Available online at www.sciencedirect.com
ScienceDirect
Advances in Space Research 41 (2008) 1–11
ADVANCES IN SPACE RESEARCH (a COSPAR publication)
www.elsevier.com/locate/asr

Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers

C.J. Merchant [a,*], D. Llewellyn-Jones [b], R.W. Saunders [c], N.A. Rayner [c], E.C. Kent [d], C.P. Old [a], D. Berry [d], A.R. Birks [e], T. Blackmore [c], G.K. Corlett [b], O. Embury [a], V.L. Jay [e], J. Kennedy [c], C.T. Mutlow [e], T.J. Nightingale [e], A.G. O'Carroll [c], M.J. Pritchard [e], J.J. Remedios [b], S. Tett [a]

[a] *School of GeoSciences, The University of Edinburgh, Crew Building, Edinburgh EH9 3JN, UK*
[b] *Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK*
[c] *Met Office, Fitzroy Road, Exeter EX1 3PB, UK*
[d] *National Oceanography Centre Southampton, European Way, Southampton SO14 3ZH, UK*
[e] *Space Science and Technology Department, Rutherford Appleton Laboratory, Chilton OX11 0QX, UK*

Received 29 September 2006; received in revised form 7 July 2007; accepted 27 July 2007

Abstract

We describe the approach to be adopted for a major new initiative to derive a homogeneous record of sea surface temperature for 1991–2007 from the observations of the series of three along-track scanning radiometers (ATSRs). This initiative is called (A)RC: (Advanced) ATSR Re-analysis for Climate. The main objectives are to reduce regional biases in retrieved sea surface temperature (SST) to less than 0.1 K for all global oceans, while creating a very homogenous record that is stable in time to within 0.05 K decade^{-1}, with maximum independence of the record from existing analyses of SST used in climate change research. If these stringent targets are achieved, this record will enable significantly improved estimates of surface temperature trends and variability of sufficient quality to advance questions of climate change attribution, climate sensitivity and historical reconstruction of surface temperature changes. The approach includes development of new, consistent estimators for SST for each of the ATSRs, and detailed analysis of overlap periods. Novel aspects of the approach include generation of multiple versions of the record using alternative channel sets and cloud detection techniques, to assess for the first time the effect of such choices. There will be extensive effort in quality control, validation and analysis of the impact on climate SST data sets. Evidence for the plausibility of the 0.1 K target for systematic error is reviewed, as is the need for alternative cloud screening methods in this context.

Keywords: Sea surface temperature; Climate change; Radiometers; ATSR

1. Introduction

Both science and society need changes in global sea surface temperature (SST) over recent decades to be more robustly quantified, at global and regional scales. This paper describes a new five-year project to address this need, called (A)RC – (Advanced) along-track scanning radiometer Re-analysis for Climate.

Global SST is a major component of the observational record used to assess climate change (IPCC, 2001). Analyses (historical reconstructions) of global SST have been made at climate research centres internationally (e.g., Kaplan et al., 1998; Rayner et al., 2006). These are based on ship and buoy measurements that are largely in common between the analyses. For recent decades, some analysts (e.g., Reynolds et al., 2002; Rayner et al., 2003) also blend

* Corresponding author. Tel.: +44 131 650 5097; fax: +44 131 662 0478.
E-mail address: c.merchant@ed.ac.uk (C.J. Merchant).

0273-1177/$30 © 2007 COSPAR. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.asr.2007.07.041

in satellite-derived SSTs that are usually themselves empirically tied to subsets of the same in situ measurements (e.g., the Pathfinder data set derived from the advanced very high resolution radiometer, AVHRR). Each data source may be characterized by different random and systematic errors which change in time (e.g., Kent and Taylor, 2006), and the balance of data sources in the observing network also evolves, potentially introducing artefacts into the SST record. SST analyses have larger errors over the regions of the ocean that have been sparsely sampled, such as the seas south of 20°S. In 2003, no less than 25% of the global ocean did not have even a single in situ SST measurement reported (i.e., available to the Hadley Centre) within a 5° latitude–longitude cell in every calendar month, but did have SSTs available in every month observed by the advanced ATSR (AATSR).

Extensive research has attempted to characterize and minimize artefacts in existing SST analyses. Independent validation of the details of temperature changes apparent in existing analyses is highly desirable, and we present in this paper our approach to developing a high-quality, independent SST record by exploiting the along-track scanning radiometers (we will hereafter refer to this series of sensors as the (A)ATSRs, encompassing the first along track scanning radiometer, ATSR, the second, ATSR-2, and the subsequent AATSR, which is currently operational on the Envisat platform). The new SST record must span more than 15 years and to meet stringent requirements on accuracy and homogeneity. The target is stability (constancy in the absence of real SST changes) of better than 0.05 K decade^{-1} in regional averages, in order to quantify observed regional and global trends that are of magnitude 0.2 K decade^{-1} and to support studies of climate-change attribution. Thorough characterization of the random and systematic errors in the record is also needed.

Creating a homogeneous SST record with the required stability is challenging but feasible using (A)ATSR data, and we present in this paper evidence to support this view. (A)ATSR data have several features that make them uniquely suitable. Firstly, (A)ATSR brightness temperatures are exceptionally well calibrated. The two-point calibration system is specified to drift less than 0.03 K over each sensor lifetime (1-standard-deviation drift over ~5 years; Mason et al., 1996; Smith et al., 2001). Secondly, the instrument response functions are sufficiently well characterized to support accurate forward modelling of observations, which means that cloud screening and SST estimation can be designed globally without empirical tuning: validations of existing (A)ATSR SSTs show consistency *between missions* of ±0.2 K (without measures to homogenise the SST record). Thirdly, the (A)ATSRs' dual-view capability allows SSTs to be robust to atmospheric conditions, including extreme water vapour and stratospheric aerosol (impacts <0.1 K). Finally, (A)ATSRs meet other criteria for satellite monitoring of global climate, such as consistent overpass time and existence of adequate overlap periods.

We proceed in this paper as follows. In Section 2, we give an overview of our research objectives and explain how this initiative is distinctive from the approach and potential of the AVHRR Pathfinder (Kilpatrick et al., 2001). In Section 3, we outline our overall approach, emphasizing innovative aspects of our plans. Our approach is founded on recent insights and progress in radiative transfer modelling, cloud screening and SST retrieval theory, and these are discussed in Section 4; this is therefore the main part of the paper where completed original work is reported.

2. Objectives for global SST

2.1. Requirements for advanced climate applications

Our targets for the new (A)RC SST record have been defined by posing the question: what would be the characteristics of a satellite SST record that would bring a significant advance in climate change research? Colleagues at the Hadley Centre for Climate Prediction and Research, defined the following list of requirements:

1. Minimum length of record required of 15 years.
2. High degree of independence of in situ records.
3. Target stability 0.05 K decade^{-1}, regionally
   i. Discontinuities understood and removed.
   ii. Inconsistencies between sensors <<0.1 K.
4. Biases reconciled to within <0.1 K, all regions (i.e., on scales of a few thousand kilometres).
5. "Bulk" and "skin" SST estimates available.
6. Comprehensive error characterization with respect to:
   i. Retrieval errors (random and systematic).
   ii. Other errors (sampling, cloud screening methodology, etc.).

