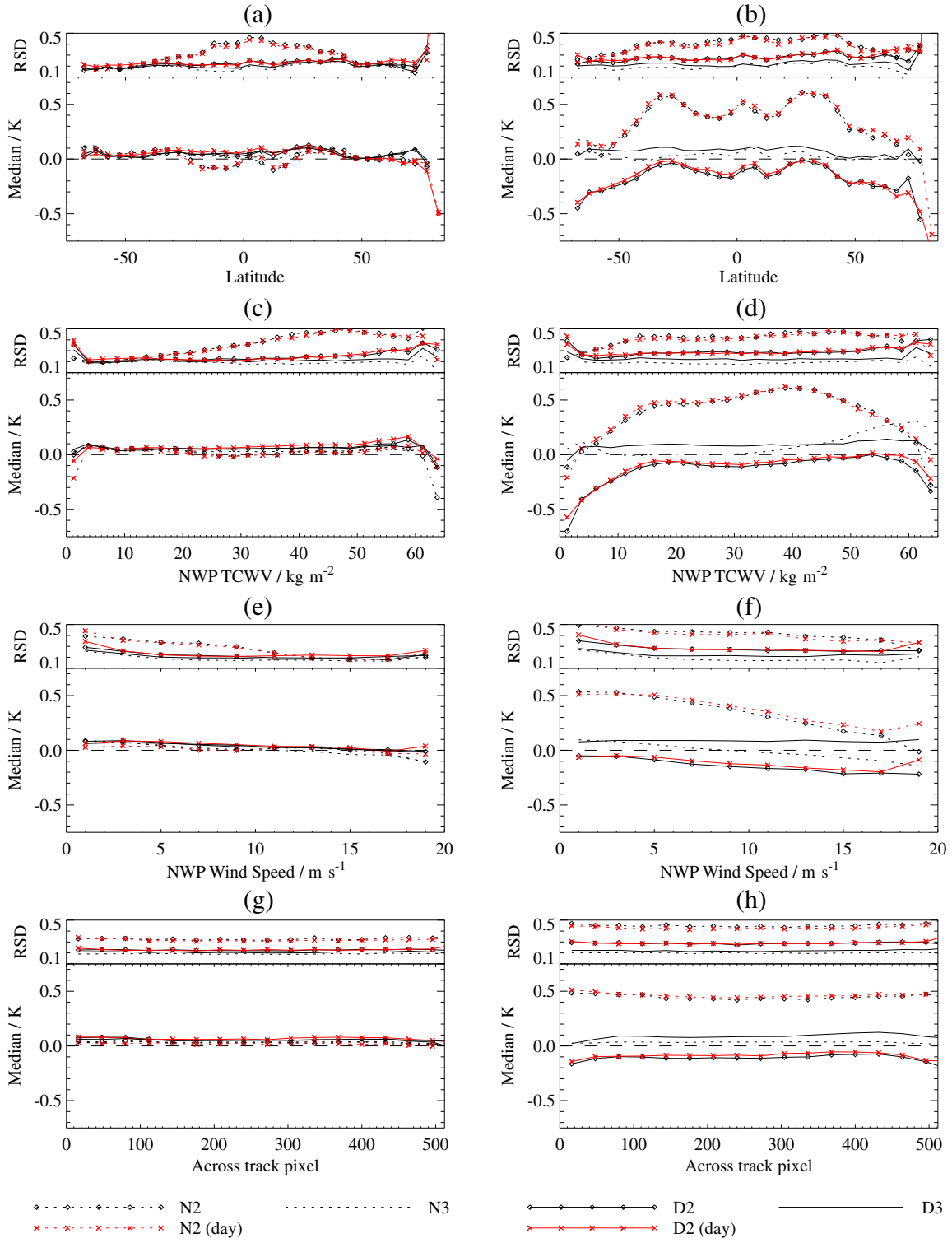

Fig. 10. AATSR estimated $SST_{0.2m}$–drifters as a function of latitude (a,b), TCWV (c,d), wind speed (e,f), and pixel position (g,h). Left column (a,c,e,g) shows ARC retrievals, right column (b,d,f,h) shows operational retrievals. Dashed with symbol—N2; dashed—N3; solid with symbol—D2; solid—D3. Black indicates night-time data; red (with X symbol) indicates day-time data.

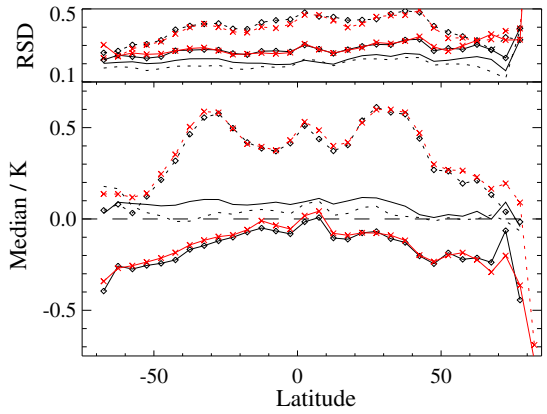

Fig. 11. As Fig. 10b, but including the latitudinal-dependent offset to the D2 SST retrieval (see main text).

limited geographical coverage of the GTMBA arrays which cover only tropical regions. These present difficulties for the split-window N2 retrievals due to the high levels of atmospheric water vapour.

4.2. Regional AATSR

Fig. 10 shows the AATSR estimated SST $_{0.2m}$–drifter SST against four different variables: latitude, NWP total column water vapour (TCWV), NWP wind speed, and across-track pixel position. ARC retrievals have improved inter-algorithm consistency, achieved independently of in situ observations by design (Embury & Merchant, this issue). ARC N3, D2, and D3 estimates are mutually consistent to within 0.05 K in agreement with the global statistics in Table 9. Latitudinal, TCWV and wind speed dependencies are also reduced.

In the latitudinal plots the greatest change between operational (Fig. 10b) and ARC retrievals (Fig. 10a) is seen for N2 and D2 where the operational retrievals have large systematic variations ~0.5 K. With the updated ARC retrievals the majority of these systematic variations are eliminated, with the exception of the N2 latitude bias pattern between 30S and 30N whose latitudinal structure remains. This is also apparent in the RSDs: in Fig. 10b the N2-drifter RSD is high (0.3–0.5 K) at all latitudes, but in Fig. 10a it is reduced to ~0.2 K in mid-latitudes while remaining high in the tropics. These retrieval issues are associated with the high atmospheric water vapour in the tropics which are particularly problematic for single-view split-window SST retrievals (Merchant et al., 2009). The latitudinal variation of the operational D2 SSTs is a known problem with AATSR retrievals—commonly shown in D2–D3 difference plots (Merchant et al., 1999)—and is an intrinsic property of the form of the linear retrieval used operationally (Merchant et al., 2006). A latitude dependent bias correction has been devised (Birks, 2006), which has been applied to the operational D2 SST in Fig. 11, although it does not resolve the underlying problem which is variability with atmospheric state.

There are also two anomalies associated with Arctic retrievals. Firstly, the day-time N2 and D2 retrievals are ~0.5 K cooler than drifters north of 80° N. A similar effect is present in the operational SSTs, but less obvious due to other zonal variations. However, this value is based on only 24 matches to two separate drifters. It is difficult to assess, therefore, whether the biases arise from retrieval problems related to extreme atmospheric conditions, prevalent cloud contamination, or other factors. Secondly, there is a ~0.1 K difference between night-time D2 and D3, and the other retrievals within the Arctic Circle. This may be due to a cloud screening failure—the nadir-only and day-time retrievals are all consistent, while the dual-view night-time retrievals are warmer in comparison. Such an effect would occur if there is a small amount of undetected cloud in the forward view.

Fig. 10d shows the large systematic variability of the operational N2 and D2 retrievals as a function of TCWV. The N3 retrieval also shows a dependence on TCWV over 30 kg m^{-2} reaching ~0.3 K for the wettest atmospheres. Using the ARC retrievals (Fig. 10c) eliminates the majority of this dependence with the exception of TCWV values less than 5 kg m^{-2} or greater than 60 kg m^{-2}. Retrieval-drifter

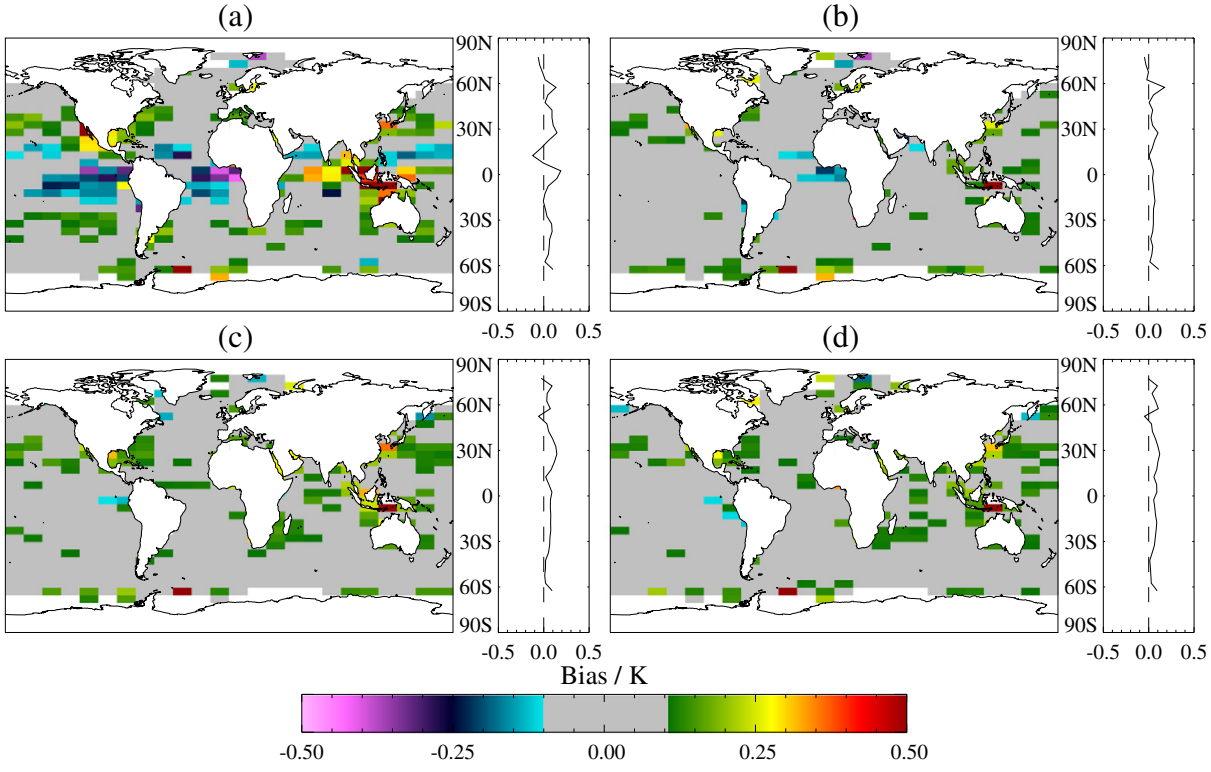

Fig. 12. Median of difference between AATSR-estimated SST$_{0.2m}$ and drifting buoy SST for (a) N2, (b) N3, (c) D2, and (d) D3 retrievals using new coefficients.

differences are 0.1–0.2 K colder for these extremes. As shown in Embury and Merchant(this issue), this is partly expected given the retrieval design which leads to inherent biases at the extremes of the atmospheric conditions used to generate the coefficients. For example, the ARC D2 retrieval is expected from simulation to have a cold bias of ~0.2 K for TCWV ~65 kg m^{-2}; N3 and D3 are, in simulation, less affected and the N2 more so, as observed in practice. There may also be some bias because extremely humid atmospheres are expected to be associated with less reliable cloud masking with a higher chance of residual cloud affecting brightness temperatures used for SST retrievals and causing cold biases.

For moderate TCWV values, the ARC N3, D2, and D3 retrievals show a slight linear trend from ~0.05 K for dry conditions to ~0.1 K for wet conditions—comparable to the operational D3 retrieval. While the ARC N2 retrieval does match the others for dry conditions, it diverges slightly for intermediate values of TCWV (20–40 kg m^{-2}) with N2-drifter differences around 0.0 K. At higher values of TCWV the N2–drifter differences increase closer to the other retrievals. However, when viewing a sufficiently opaque atmosphere, the retrieval relies more on prior information (the climatological correlation of SST and lower atmospheric temperatures) with decreased sensitivity to the observations (Merchant et al., 2009).This can also be seen in the N2–drifter RSD, which increases with TCWV, and is little different from the operational N2–drifter RSD above about 40 kg m^{-2}.