Having stated the requirements, we outline how this new project will meet each of them in turn.

A near-continuous record for mid-1991 to end-2007 will be created from the (A)ATSRs in the timeframe of our project. It is worthwhile commenting on the value of such a record to climate research, since it is sometimes stated that this length of time-series is too short to be of great value.

This sixteen-year interval is crucial in analysing climate trends because the 1990s to the present are an important period of relatively rapid change. Observed global rates of change since 1991 have, on current estimates, a low (<1%) probability of arising in any given 15 year period by unforced decadal variability alone (Collins et al., 2001). The sensitivity of climate to greenhouse gas forcing is a critical parameter in assessing the risks of future climate change, and more precise determination of recent trends will further efforts to quantify this parameter.

Nonetheless, SST records pre-dating the satellite era are required for many purposes, and can only be based on historical in situ data. There are two important means by which the (A)RC SST record will benefit the long-term data sets.

The first is that (A)RC SSTs, because of their low bias and high global-and-temporal consistency, will help to characterize biases and artefacts in the recent in situ SST (Kent and Kaplan, 2006). For example, characterization of biases in recent dipped-bucket observations from ships may be valid back to the 1970s. The second relates to reconstruction methods used to create spatially complete fields of past SST where SST observations are absent for significant domains in space and time (e.g. Kaplan et al., 1998; Rayner et al., 2003; Smith and Reynolds, 2004). These rely on knowledge of modes (spatio-temporal patterns) of SST variability on monthly to decadal time-scales, evaluated from more recent decades which have the most complete data. The (A)RC SST record may improve the robustness with which these modes are evaluated, especially for sparsely-observed regions.

(A)RC SSTs will also be used at the Hadley Centre to prescribe SSTs in contemporary simulations using atmospheric general circulation models, for climate model validation, to re-evaluate inter-annual ocean SST variability, and for assimilation into ocean-only and coupled model runs for decadal forecasting.

There will be many more applications in the wider oceanography and climate communities, some of which require a high degree of independence of other climate data, and of in situ SST in particular. Complete independence is, however, unattainable: *all* cloud screening and retrieval schemes embody knowledge of the range of possible states of the surface and atmosphere that must ultimately be traceable back to in situ measurements, even if such knowledge is embodied indirectly (e.g., in the threshold set in a test for detecting cloud-affected radiances). We aim in this project to maximize the degree of independence of the (A)RC SSTs from in situ temperatures used in climate data sets, but are prepared to sacrifice some independence in order to achieve our accuracy and stability targets. Section three discusses how some in situ SSTs will be used to ensure the stability of the new record.

The target for SSTs biases to be <0.1 K in all regions throughout the record is challenging, but we provide evidence in section four to show that recent progress makes it feasible.

Satellite derived SSTs are actually sensitive to the radiometric temperature of the ocean surface, which is intimately related to the ocean skin temperature. The bulk water temperature at depths of order a metre can differ significantly from the skin temperature because of the thermal skin effect (Saunders, 1967) and near-surface diurnal warming of water (e.g., Stramma et al., 1986). The day and night dependencies on wind speed of these effects were explored with ATSR data by Murray et al. (2000). (A)ATSR SSTs are skin temperature estimates based on radiative transfer modelling (Závody et al., 1995; Merchant and Le Borgne, 2004). Skin temperature is useful in its own right for some applications (such as calculating air-sea fluxes of heat and gas, e.g., Fairall et al., 2003), whereas the bulk temperature is usually required for climate applications (since it is more representative of the heat content of the upper ocean).

The last objective is to characterize the systematic uncertainties and random errors in the (A)RC SSTs comprehensively. A full discussion of this must be reserved for a future article. Here, we comment only that an important contribution to assessing systematic uncertainties is made by the "multiple stream" approach described in Section three.

(A)ATSR SSTs currently exist and are distributed by ESA and NEODC (the Earth Observation Data Centre of the UK Natural Environment Research Council, NERC). These SST products are established through validation to be of very good quality with errors usually better than ~0.3 K. Nonetheless, to meet our targets requires significant improvements over existing SST products, as summarized in Table 1.

2.2. Wider context

The project is currently envisaged to re-analyse the (A)ATSR record for 1991–2007. The ATSR-2 on ERS-2 remains operational and within radiometric specification more than 10 years after launch. Envisat appears capable of exceeding its design lifetime of 5 years (its launch was in 2002), and thus there is some prospect of AATSR SST acquisition until 2009, which would give an 18 year record overall. It looks unlikely that an ATSR-class SST sensor will be flown in time to overlap with AATSR in order to give continuity beyond 2009.

Prior to 1991, the main available SST sensor is the AVHRR. Reprocessing efforts using AVHRR data include Pathfinder and a re-analysis activity within the Global Ocean Data Assimilation Experiment High Resolution SST pilot project (Donlon et al., 2007). These re-analyses are very welcome, but are unlikely to result in an SST time-series prior to 1991 with accuracy, stability or degree of independence from in situ measurements comparable to the (A)RC SST record. Experience in defining AVHRR SSTs by radiative transfer modelling has shown that the single-view and poorer calibration, characterisation and instrumental stability of AVHRRs lead to variable regional biases that are ~0.5 K. For this reason, most AVHRR SST products are based on empirical regression to in situ data, i.e., do not attempt independence from the in situ record. Since AVHRR SST retrieval algorithms are tuned to the regions where the in situ data are prevalent, there is significant uncertainty about the variation of bias in time and space of AVHRR SSTs in regions of ocean that are sparsely observed by direct measurements of temperature. For these reasons, it is difficult to see how the requirements discussed in Section 2.1 can be met for the pre-(A)ATSR era.

3. Overview of approach

3.1. Practicalities

The (A)RC SSTs will be generated from a uniform archive of level 1b data (calibrated, geolocated top-of-atmosphere brightness temperature and reflectance) in

Table 1
Estimates of aspects of current (A)ATSR SSTs and how they will be improved

Aspect	Current value	Means of improvement	Expected outcome
"Global" bias compared to all buoy data	Between −0.05 and +0.17 K (depending on sensor, algorithm etc.)	New emissivity model used for simulations	Sensitivity study shows all retrievals are affected such as to reduce magnitude of bias by 0.05–0.1 K
		Line-by-line (rather than parameterised) transmittance model, and full treatment of scattering by aerosols	Improvements ~0.05 K
Consistency of "global" biases between sensors	~0.1 K	Make homogeneous using overlap periods	Consistency to ~0.01 K
Drifts in "global" SST (various causes)	~0.07 K yr^{-1} (ATSR-1, mainly a detector- temperature effect)	Full correction of ATSR-1 detector-temperature effect Account for trends in trace gases	Trend artefacts reduced to ~0.005 K yr^{-1}
	~0.02 K yr^{-1} (ATSR-2, AATSR)	Bring sensors into consistency using overlap periods Final step: tie SST to selected, stable in situ measurements	
Persistent regionally varying retrieval biases	−0.2 to +0.25 K	Use new non-linear retrieval method that reduces these errors Account for spatial variations of trace gases (climatologically)	Reduced to ±0.05 K
Localized biases (high latitude and Saharan dust region)	~0.1 K on average	Improved realism of simulations, including extreme polar atmospheres and mineral aerosol. Better flagging of severe dust events	Achievable reduction of bias is uncertain at present
Occasional biases from erroneous cloud screening	One-off errors exceeding 1 K in about 1% of 1° × 1°-averaged AATSR SSTs	Post-hoc adjustments to operational screening and use of new probabilistic screening	Rate of occurrence reduced to <0.1%

"Envisat format", being created at the NERC Earth Observation Data Centre (NEODC). The input data volume is 60 TB and the processing chain, after quality checking and acquisition of auxiliary data, will involve:

1. post-hoc revisions to operational cloud screening,
2. new probabilistic cloud screening (involving fast forward modelling of brightness temperatures & reflectance, and estimation of clear-sky probability for each pixel; see Section 4.2),
3. 1 km pixel-resolution SST retrieval (linear estimate & non-linear correction; see Section 4.1), using pixels assessed as clear-sky by each cloud detection method,
4. averaging of pixel-resolution SSTs to 0.1° latitude–longitude cells,
5. full SST error modelling.

As explained in Section 3.2, each of these steps is required to be done using a full and restricted set of channels (observed bands of wavelength). The main production phase is estimated to require about one year of processing on 10 compute-servers, assuming a reasonable load factor.

3.2. Multiple processing streams

In any such exercise there are choices about how data are treated that may influence the final outcome (in this case the SST record) to a degree that is difficult to quantify a priori. Two fundamental choices face us in (A)RC: the method of cloud detection and the choice of channels.

The operational cloud detection scheme is based on the work of Zavody et al. (2000), and is well developed. It is a threshold-based scheme in which various tests for cloud are applied to channel brightness temperatures, brightness temperature differences, local standard deviation and larger-scale measures of coherence. An alternative scheme has recently been reported (Merchant et al., 2005) in which the probability that a pixel is clear-sky is estimated much as any other parameter might be retrieved (see also Section 4.2). Each scheme has a comparable efficiency, with particular strengths and weaknesses; the potential impact on the SST record of choosing one over the other is as yet unclear.

The availability of channels varies through the period, as summarized in Table 2. Given the requirement for homogeneity of the new SST record, we could decide to use only the channels that are available throughout the time-series (11, 12 and 1.6 μm) – the "minimal channel set". However, the best result for any given observation usually comes from using all the information available – the "maximal channel set". For example, dual-view three-channel retrievals of SST usually outperform dual-view two-channel and single-view SSTs, when applied to night-time data.

Since the consequences are unpredictable and since we wish to evaluate the magnitude of systematic uncertainty

Table 2
Channel availability for the (A)ATSRs

Channel/μm	12	11	3.7	1.6	.87	.65	.55
ATSR	Y	Y	Y[a,c]	Y	N	N	N
ATSR-2	Y	Y	Y[c]	Y	Y[b]	Y[b]	Y[b]
AATSR	Y	Y	Y[c]	Y	Y	Y	Y

[a] Channel failed less than 1 year after launch.
[b] Coverage in this channel limited by down-linking restrictions.
[c] Thermal channel not normally used for SST in day-time images.

(or at least, systematic influence) from such decisions, we will process four streams to SST: using the minimal and maximal channel sets and both cloud detection schemes. The spread of differences between the SST records is informative about the sensitivity of the outcome to such choices. One stream will be identified as preferred and will go forward to full validation, assessment and distribution.

3.3. Homogeneity

Homogeneity is critical for any record used for climate research. The current operational and archived (A)ATSR SSTs are not homogeneous in their generation. During satellite operations, improvements are made to the processing, and problems are identified and solved. Sensor characteristics drift over time and there are always discontinuities between different sensors in a series. Moreover, knowledge of atmospheric radiative transfer has increased during the (A)ATSR missions. We will create the first truly homogeneous (A)ATSR SST record by reprocessing with consistent cloud detection and new estimators (i.e., "retrieval coefficients") for SST. These new estimators are being developed for all sensors using fully consistent spectroscopy and line-by-line radiative transfer modelling for the transmittance of atmospheric gases. These simulations will take into account the changing concentrations of atmospheric trace gases over the period (both trends and significant seasonal-regional variability). Simulations of the time-dependent radiative effects of marine and stratospheric aerosols are being performed (with multi-stream simulation of scattering effects).

The new record will also benefit from greater completeness of coverage due to the consolidation of the orbital passes undertaken as part of the creation of the uniform archive.

3.4. Stability

The target stability of the (A)RC SSTs in the absence of real changes is 0.05 K decade^{-1}. For ATSR there was a drift of 0.7 K decade^{-1} in the latter part of ATSR's life (relative to the Tropical Atmosphere Ocean array moorings, identified in unpublished work by Merchant). The reason for this relatively large drift is as follows. For operational reasons, a rise in temperature of the actively cooled detectors was allowed. This rise is known to move the edge of the response of the 12 μm channel detector into the filter pass band of the channel. This modifies the spectral response of the detectors and requires modifications to the calibration curves used to convert detector counts to brightness temperature. These effects were not corrected operationally. A partial correction for the change in spectral response uses detector-temperature dependent coefficients, and this has been shown to reduce the drift of ATSR SSTs relative to SSTs from moorings. Further work on this drift will be undertaken within (A)RC to correct it as fully as possible. O'Carroll et al. (2006b) compare ATSR-2 SSTs with matched buoy SSTs obtained at the Met Office and find that the drift over 1995–2000 was +0.2 K decade^{-1} in night-time retrievals.

Thus, an improvement in stability of the observing systems by a factor of four and fourteen is required to meet our target for ATSR-2 and ATSR respectively. In addition to the work on the ATSR detector temperature trend, this improvement will partly be achieved by exploiting overlap periods between sensors. We will analyze overlap periods to identify discrepancies between consecutive sensors, and, if required and as appropriate, devise drift corrections bringing the sensors into consistency, using matched observations of ATSR with ATSR-2 and of ATSR-2 with AATSR.