Wind speed dependence is shown in Fig. 10e and f. The operational retrievals all have widely different sensitivities to the wind speed, with N2 the most sensitive and D3 showing no dependence. With the ARC coefficients all four retrievals have the same small dependence on wind speed seen in Section 3.1. The consistency of the residual dependence between the four different ARC retrievals is encouraging. As discussed in Section 3.1, it is likely that a significant part of the residual dependence is an artefact of the skin-effect adjustment, implying that the wind-speed dependence of the new retrievals is yet smaller.

Fig. 10g and h show the retrieval–drifter differences as a function of across track pixel position. Systematic biases in the operational retrievals across track arise from assuming the satellite viewing geometry is a fixed function of the across track pixel position. The N2 retrieval–drifter difference is approximately 0.1 K lower in the centre swath than at the edges, while the D2 and D3 retrievals show a slight "m" shape similar to the simulations shown in Embury and Merchant (this issue). This viewing geometry dependence may also be contributing to the north–south asymmetry affecting the operational D3 (Fig. 10b). The ARC scheme removes these across track biases.

Figs. 12 and 13 show global distributions of the median and RSD of difference between AATSR-estimated $SST_{0.2m}$ and drifter SST. The N3, D2, and D3 retrievals all have good agreement with the in situ data with cell-mean differences mostly <0.1 K and RSDs <0.2 K. The largest satellite–drifter differences are around Svalbard, Indonesia, and in the Southern Ocean and these are all locations where satellite SST retrievals are challenging and there is poor in situ coverage. In all three cases the number of in situ matches is very low (see Fig. 2) as is the number of individual buoys making the reports. The N2 retrieval, however, has a much more varied regional bias pattern, especially along the equator following the high equatorial RSD and zonal differences seen in Fig. 10a. Here we see the equatorial satellite–drifter differences are negative in the Pacific and Atlantic Oceans, but positive in the Indian Ocean and around Indonesia. Similarly the N2 RSD is high at ~0.5 K throughout most of the tropics.

The degree of statistical significance of the median difference of each cell from zero was evaluated using Student's *t* test, assuming that all matches have independent errors. (Correlated errors between different matches to a given buoy mean that the significance levels found in this way are over-estimates.) The test indicated that the calculated median differences are statistically significant at the 90% confidence level for 65 to 70% of latitude–longitude cells depending on the type of retrieval. Over half (50–55%) are significant at the 99% level. In this context, statistically significance indicates that the median satellite–drifter difference is non-zero, but these differences are still mostly less than 0.1 K except for the N2 retrieval. Cells where the observed median differences are not statistically significant include the three areas of low in situ coverage and large (~0.5 K)

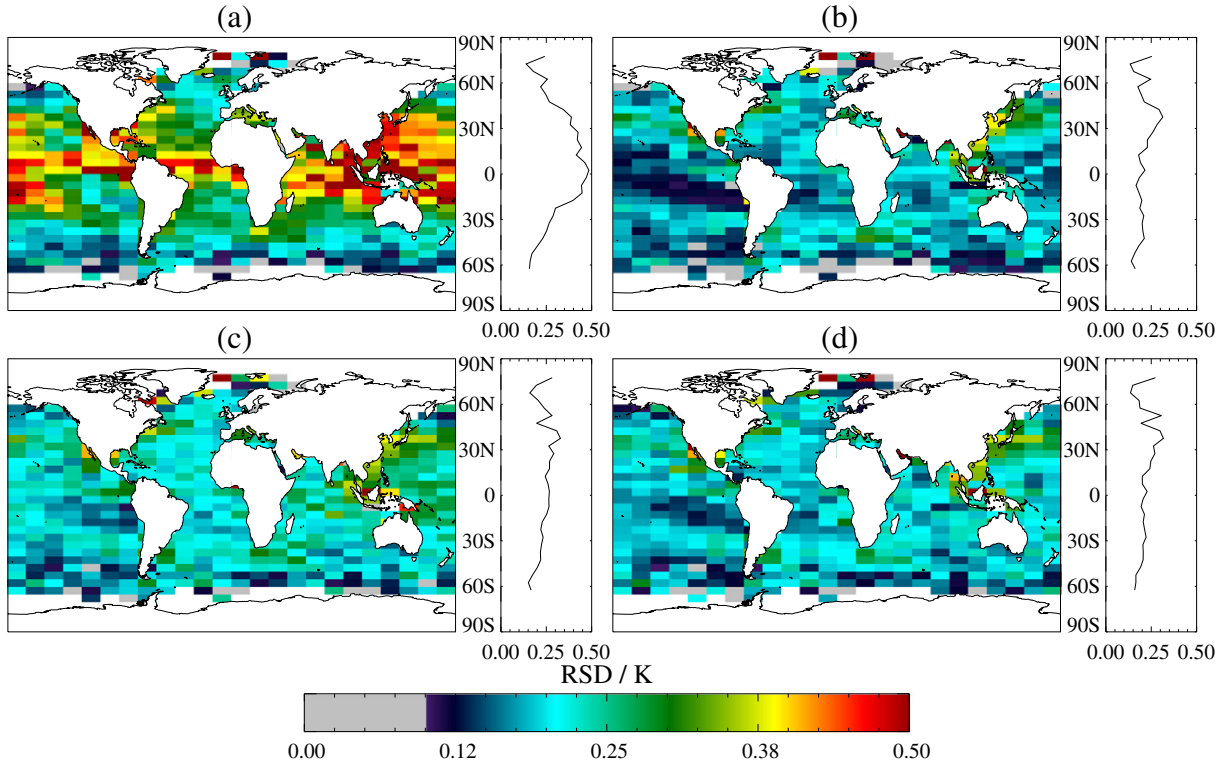

Fig. 13. Robust standard deviation of difference between AATSR-estimated $SST_{0.2m}$ and drifting buoy SST for (a) N2, (b) N3, (c) D2, and (d) D3 retrievals using new coefficients.