Progress thereafter is unlikely to be achieved without reference to in situ SSTs. As a final step toward minimizing trend artefacts in the (A)RC SST, therefore, we may adjust the satellite SSTs with reference to selected high-quality in situ SST data. However, this must be implemented with caution. We will assess the in situ record post-1991 to identify a subset with temporal stability better than our estimate of the temporal stability of the (A)RC SST. It is not guaranteed that a subset with adequate stability exists, and if it does it is likely to cover a geographically limited region. The evolution of difference between the (A)RC bulk SST and the in situ subset may identify residual jumps or drifts in the new record: whether this can be done to 0.05 K decade^{-1} will be assessed.

Because the in situ subset of data is likely to come from geographically limited areas of ocean, the question arises of whether we can be confident that temporal adjustments derived to improve the (A)RC SST stability relative to the in situ subset are valid globally. It is difficult to prove this in general, although the relationship of biases in different geographical areas can be explored by radiative transfer modelling where candidate mechanisms for the residual biases can be identified. Moreover, extra insight and confidence can be gained by considering inter-algorithm differences in SST globally, an example of which follows.

The main algorithms for SST estimation are referred to as "dual-view three-channel" (or "dual-3", which uses the 3.7, 11 and 12 μm observations in both forward and nadir view, i.e., at zenith angles of approximately 55° and 0°, respectively) and "dual-view two-channel" (or "dual-2", using both views, and only 11 and 12 μm). Both dual-3 and dual-2 are used: the dual-3 is not applicable to day time images where solar irradiance may modify the brightness temperature in the 3.7 μm channel significantly, but is preferred over dual-2 for night-time images because of its lower systematic and random errors.

Ideally, the dual-3 and dual-2 estimators should give similar results when applied to night time data. In practice there are discrepancies of order tenths of kelvin that are progressively becoming better understood (see Section 4.1). We aim to reduce these discrepancies (i.e., the relative systematic errors between dual-3 and dual-2) to <0.1 K. If this is achieved, it gives some confidence (in conjunction with validation against in situ measurements) that *absolute* systematic errors in dual-3 and dual-2 SSTs have also been reduced to <0.1 K. We may assert this on the basis of this self-consistency check for the following reason. Dual-3 and dual-2 SSTs do use channels in common (11 and 12 μm), but in fact have a reasonable level of independence, the dual-3 SST retrieval being dominated by the 3.7 μm observations. Since all channels tend to be differently sensitive to potential causes of bias, the degree of self-consistency between dual-3 and dual-2 SSTs is indicative of the degree of validity of the radiative transfer simulations that underpin the retrievals – there is low probability of precise agreement (<0.1 K) occurring in all regions by chance. Where in situ observations of sufficient accuracy are not available, checking for self-consistency between algorithms is the most informative approach available. In future, this approach will be strengthened by comparing dual-2 SSTs with SSTs based solely on a dual-view retrieval using the independent 3.7 μm channel.

In Section 4.1, progressive improvements in dual-3/dual-2 consistency in recent years are described which demonstrate that 0.1 K is now a plausible target. In this context of low, consistent bias globally, applying a small temporal correction for stability with reference to a limited subset of high quality in situ observations is a justifiable approach.

3.5. Validation and verification of the (A)RC SST record

Systematic errors associated with changing channel availability and alternative cloud-screening schemes will be evaluated from the differences in SSTs between alternative streams. Changes in these differences over time are informative about discontinuities and trend artefacts within each mission. These terms will be added to the retrieval error model.

The Met Office operational system for validating AATSR bulk SSTs against quality-controlled global drifting buoys will be applied to the new SST records. Although somewhat limited by the accuracy of the in situ data collected in this system, the data are numerous and are useful for comparing different SSTs, and examining the distributions of biases in time and space and as a function of wind speed, insolation, etc.

Validation of skin SST by co-incident radiometric SST measurement from ships has been performed for all (A)ATSRs, and in recent years more attention has been paid to inter-calibration of in situ measurements (e.g., Barton et al., 2004). These matches will be re-visited with the new skin SSTs (both 0.1°-averaged skin SSTs and full-resolution images). Given our improvements to skin SST retrieval in reducing regional systematic errors, the interpretation of the validation results in terms of absolute accuracy of skin SSTs should be clearer than previously.

Lastly, various analyses will be undertaken to assess the (A)RC SSTs compared to the Hadley Centre analysis of in situ SSTs. The increase in spatio-temporal coverage relative to in situ data will be characterized, and monthly mean SST differences from existing climate SST fields will be assessed at 5° spatial resolution. Other parameters describing the two SST records will be calculated and compared: for example, the anomaly time-series, variance, local linear trends, zonal means, auto-correlations and 5-year mean anomaly fields.

4. Review of recent developments

4.1. Improvements in systematic errors towards <0.1 K target

Current (A)ATSR 1 km SSTs have scatter of 0.2–0.5 K (depending on region and algorithm) with regional systematic biases (e.g., Merchant et al., 2006) up to ~0.3 K (excluding biases from cloud-screening failures and severe tropospheric aerosol events). We aim to reduce systematic errors to <0.1 K for all regions, sensors and algorithms, and in this sub-section we review the progress already made towards this target, and the areas where problems remain. We focus here on the global scale, using zonal dual-2/dual-3 difference as an indirect measure of bias, as discussed above. For results of validation of (A)ATSRs against in situ data see, for example, Corlett et al. (2006); Noyes et al. (2006) and O'Carroll et al. (2006a).

Fig. 1 presents the past and anticipated progress in reducing systematic error for (A)ATSRs, showing the feasibility of this ambitious target. This figure is based on six months (December 2002 to May 2003) of routinely distributed AATSR "Meteo" product, filtered to retain night-time data only, accumulated and averaged in 5° bands of latitude, to create the thin solid line labelled "Current". This line is indicative of the systematic error in operational AATSR SSTs, and shows that there is an overall offset with dual-2 SSTs cooler than dual-3 SSTs by ~0.2 K. The other lines in the figure are calculated by perturbation of the "Current" line, as follows. The retrieval equations are all of form

$$\hat{x} = a_0 + \mathbf{a}^{\mathrm{T}}\mathbf{y}$$

where a_0 is the offset coefficient, and $\mathbf{a}^{\mathrm{T}} = [a_1, \ldots, a_n]$ is a vector of n weighting coefficients that each multiply one of the n brightness temperatures (BTs) in the observation vector $\mathbf{y}$ (The coefficients vary weakly with across-track distance to account for scan-angle changes). The different elements in $\mathbf{y}$ correspond to observations in the different AATSR channels and two view angles. $\mathbf{a}^{\mathrm{T}}\mathbf{y}$ is the inner