Table 11
Global ATSR-2 estimated $SST_{0.2m}$–drifter in Kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

	N2	N3	D2	D3
Day (46,017)				
ATS*	0.574 (0.438)		−0.004 (0.357)	
ARC	0.064 (0.350)		0.138 (0.315)	
Night (36,461)				
ATS*	0.575 (0.446)	0.129 (0.238)	−0.008 (0.335)	0.133 (0.260)
ARC	0.095 (0.351)	0.137 (0.235)	0.143 (0.291)	0.143 (0.250)

satellite–drifter differences discussed previously. The remaining cells which are 'not significant' correspond to cases where the regional biases are very low (~0.01 K for the N3, D2 and D3 retrievals).

One of the targets for ARC (Merchant et al., 2008a) was for regional biases to be within 0.1 K. On the basis of Fig. 12, this objective has been met for AATSR dual–view SSTs relative to drifting buoys across 80% of the global ocean.

4.3. ATSR-2

Table 11 shows global statistics comparing satellite estimated $SST_{0.2m}$ against drifter SST for ATSR-2. As with the AATSR results (c.f. Table 9), the ARC retrieval coefficients result in a lower satellite–drifter RSD and more consistent median differences. However, the satellite estimates of SST are now ~0.14 K warmer than drifters or ~0.1 K warmer than the corresponding AATSR comparison. Furthermore the N2 is ~0.05 K cooler than the other retrievals, a much larger discrepancy than seen for AATSR. Table 12, comparing satellite estimated $SST_{1.0m}$ against in situ SSTs from the Global Tropical Moored Buoy arrays, again shows while the ARC SSTs are more consistent in satellite–buoy differences, the ARC SSTs are significantly warmer ~0.16 K than in situ data. The RSDs for the ATSR-2 N3 and D3 estimates shown in Table 12 are only slightly higher ~0.01 K (equivalent to ~0.05 K independent noise) than the corresponding AATSR values in Table 12, while for the D2 RSD the ATSR-2 value is ~0.06 K higher (equivalent to 0.16 K independent noise) than AATSR.

Fig. 14 is the ATSR-2 equivalent to Fig. 10 and shows the ATSR-2 estimated $SST_{0.2m}$–drifter SST. The same comments made concerning Fig. 10 apply here too, though the satellite–drifter differences are ~0.1 K warmer for the ARC ATSR-2 estimates than the AATSR ones. The ATSR-2 SST estimates are more sensitive to TCWV than the AATSR estimates, in Fig. 14c, the satellite–drifter differences go from ~0.1 K for dry atmospheres to ~0.2 K for wet (values were ~0.05 K and ~0.10 K for AATSR). This also shows in the latitudinal differences, Fig. 14a, where the ATSR-2 tropical estimates are ~0.2 K warmer than drifters. Behaviour of RSD with latitude and TCWV is also different for ATSR-2. While N3 and D3 RSDs were almost independent of TCWV with AATSR, they are now higher (~0.05 K) for low TCWV and decrease to close to the AATSR values for high TCWV; D2 RSDs, however, are elevated regardless of the TCWV value.

Table 12
Global ATSR-2 estimated $SST_{1.0m}$–GTMBA in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

	N2	N3	D2	D3
Day (6097)				
ATS*	0.525 (0.381)		−0.103 (0.263)	
ARC	0.045 (0.358)		0.121 (0.243)	
Night (1675)				
ATS*	0.586 (0.438)	0.208 (0.176)	−0.071 (0.265)	0.123 (0.160)
ARC	0.134 (0.407)	0.181 (0.134)	0.158 (0.238)	0.158 (0.156)

5. Discussion

In the ARC project we aim to retrieve skin SSTs from the satellite observations, and also to generate SST estimates at depths representative of buoy observation depths at a standardised local time. Therefore, in this initial validation, we have considered in detail the relationships between skin SST retrieved from AATSR and in situ observations, with respect to the skin effect, near-surface stratification and the diurnal heating and cooling cycle.

To estimate the skin effect we use the Fairall et al. (1996) model tuned for use with the operational AATSR D3 retrievals in a previous study. Without accounting for the skin effect, there is apparently a ~0.07 K discrepancy between day and night-time SST retrievals relative to in situ observations. Our interpretation is that this relative bias is the average mid-morning modification of the skin effect by insolation. Correction with the Fairall skin model (which accounts for fluxes into the skin layer) reduces the day–night discrepancy to ~0.01 K.

The dependence of the Fairall skin effect model with wind speed is not entirely consistent with the data, with a wind-speed dependent residual of 0.1 K at about 3 m s^{-1} falling close to zero by 10 m s^{-1}. While having a positive effect on bias, applying the Fairall model does slightly increase the scatter of the resulting $SST_{subskin}$–drifter differences. There are various possible reasons for these inconsistencies. There may be wind-speed related biases in the skin SST estimated from AATSR, either directly via inadequately modelled surface emissivity or via confounding of wind-speed with some other factor that causes bias. Moreover, the greater noise apparent in the Fairall model estimates may arise from error in the input fluxes rather than limitations of the model. However, we think the most likely reason is that the Fairall model parameters used in this study were originally intended to use with operational AATSR D3 retrievals. The residual wind-speed dependence seen here is likely an artefact of the different wind-speed sensitivity of the operational SST_{skin} retrievals.

There are two factors which suggest the ARC D3 skin SSTs are less likely to be in error than the operational D3 skin SSTs. Firstly, in validation all the operational retrievals have widely different responses to wind speed while the ARC retrievals are all consistent. Secondly, the independent night-time skin correction of Donlon et al., (2002) has a wind-speed dependence that is consistent with the ARC observations, and does reduce the scatter of $SST_{subskin}$–drifter differences when applied to night-time data.

The second step involved in adjusting $SST_{subskin}$ to drifter-depth is to estimate near-surface stratification. This is based on running a Kantha–Clayson model forced by the time-history of NWP fluxes from the previous dawn (or time of minimum solar elevation). Diurnal stratification within the uppermost few tens of centimetres of the ocean around 10 a.m. local time is generally modest (<0.1 K on average) but is detectable in the observations for wind speeds less than ~3 m s^{-1}. The Kantha–Clayson model we use for thermal stratification does reconcile daytime and nighttime satellite–buoy differences for low wind speeds (<3 m s^{-1}) to ~0.02 K. Alternatively, users of ARC SSTs for climate applications may prefer to exclude potential thermal stratification events based on the model predictions: we find that this approach also reduces the day–night differences and reduces the daytime RSD from 0.24 K to match the night-time value of 0.23 K.