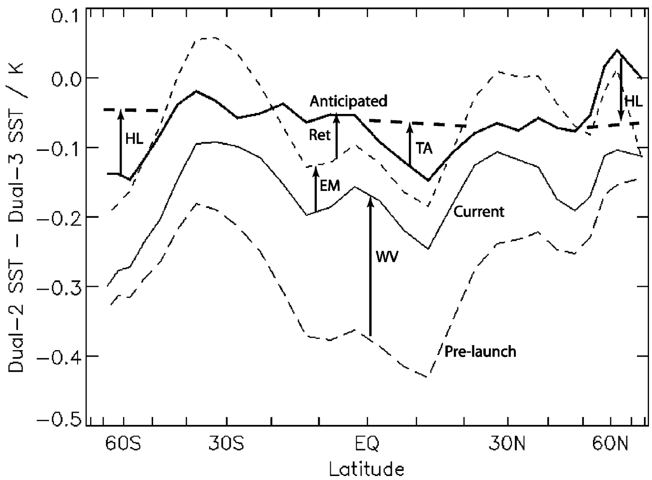

Fig. 1. Zonal difference (relative bias) between different types of SST retrieval, as an indicator of the magnitude of absolute SST bias. The plot presents differences between SSTs retrieved using "dual-2" and "dual-3" coefficients when applied to 6 months of AATSR night-time brightness temperatures. This *relative* bias is a useful measure of the magnitude and geographical variation of *absolute* bias. Pre-launch curve: Light dashed line: SSTs based on pre-1996 spectroscopy suffered a water-vapour related bias, reduced (see improvement labelled "WV") by new characterisation of the water vapour continuum absorption. For ATSR, previous work also addressed ~1 K biases from the Pinatubo aerosol (not shown). Current: Thin solid line: results from current AATSR operational coefficients, showing relative bias ~0.2 K in the global mean, and zonal structure ~0.1 K in amplitude. Anticipated: Heavy solid line: this curve shows the estimated improvements from using a new model of sea surface emissivity (improvement "EM") and correction of prior and non-linearity errors (improvement "Ret"). Two biases then remain: tropospheric aerosol, principally Saharan dust, differentially affects dual-2 and dual-3 SSTs around 10°N; and there is bias of uncertain origin in high latitudes. Assuming methods to correct these biases ("TA" and "HL") are forthcoming, global and zonal biases will be reduced to <0.1 K.

product of the weighting and observation vector, equivalent to the summation $\sum_1^n a_i y_i$. If the coefficients are applied to AATSR observations that are consistent with the simulated AATSR BTs used to define the coefficients, there is no global bias (although, as we will see below, a local variation in bias of known origin remains). The simulations used to define coefficients may be subject to forward model error that can be expressed:

$$\varepsilon_{\mathbf{y}} = \varepsilon_F + \frac{\partial F}{\partial \mathbf{x}}\varepsilon_{\mathbf{x}} + \frac{\partial F}{\partial \mathbf{b}}\varepsilon_{\mathbf{b}}$$

where the terms on the right hand side represent (from left to right): error in the radiative transfer model itself (approximations and numerical errors); departures from realistic specification of the atmospheric profiles used for simulations; and error in forward model parameters, such as emissivity or absorption by radiatively active gases. F is the forward model, **b** contains parameters embedded in the forward model, **x** represents the atmospheric state, and ε_F, ε_x and ε_b are systematic errors in F, **x** and **b**, respectively. The contribution to error from forward model parameters is isolated by using identical radiative transfer runs except for different values of parameters of interest. If we define $\mathbf{y}_p$ as the brightness temperatures simulated after perturbing a parameter, b, of the forward model by Δb, then

$$\mathbf{e}_b = \mathbf{y} - \mathbf{y}_p \cong \frac{\partial F}{\partial b}\Delta b$$

quantifies the error in brightness temperature from a parameter error of size Δb. Substituting this in the retrieval equation shows that the associated SST error, $\mathbf{e}_{\mathrm{SST}}$ is then:

$$\mathbf{e}_{\mathrm{SST}} = \mathbf{a}^{\mathrm{T}}\mathbf{e}_b$$

This is similar to the method in Merchant et al. (1999) for assessing the impact of stratospheric aerosols on SST retrievals.

The first perturbation to the "Current" line in Fig. 1 illustrates the effect that improvements in forward model parameters in the spectroscopy of water vapour brought to (A)ATSR SSTs (Merchant and Harris, 1999). SSTs based on pre-1996 spectroscopy showed a water-vapour related bias that was reduced (see transition labelled "WV" in Fig. 1) when the "CKD 2.2" (Han et al., 1997) and subsequent formulations were adopted.

The next improvement in AATSR biases is expected to come from revising the forward model parameters defining the thermal emissivity of the ocean surface. Until now, the model for emissivity of the ocean surface has depended on the satellite zenith angle and wind speed (or surface roughness, Watts et al., 1996). Recent work (Newman et al., 2005) has identified that significant errors in simulated brightness temperatures may arise from neglecting the temperature dependence of emissivity in simulations. Using simulations based on the different emissivity parameterisations, we have estimated zonally-averaged $\mathbf{e}_{\mathrm{SST}}$ and derived the expected bias in SST. By subtracting this bias (transition "EM" in Fig. 1) we derive a new curve for the zonal error (short dashed line) that has a reduced global mean difference (<0.1 K), while showing a greater amplitude of zonal variation.

These zonal variations are not mainly a result of further forward modelling errors, but are principally explained by prior and non-linearity error intrinsic to the linear retrieval method (Merchant et al., 2006). There are various means by which these systematic errors can be greatly reduced which will be evaluated. That they can largely be eliminated is clear, given that they can be successfully reproduced in simulation. The zonal structure of these systematic errors has been subtracted (transition "Ret") to create an estimate of the zonal bias curve achievable on current understanding (thick solid line, labelled "Anticipated"). Between 50°S and the equator and between 20°N and 50°N, this curve is within ±0.02 K of −0.05 K and therefore within our target. There are deviations from −0.05 K of up to ±0.1 K, however, at latitudes poleward of 50°N and 50°S, and between 0°N and 20°N. The origin(s) of the former "high latitude" biases is (are) presently unclear; the neglect of seasonal variations in trace gas concentrations at high latitudes may contribute. The north equatorial bias is at least partly due to Saharan dust (aerosol) causing differential biases in the dual-2 SSTs and dual-3 SSTs (Noyes et al., 2006). In general, the (A)ATSRs have been shown to be relatively tolerant of tropospheric aerosols (Vazquez-Cuervo et al., 2004) , with Saharan dust being the principal concern. It is not known at present how to solve either the high latitude or Saharan dust biases. There is effort within (A)RC to seek new solutions. In the case of Saharan dust, the recent development of a Saharan Dust Index (SDI, Merchant et al., 2006) for Meteosat Second Generation satellites will allow matched SDI-AATSR observations to be exploited to characterize the effects of dust on AATSR brightness temperatures.