The final step in adjusting skin to drifter-depth SST accounts for the diurnal cycle that is associated with warming during the morning overpass and cooling during the evening overpass. There are two reasons to address this. First, the uncorrected diurnal cycle increases scatter in validation against observations collected over a time window around the satellite time. Moreover, if satellite–buoy time differences are asymmetrically distributed or include the time of minimum or maximum diurnal cycle, the effect can introduce bias artefacts between day and night satellite SSTs. While these problems

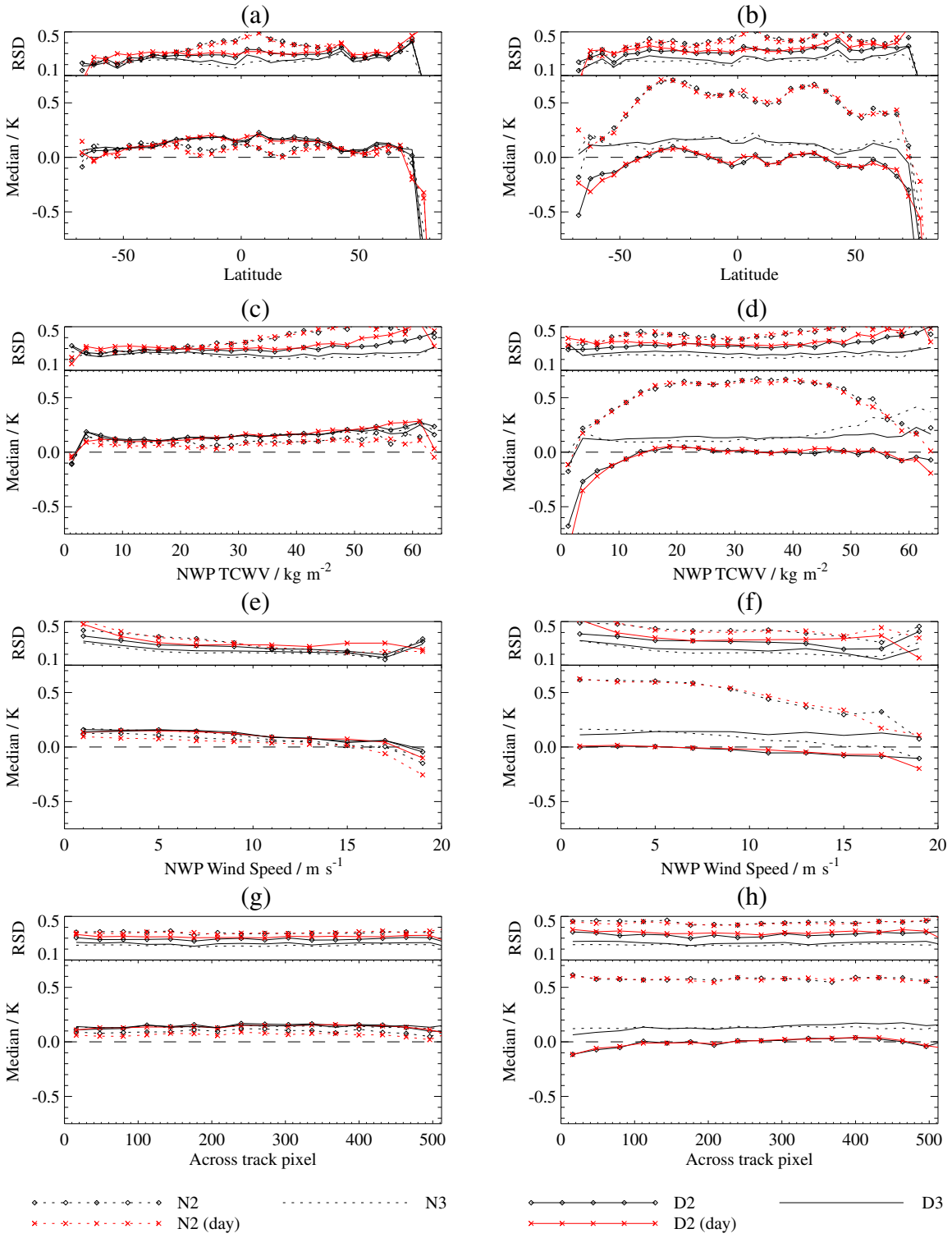

Fig. 14. ATSR-2 estimated $SST_{0.2m}$–drifters as a function of latitude (a,b), TCWV (c,d), wind speed (e,f), and pixel position (g,h). Left column (a,c,e,g) shows ARC retrievals, right column (b,d,f,h) shows operational retrievals. Dashed with symbol—N2; dashed—N3; solid with symbol—D2; solid—D3. Black indicates night-time data; red (with X symbol) indicates day-time data.

can be avoided by using very harsh matching criteria between satellite and in situ (i.e. time differences <1 h or better) this can drastically reduce the number of matches available for validation. Second, there is a difference in overpass time of about 30 min between AATSR and ATSRs 1 and 2, which must be adjusted for in an SST time series for climate applications.

We have explored heating and cooling rates around the satellite observation time both in simulation (using the Kantha–Clayson model referred to above) and empirically. We determined the local surface heating (or cooling) rate during the AATSR overpass from observed time-dependence of satellite-in situ differences, as a function of solar zenith angle and surface wind speed. The observed heating rates show

excellent functional agreement with those calculated with the Kantha–Clayson model.

Because we have a large validation dataset in which most buoys reports are within an hour of the satellite observation time, correcting for the satellite–buoy observation time difference in the AATSR–drifter match-up data base results in only slightly improved RSDs. However, the AATSR–GTMBA matches are limited in number and not so favourably sampled in time, and correcting for the time-difference results in significantly reduced RSDs and better day/night bias agreement, while using harsher matching criteria results in higher RSDs and worse biases.