4.2. Alternative cloud screening

Merchant et al. (2005) have compared ATSR-2 operational cloud screening with a new method that exploits prior fields of SST and atmospheric state combined with forward modelling to estimate the probability that each image pixel is cloud-free. Pixels for which this probability is low can then be excluded from the SST retrieval step. This was done for night-time imagery, and showed, in most cases, a modest improvement over the operation cloud masks in terms of false alarm rate (flagging where cloud is absent) and hit rate (flagging where cloud is present). However, in 7% of images, problems of widespread false flagging were evident in the operational mask, and could be associated with errors exceeding 1 K in SST in 1° latitude–longitude cells.

Further work has extended the new cloud-screening method to AATSR and to day-time imagery, and here we give an example highlighting situations where the operational AATSR mask produces high false alarm rates. The example is of AATSR imagery collected over Korea on the 18 October 2003, 01:17 UTC. The image (Fig. 2a) shows cloud over much of the lower third of the image and cloudy areas in the upper right of the image. The retrieved surface temperature (Fig. 2b) shows strong oceanic fronts formed throughout this region, a tricky context for cloud screening involving any image-textural analysis. The operational cloud-mask (Fig. 2c) produces the highest false alarm rate in the central section of the image where the strongest ocean fronts occur. There are also areas throughout the image where false alarms occur over 32×32 pixel blocks (especially on fronts in the top third of the image). The alternative cloud-mask (Fig. 2d), defined using a threshold of 50% on the estimated probability of a pixel being cloud-free, avoids such problems over the full region.

The high false alarm rates in the operational mask in this instance are due to the 1.6 and 11–12 μm histogram

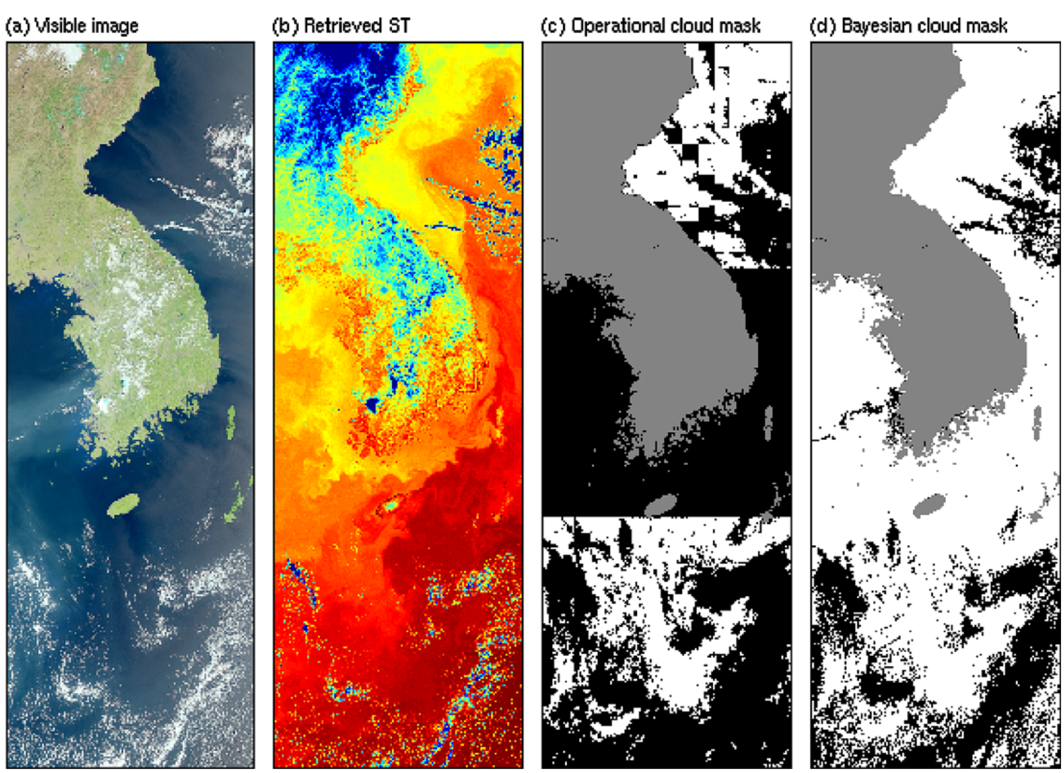

Fig. 2. Comparison of the operational cloud mask with the alternative probabilistic cloud mask. (a) False-colour image formed from the 0.67, 0.87 and 1.6 μm reflectance channels. A logarithmic scaling has been used to highlight the aerosols. (b) False colour retrieved surface temperature scaled to highlight the oceanic fronts. Scale limits are 270–300 K, blue is cold, red is warm. (c) Operational cloud mask formed from a series of threshold tests. Black = cloud, white = clear, grey = land. The central 512×512 section has been completely masked resulting in a very high false alarm rate for this image. (d) Cloud mask formed by setting a 50% threshold on the calculated probability of a pixel being clear. The clear sky probability was calculated using atmospheric fields from the ECMWF operational forecast model, with the prior BTs modelled using the radiative transfer model, RTTOV-8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tests (see Zavody et al., 2000 for a description of these tests). The 1.6 μm histogram test assumes that clouds scatter 1.6 μm radiation strongly, whereas the reflectivity of the sea surface is very low away from sun-glint regions. Due to the spatial variability in aerosol loading a dynamic threshold is estimated based on histograms produced from 32×32 blocks of pixels. Comparing the 1.6 μm histogram mask (Fig. 3a) with the 1.6 μm reflectance values (Fig. 3b) it is apparent that the false alarms from this test are associated with aerosol scattering, producing the blocky features in the operational mask. The test is, in a sense, behaving correctly, but the masking is unnecessary in this case since these aerosols do not significantly bias the SST retrieval.

The 11–12 μm histogram test (Fig. 3c) has effectively masked all of the remaining pixels in the central 512×512 section of this image. The aim of the test is to capture low stratus cloud that has escaped detection by the other tests. The test relies on the fact that the 11–12 μm difference is lower for cloudy pixels than pixels over clear sea. The histogram in this case is formed from all of the remaining clear pixels in the current 512×512 section of imagery being processed. Our independent re-creation of this histogram-based test (from the 11–12 μm difference data shown in Fig. 3d) did not flag this section of the image.

The above is an example of the need for the re-analysis to include revisions to the operational cloud mask. This has been confirmed by further statistical analysis within the (A)RC project of ATSR-2 operational cloud masks recomputed within the new unified archive at NEODC (Merchant et al., 2005 , worked from a previous release of ATSR-2 imagery). A set of 225 ATSR-2 scenes have been screened by an expert without reference to the operational cloud mask. Compared to the expert masks, the operational screening has a high hit rate (95% of pixels judged cloudy by the expert flagged by operational screening), with 54% of the image segments tested giving hit rates >95%. On the other hand, the false alarm rate (clear pixels flagged as cloudy) was also high at 31%. Since the swath of the (A)ATSRs is relatively narrow, losing clear-sky ocean pixels to false alarms at this rate is particularly undesirable. Major failures in the operational cloud mask were again observed in 7% of the imagery, these mostly being false alarms over large areas. This may partly arise because of a conservative approach to setting cloud screening thresholds which has, on the other hand, yielded (A)ATSR SST products that validate extremely well; nonetheless, it would be beneficial to reduce the rate of false alarms. A breakdown of the false alarm pixels showed that 49% were flagged by the 11 μm spatial coherence test, 49% were flagged by the 11/12 μm dual view difference test and

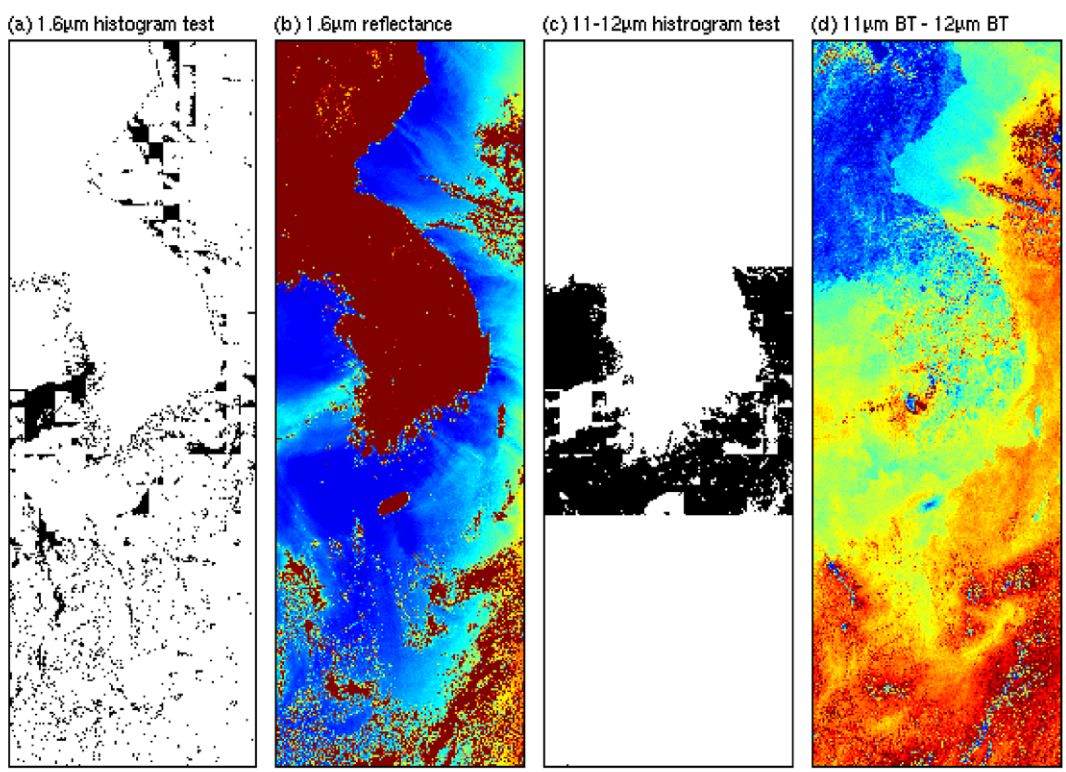

Fig. 3. The two histogram test that produced the highest false alarm rates in the operational cloud mask. (a) Pixels flagged as cloudy (black) by the 1.6 μm histogram test. This test operates on 32×32 pixel grids. (b) 1.6 μm reflectance scaled (0–5%) to highlight the aerosols, blues = low and red = high reflectance. Comparing (a) and (b) boxy flagging in the cloud mask corresponds to the variability in the reflectance due to the aerosols. (c) Pixels flagged as cloudy by the 11 μm BT – 12 μm BT difference histogram test. The combined flagging from (a) and (c) accounts for 80% of the flagging in the central 512×512 section of the image. (d) The 11 μm BT – 12 μm BT difference (scaled on 0–2.5 K). The BT differences due to the oceanic fronts range from 0.8 to 2.3 K over the central 512×512 section of the image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

35% by the 3.7/11 μm dual view difference test. (This sums to >100% because some pixels are "falsely" flagged by more than one test. See Zavody et al., 2000 , for descriptions of these tests.) The flagging of clear pixels by dual-view tests is expected because clouds obscure different surface locations in the nadir and forward views. Excluding dual-view tests, the main origin of false alarms is from the 11 μm spatial coherence test. However, effort is dedicated within the (A)RC project to explore post-hoc adjustments to the 11 μm spatial coherence test to reduce the false alarm rate and maximize the return of valid SSTs.

Both the new probabilistic and the revised operational cloud screening will be implemented within (A)RC. Each scheme has a comparable efficiency, with particular strengths and weaknesses. The impact on the SST record of choosing one method over the other will be characterised, giving insight about the magnitude of systematic influence on the final SST record from this choice (see Section 3.2).

5. Conclusion

The (A)RC project will deliver a unique record of sea surface temperature during the 1990s and early years of the 21st century – a period of relatively rapid climate change and of increasing societal concern about future climate. (A)RC SSTs will build on the success of the series of (A)ATSRs as stable and accurate sensors observing Earth's thermal radiation in wavelengths that allow determination of surface temperature. Existing (A)ATSR SSTs already have a deserved reputation for stability and precision; the (A)RC project has ambitious objectives of building on this heritage to create a record that will significantly advance climate research.

The objectives are ambitious, but there is reason to consider them feasible. In Section 3 of this article, we have summarized the approach being taken within (A)RC to obtain homogeneity and stability of the SST record and a comprehensive understanding of errors. In Section 4, we have reported recent progress towards reducing SST retrieval errors further and towards having available an alternative approach to the important step of cloud screening.

We have described in Sections 1 and 2 the motivation for (A)RC and the requirements for (A)RC SSTs to contribute significantly to climate research. This context is, of course, crucial, since the ultimate criterion for success of the project is the degree to which it meets the need for more robust quantification of global and regional sea surface temperature over recent decades.

Acknowledgements

(A)RC is funded jointly by the Natural Environment Research Council, the Ministry of Defence, and the Department of Environment, Food and Rural Affairs. This support is gratefully acknowledged.

References

Barton, I.J., Minnett, P.J., Maillet, K.A., Donlon, C.J., Hook, J.S., Jessup, A.T., Nightingale, T.J. The Miami2001 infrared radiometer calibration and intercomparison. Part II: Shipboard results. J. Atmos. Ocean. Technol. 21 (2), 268–283, 2004.

Collins, M., Tett, S.F.B., Cooper, C. The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. Clim. Dyn. 17 (1), 61–81, 2001.

Corlett, G.K., Barton, I.J., Donlon, C.J., Edwards, M.C., Good, S.A., Horrocks, L.A., Llewellyn-Jones, D.T., Merchant, C.J., Minnett, P.J., Nightingale, T.J., Noyes, E.J., O'Carroll, A.G., Remedios, J.J., Robinson, I.S., Saunders, R.W., Watts, J.G. The accuracy of SST retrievals from AATSR: an initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. Adv. Space Res. 37 (4), 764–769, 2006.