Our initial validation results for ARC $SST_{0.2m}$ from AATSR against drifter measurements is encouraging. Global biases are around 0.03 K for nadir-only retrievals and 0.05 K for dual-view retrievals. Scatter is also low with RSD values of 0.33, 0.19, 0.23, 0.21 for the N2, N3, D2, and D3 algorithms respectively. However, there is a limitation to the validity of these RSD results. As shown by O'Carroll et al. (2008), quality controlled drifting buoys reported on the GTS themselves have an error of ~0.2 K, so the RSDs of the ARC–drifter comparisons may predominantly reflect the quality of the in situ data rather than the SST retrievals. N3 and D3 SSTs in particular are expected (from simulation) to have scatter markedly less than 0.2 K (Embury & Merchant, this issue). Using the GTMBA dataset based on the more accurate ATLAS moorings, RSDs of 0.38 (N2), 0.13 (N3), 0.18 (D2), and 0.15 (D3) are obtained, which are a better estimation of the point scatter error attributable to the ARC SST depth estimates. The SST depth estimates are formed by adding adjustments for skin and stratification effects to the primary ARC observations, which are skin SST retrievals. With some limitations that have been noted, these adjustments seem effective for systematic aspects of skin-depth SST difference. However, the adjustments tend to add a small amount of scatter, since they have random uncertainty of their own (e.g., see Table 4). The RSDs quoted above therefore give us upper limits to the random uncertainty inherent in the ARC SST_{skin} estimates.

Systematic biases have been observed for previous (operational) ATSR-series SSTs against latitude, TCWV, wind speed and viewing geometry. The magnitude of these systematic effects has been reduced using ARC SSTs to <0.1 K for all except the most extreme regimes, namely, where TCWV <5 kg m^{-2}, TCWV >60 kg m^{-2}, and for latitudes polewards of about 65°. Regional validation show $SST_{0.2m}$–drifter differences are <0.1 K for the N3, D2, and D3 retrievals for 80% of the area of the global ocean. The exceptions are high-latitudes and Indonesian waters, both of which have very sparse in situ coverage. The N2 is the only retrieval with significant regional SSTs biases exceeding 0.1 K; the greatest variations are seen in the tropics, where N2 SSTs are several tenths of a degree cool in the Pacific and Atlantic, and several tenths of a degree warm in the Indian Ocean and around Indonesia. These patterns have some resemblance to those described for AVHRR in Merchant et al. (2009) and it should be noted that the limitations of the Pathfinder "split-window" SSTs discussed in that paper will apply to some degree also to ARC N2 SSTs, despite the ARC coefficients being banded by TCWV.

Initial validation of ARC ATSR-2 SSTs show a reduction in functional and regional biases comparable to that found for AATSR. Likewise, there are also reduced RSDs compared to operational algorithms. The global $SST_{0.2m}$–drifter differences, while more consistent than the operational retrievals, are around 0.1 K warmer than their AATSR equivalents resulting in a median $SST_{0.2m}$–drifter difference of ~14 K. However, as shown in Fig. 15 the ATSR–2$SST_{0.2m}$–drifter differences change significantly over time. All the ATSR-2 retrievals are somewhat warmer, ~0.15 K, than drifters during the early part of the ATSR-2 mission (1996–1998) but the N2 retrieval in particular cools towards the end of the ATSR-2 mission. During the overlap between the two satellites only the N3 and D3 retrievals show the ~0.1 K difference between ATSR-2 and AATSR, for D2 the difference is reduced to ~0.05 K, and N2 is <0.01 K.

There are several causes which could be contributing to the ATSR2–AATSR discrepancy. It is most likely to arise from calibration difference between the two instruments. Given that the process of SST retrieval

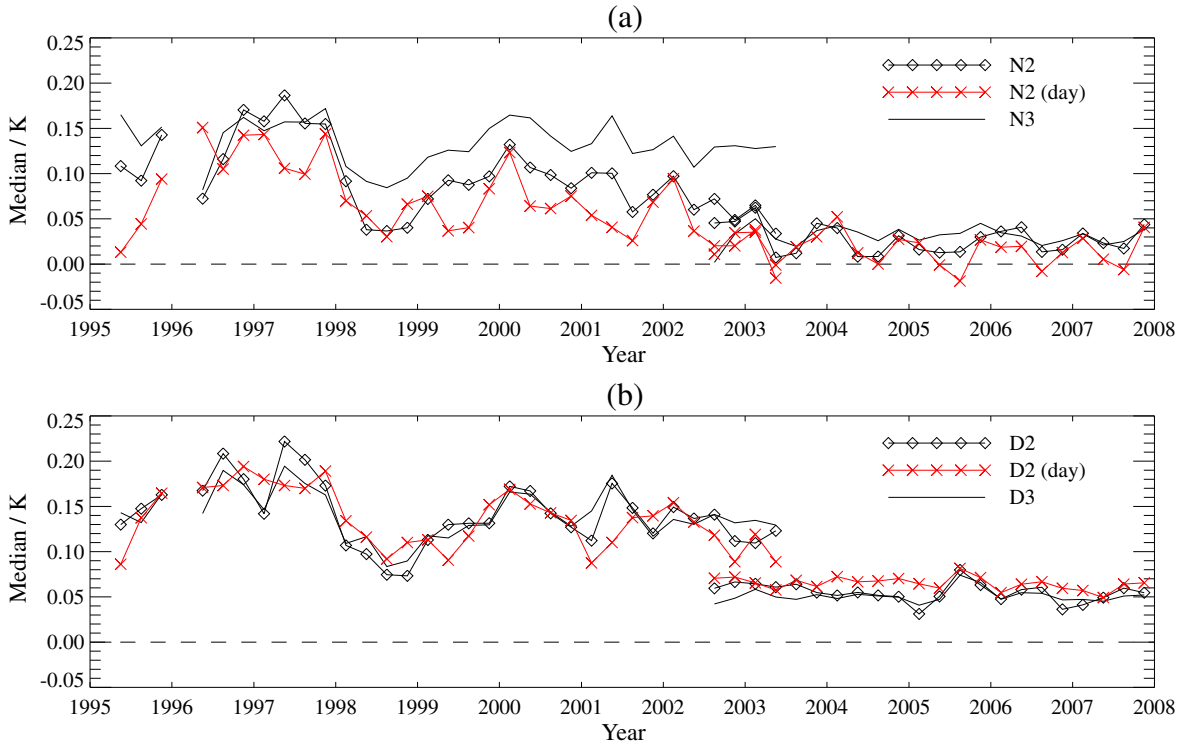

Fig. 15. Time series of ATSR-2 (up to mid-2003) and AATSR (from mid-2002) estimated $SST_{0.2m}$–drifters. (a) Nadir-only retrievals: solid with symbol N2; solid N3. (b) Dual-view retrievals: solid with symbol D2; solid D3. Red lines (with X symbol) indicates day-time data.

tends to amplify any calibration errors (Merchant & Le Borgne, 2004), this is quite plausible, and could arise by chance even if all channel calibrations were accurate to <0.1 K. Furthermore a change in calibration over time could lead to the inter-algorithm differences seen with ATSR-2. Another cause (which we consider less plausible) is inhomogeneity in the ARC processing inputs over time. The ARC processing uses prior NWP information for both cloud detection and to determine the TCWV that selects the SST coefficients used. The NWP fields are taken from ERA-40 and, after the end of the ERA-40 time series in August 2002, ECMWF-operational analyses. We consider this potential cause less plausible for a 0.1 K offset, since neither cloud detection nor SST retrieval are strongly responsive to the NWP state. Lastly, it could be an artefact of changes in the in situ record itself. The number of drifter matches in 2000 was double that in 1997. This can affect ATSR2–drifter differences if the new additions are altering the geographical distribution of the buoys as the ATSR-2 SST retrievals have a greater regional variability (~0.1 K) than the AATSR ones. As part of future work to ensure homogeneity of the ARC SST record a detailed overlap analysis will be performed using common in situ data so that ATSR-2/AATSR SSTs (and brightness temperatures) can be directly compared, and ATSR-2/AATSR SSTs will be compared against quality controlled in situ data selected for high stability.

This initial validation has established a high degree of consistency between ATSR skin SST retrievals, day and night, for ATSR-2 and AATSR, over most of the regimes of the global oceans. The subtle variability of the ocean thermal skin and diurnal stratification has been estimated to adjust the ATSR skin SST retrievals to form SST_{depth} estimates, which are more directly comparable to the buoy observations available for validation. We have been able to validate the systematic behaviour of the skin and stratification models used, and have shown that ATSR-2 and AATSR SST_{depth} can be adjusted to a common satellite overpass time with good confidence, thereby minimising aliasing of the diurnal cycle into the multi-mission record of SST. This all represents significant progress towards an independent "climate quality" (Merchant et al., 2008a) SST record capable of refining our knowledge of marine climate change over recent decades.

Acknowledgements

The ATSR Reprocessing for Climate project is jointly funded by the Natural Environment Research Council, the Ministry of Defence, and the Department of Energy and Climate Change (previously Defra). Archived ATSR data and computer facilities are provided by the NERC Earth Observation Data Centre (http://www.neodc.rl.ac.uk) and we would like to thank Matt Pritchard for his work on the multi-mission archive and his assistance with the ARC project. We would also like to thank John Kennedy for access and support for the Met Office in situ SST dataset, and Karsten Fennig for adapting the Met Office SST skin-to-depth correction code for use in the ARC project.

References

Barton, I. J. (1998). Improved techniques for the derivation of sea surface temperatures from ATSR data. *Journal of Geophysical Research*, *103*(C4), 8139–8152 URL. http://www.agu.org/pubs/crossref/1998/97JC02569.shtml

Birks, A. R. (2006). Latitude Dependent Bias Correction. AATSR Technical Note. Didcot, UK: Rutherford Appleton Laboratory URL. http://envisat.esa.int/pub/ESA_DOC/ENVISAT/AATSR/Latitude_Dependent_Bias_Correction.pdf

Bourlès, B., Lumpkin, R., McPhaden, M. J., Hernandez, F., Nobre, P., Campos, E., et al. (2008). The pirata program: History, accomplishments, and future directions. *Bulletin of the American Meteorological Society*, *89*(8), 1111 URL. http://dx.doi.org/10.1175/2008BAMS2462.1

Corlett, G., Barton, I., Donlon, C., Edwards, M., Good, S., Horrocks, L., 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 URL. http://dx.doi.org/10.1016/j.asr.2005.09.037

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., et al. (Feb. 2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*(4), 353–369 URL. http://dxdoi.org/10.1775%2F1520-0442%282002%29015%3C0353%3ATIVOSS%3E2.0.C0%3B2

Embury, O., Merchant, C. J., this issue. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. Remote Sensing of Environment vol (num), pp–pp. URL http://dx.doi.org/doi

Embury, O., Merchant, C. J., Filipiak, M. J., this issue. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer. Remote Sensing of Environment vol (num), pp–pp. URL http://dx.doi.org/10.1016/j.rse.2010.10.016

Emery, W. J., Baldwin, D. J., Schlüssel, P., & Reynolds, R. W. (2001). Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, *106*(C2), 2387–2405 URL. http://dx.doi.org/10.1029/2000JC000246

Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B., & Young, G. S. (1996). Cool-skin and warm-layer effects on sea surface temperature. *Journal of Geophysical Research*, *101*(C1), 1295–1308 URL. http://dx.doi.org/10.1029/95JC03190

Gentemann, C. L., Minnett, P. J., Le Borgne, P., & Merchant, C. J. (2008). Multi-satellite measurements of large diurnal warming events. *Geophysical Research Letters*, *35*(22) URL. http://dx.doi.org/10.1029/2008GL035730

Gentemann, C. L., Minnett, P. J., & Ward, B. (2009). Profiles of ocean surface heating (POSH): A new model of upper ocean diurnal warming. *Journal of Geophysical Research*, *114*, C07017 URL. http://dx.doi.org/10.1029/2008JC004825

Horrocks, L. A., Candy, B., Nightingale, T. J., Saunders, R. W., O'Carroll, A., & Harris, A. R. (2003). Parameterizations of the ocean skin effect and implications for satellite-based measurement of sea-surface temperature. *Journal of Geophysical Research*, *108*(C3) URL. http://dx.doi.org/10.1029/2002JC001503

Horrocks, L. A., Harris, A. R., & Saunders, R. W. (2003). Modelling the diurnal thermocline for daytime bulk SST from AATSR. *Forecasting Research Technical Report 418, MetOffice* URL. http://archipelago.uma.pt/pdflibrary/Horrocksetal2003TechinicalReport.pdf

Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.