Donlon, C., Robinson, I., Casey, K.S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., Barton, I., Beggs, H., Poulter, D.J.S., Merchant, C.J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R.W., Kawamura, H., Rayner, N. The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature pilot project. Bull. Am. Met. Soc. 88 (8), 1197–1213, 2007.

Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J.B. Bulk parameterization on air–sea fluxes: Updates and verification for the COARE algorithm. J. Climate. 16, 571–591, 2003.

Han, Y., Shaw, J.A., Churnside, J.H., Brown, P.D., Clough, S.A. Infrared spectral radiance measurements in the tropical Pacific atmosphere. J. Geophys. Res. 102, 4353–4356, 1997.

Kaplan, A., Cane, M.A., Kushnir, Y., Clement, A.C., Blumenthal, M.B., Rajagopalan, B. Analyses of global sea surface temperature 1850–1991. J. Geophys. Res. 103, 18567–18589, 1998.

Kent, E.C., Kaplan, A. Toward estimating climatic trends in SST. Part III: Systematic biases. J. Atm. Ocean. Tech. 23 (3), 487–500, doi:10.1175/JTECH1845.1, 2006.

Kent, E.C., Taylor, P.K. Toward estimating climatic trends in SST. Part I: Methods of measurement. J. Atm. Ocean. Tech. 23 (3), 464–475, doi:10.1175/JTECH1843.1, 2006.

Kilpatrick, K.A., Podesta, G.P., Evans, R. Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. J. Geophys. Res. 106 (C5), 9179–9197, 2001.

Mason, I.M., Sheather, P.H., Bowles, J.A., Davies, G. Black-body calibration sources of high accuracy for a spaceborne infrared instrument: the Along Track Scanning Radiometer. App. Optics. 35, 629–639, 1996.

Merchant, C.J., Embury, O., Le Borgne, P., Bellec, B. Saharan dust in night-time thermal imagery: detection and reduction of related biases in retrieved sea surface temperature. Rem. Sens. Env. 104, 15–30, 2006.

Merchant, C.J., Harris, A.R., Maturi, E., MacCallum, S. Probabilistic physically-based cloud screening of satellite infra-red imagery for operational sea surface temperature retrieval. Q. J. R. Meteorol. Soc. 131, 2735–2755, 2005.

Merchant, C.J., Harris, A.R., Murray, M.J., Zavody, A.M. Toward the elimination of bias in satellite retrievals of skin sea surface temperature 1. Theory, modeling and inter-algorithm comparison. J. Geophys. Res. 104 (C10), 23565–23578, 1999.

Merchant, C.J., Harris, A.R. Toward the elimination of bias in satellite retrievals of skin sea surface temperature 2. Comparison with in situ measurements. J. Geophys. Res. 104 (C10), 23579–23590, 1999.

Merchant, C.J., Horrocks, L.A., Eyre, J., O'Carroll, A.G. Retrievals of sea surface temperature from infra-red imagery: origin and form of systematic errors. Quart. J. Royal Met. Soc. 132, 1205–1223, 2006.

Merchant, C.J., Le Borgne, P. Retrieval of sea surface temperature from space based on modeling of infrared radiative transfer: capabilities and limitations. J. Atmos. Ocean. Technol. 22 (11), 1734–1746, doi:10.1175/JTECH1667.1, 2004.

Murray, J.M., Allen, M.R., Merchant, C.J., Harris, A.R., Donlon, C.J. Direct observations of skin-bulk SST variability. Geophys. Res. Lett. 27 (8), 1171–1174, 2000.

Newman, S.M., Smith, J.A., Glew, M.D., Rogers, S.M., Taylor, J.P. Temperature and salinity dependence of sea surface emissivity in the thermal infrared. Q. J. Roy. Meteorol. Soc. 131 (610), 2539–2557, 2005.

Noyes, E.J., Minnett, P.J., Remedios, J.J., Corlett, G.K., Good, S.A., Llewellyn-Jones, D.T. The accuracy of the AATSR sea surface temperatures in the Caribbean. Rem. Sens. Env. 101 (1), 38–51, 2006.

O'Carroll, A.G., Watts, J.G., Horrocks, L.A., Saunders, R.W., Rayner, N.A. Validation of the AATSR meteo product sea surface temperature. J. Atmos. Ocean. Technol. 23 (5), 711–726, 2006a.

O'Carroll, A.G., Saunders, R.W., Watts, J.G. The measurement of the sea surface temperature by satellites from 1991 to 2005. J. Atmos. Oceanic Technol. 23 (11), 1573–1582, 2006b.

Rayner, N.A., Parker, D.E., Horton, E.B., Folland, C.K., Alexander, L.V., Rowell, D.P., Kent, E.C., Kaplan, A. Global analyses of sea surface temperature, sea ice and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108 (D14). Art. No. 4407, 2003.

Rayner, N.A., Brohan, P., Parker, D.E., Folland, C.K., Kennedy, J.J., Vanicek, M., Ansell, T., Tett, S.F.B. Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the new HadSST2 data set. J. Clim. 19 (3), 446–469, doi:10.1175/JCLI3637.1, 2006.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W. An improved in situ and satellite SST analysis for climate. J. Clim. 15, 1609–1625, 2002.

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

Smith, D.L., Delderfield, J., Drummond, D., Edwards, T., Mutlow, C.T., Read, P.D., Toplis, G.M. Calibration of the AATSR instrument. Adv. Space Res. 28 (1), 31–39, 2001.

Smith, T.M., Reynolds, R.W. Improved extended reconstruction of SST (1854–1997). J. Clim. 17, 2466–2477, 2004.

Stramma, L., Cornillon, P., Weller, R.A., Price, J.F., Briscoe, M.G. Large diurnal sea-surface temperature variability – Satellite and In-situ measurements. J. Phys. Oceanogr. 16 (5), 827–837, 1986.

Vazquez-Cuervo, J., Armstrong, E.M., Harris, A. The effect of aerosols and clouds on the retrieval of infrared sea surface temperature. J. Clim. 17 (20), 3921–3933, 2004.

Watts, P.D., Allen, M.R., Nightingale, T.J. Wind speed effects on sea surface emission and reflection for the Along Track Scanning Radiometer. J. Atmos. Ocean. Technol. 13 (1), 126–141, 1996.

Závody, A.M., Mutlow, C.T., Llewellyn-Jones, D.T. A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. J. Geophys. Res. 100, 937–952, 1995.

Zavody, A.M., Mutlow, C.T., Llewellyn-Jones, D.T. Cloud clearing over the ocean in the processing of data from the along-track scanning radiometer (ATSR). J. Atmos. Oceanic. Technol. 17, 595–615, 2000.