Kantha, L. H., & Clayson, C. A. (1994). An improved mixed layer model for geophysical applications. *Journal of Geophysical Research*, *99*(C12), 25235–25266 URL. http://dx.doi.org/10.1029/94JC02257

Llewellyn-Jones, D., & Remedios, J., (this issue). The Advanced Along Track Scanning Radiometer (AATSR) and its predecessors ATSR-1 and ATSR-2: an introduction to the special issue. *Remote Sensing of Environment*.

Lukas, R. (1991). The diurnal cycle of sea surface temperature in the western equatorial pacific. *TOGA Notes*, *2*, 1–5.

Mackie, S., Merchant, C. J., Embury, O., & Francis, P. (2010). Generalized bayesian cloud detection for satellite imagery. part 2: Technique and validation for daytime imagery. *International Journal of Remote Sensing* URL. http://dx.doi.org/10.1080/01431160903051703

McPhaden, M. J. (1995). The tropical atmosphere ocean (TAO) array is completed. *Bulletin of the American Meteorological Society*, *76*(5), 739–741.

McPhaden, M. J., Meyers, G., Ando, K., Masumoto, Y., Murty, V. S. N., Ravichandran, M., et al. (2009). RAMA: the research moored array for African-Asian-Australian monsoon analysis and prediction. *Bulletin of the American Meteorological Society*, *90*(4), 459 URL. http://dx.doi.org/10.1175/2008BAMS2608.1

Merchant, C. J., Filipiak, M. J., Le Borgne, P., Roquet, H., Autret, E., Pioll, J., et al. (2008). Diurnal warm-layer events in the western Mediterranean and European shelf seas. *Geophysical Research Letters*, *35*(4) URL. http://dx.doi.org/10.1029/2007GL033071

Merchant, C. J., & Harris, A. R. (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements. *Journal of Geophysical Research*, *104*(C10), 23579–23590 URL. http://dx.doi.org/10.1029/1999JC900106

Merchant, C. J., Harris, A. R., Maturi, E., & Maccallum, S. (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society*, *131*(611), 2735–2755 URL. http://dx.doi.org/10.1256/qj.05.15

Merchant, C. J., Harris, A. R., Murray, M. J., & Závody, A. M. (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 1. Theory, modeling and interalgorithm comparison. *Journal of Geophysical Research*, *104*(C10), 23565–23578 URL. http://dx.doi.org/10.1029/1999JC900105

Merchant, C. J., Harris, A. R., Roquet, H., & Le Borgne, P. (2009). Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer. *Geophysical Research Letters*, *36*(17) URL. http://dx.doi.org/10.1029/2009GL039843

Merchant, C. J., Horrocks, L. A., Eyre, J. R., & O'Carroll, A. G. (2006). Retrievals of sea surface temperature from infrared imagery: Origin and form of systematic errors. *Quarterly Journal of the Royal Meteorological Society*, *132*(617), 1205–1223 URL. http://dx.doi.org/10.1256/qj.05.143

Merchant, C. J., & Le Borgne, P. (2004). Retrieval of sea surface temperature from space, based on modeling of infrared radiative transfer: Capabilities and limitations. *Journal of Atmospheric and Oceanic Technology*, *21*(11), 1734 URL. http://dx.doi.org/10.1175/JTECH1667.1

Merchant, C., Llewellyn-Jones, D., Saunders, R., Rayner, N., Kent, E., Old, C., et al. (2008). Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Advances in Space Research*, *41*(1), 1–11 URL. http://dx.doi.org/10.1016/j.asr.2007.07.041

Minnett, P. J. (1986). A numerical study of the effects of anomalous north atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space. *Journal of Geophysical Research*, *91*(C7), 8509–8521 URL. http://dx.doi.org/10.1029/JC091iC07p08509

Minnett, P. J. (1991). Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *Journal of Geophysical Research*, *96*(C10), 18475–18489 URL. http://dx.doi.org/10.1029/91JC01816

Murray, M. J., Allen, M. R., Merchant, C. J., Harris, A. R., & Donlon, C. J. (2000). Direct observations of skinbulk SST variability. *Geophysical Research Letters*, *27*(8), 1171 URL. http://dx.doi.org/10.1029/1999GL011133

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 URL. http://dx.doi.org/10.1175/2007JTECHO542.1

O'Carroll, A. G., Watts, J. G., Horrocks, L. A., Saunders, R. W., & Rayner, N. A. (2006). Validation of the AATSR meteo product sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, *23*(5), 711 URL. http://dx.doi.org/10.1175/JTECH1876.1

Rayner, N. A., Brohan, P., Parker, D. E., Folland, C. K., Kennedy, J. J., Vanicek, M., et al. (2006). Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the Mid-Nineteenth century: The HadSST2 dataset. *Journal of Climate*, *19*(3), 446 URL. http://dx.doi.org/10.1175/JCLI3637.1

Saunders, P. M. (1967). The temperature at the ocean–air interface. *Journal of the Atmospheric Sciences*, *24*(3), 269–273 URL. http://dx.doi.org/10.1175/1520-0469(1967)024<0269:TTATOA>2.0.CO;2

Simmons, A., Uppala, S., Dee, D., & Kobayashi, S. (2007). ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, *110*, 25–35.

Worley, S. J., Woodruff, S. D., Reynolds, R. W., Lubker, S. J., & Lott, N. (2005). ICOADS release 2.1 Data and products. *International Journal of Climatology*, *25*(7), 823–842 URL. http://dx.doi.org/10.1002/joc.1166

Závody, A. M., Mutlow, C. T., & Llewellyn-Jones, D. T. (1995). A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. *Journal of Geophysical Research*, *100*(C1), 937–952 URL. http://www.agu.org/journals/jc/v100/iC01/94JC02170/

Závody, A. M., Mutlow, C. T., & Llewellyn-Jones, D. T. (2000). Cloud clearing over the ocean in the processing of data from the Along-Track scanning radiometer (ATSR). *Journal of Atmospheric and Oceanic Technology*, *17*(5), 595–615 URL. http://journals.allenpress.com/jrnlserv/?request=get-abstract issn=1520-0426,volume=17, page=595.

Zeng, X., & Beljaars, A. (Jul. 2005). A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophysical Research Letters*, *32*, L14605 URL. http://dx.doi.org/10.1029/2005GL023030