# A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers

Christopher J. Merchant,[1] Owen Embury,[1] Nick A. Rayner,[2] David I. Berry,[3] Gary K. Corlett,[4] Katie Lean,[2] Karen L. Veal,[4] Elizabeth C. Kent,[3] David T. Llewellyn-Jones,[4] John J. Remedios,[4] and Roger Saunders[2]

[1] A new record of sea surface temperature (SST) for climate applications is described. This record provides independent corroboration of global variations estimated from SST measurements made in situ. Infrared imagery from Along-Track Scanning Radiometers (ATSRs) is used to create a 20 year time series of SST at 0.1° latitude-longitude resolution, in the ATSR Reprocessing for Climate (ARC) project. A very high degree of independence of in situ measurements is achieved via physics-based techniques. Skin SST and SST estimated for 20 cm depth are provided, with grid cell uncertainty estimates. Comparison with in situ data sets establishes that ARC SSTs generally have bias of order 0.1 K or smaller. The precision of the ARC SSTs is 0.14 K during 2003 to 2009, from three-way error analysis. Over the period 1994 to 2010, ARC SSTs are stable, with better than 95% confidence, to within 0.005 K yr$^{-1}$ (demonstrated for tropical regions). The data set appears useful for cleanly quantifying interannual variability in SST and major SST anomalies. The ARC SST global anomaly time series is compared to the in situ-based Hadley Centre SST data set version 3 (HadSST3). Within known uncertainties in bias adjustments applied to in situ measurements, the independent ARC record and HadSST3 present the same variations in global marine temperature since 1996. Since the in situ observing system evolved significantly in its mix of measurement platforms and techniques over this period, ARC SSTs provide an important corroboration that HadSST3 accurately represents recent variability and change in this essential climate variable.

**Citation:** Merchant, C. J., et al. (2012), A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers, *J. Geophys. Res.*, *117*, C12013, doi:10.1029/2012JC008400.

## 1. Introduction

[2] Sea surface temperature (SST) is a variable of central importance within climate science, meteorology and oceanography. This paper presents a new data set of SST derived from satellite observations intended primarily for climate applications. It is the outcome of the Along-track scanning radiometer Reprocessing for Climate (ARC) project, whose motivation and aims were described by *Merchant et al.* [2008].

[3] As explained in section 2.1, ARC SSTs have near-total independence from in situ SST measurements, high accuracy (∼0.1 K) and stability (<0.005 K yr$^{-1}$). These are the principal advances achieved with this data set relative to previous records of SST from satellites.

[4] Satellite and in situ SST need to be used together to quantify marine change over many decades. The ARC data set includes both the primary observation of skin SST, and an estimate of the SST at a depth of 20 cm ($SST_{0.2m}$) that is standardized with respect to the diurnal cycle and is more directly comparable to in situ measurements. We see this as another useful advance from the ARC project.

[5] Given the high stakes associated with concerns about anthropogenic climate change, it is important that climate data are trustworthy and shown to be robust. Agreement between two independent records for a given component of the climate is powerful evidence of the validity of both. This is a real issue, since the in situ-based historical data sets of marine climate are synthesized from networks of measurements that have evolved significantly over time, in terms of the mix of different measurements types and technologies [*Kent et al.*, 2010; *Kennedy et al.*, 2011b]. Few of the observing systems used in marine climate data sets were designed to have the stability necessary to observe climate change. In this situation, the risk of artifacts in the long-term changes is real, and a great deal of scientific effort has been invested in minimizing biases and estimating uncertainties from "known unknowns." But what about "unknown unknowns"? On this topic, *Immler et al.* [2010] rightly point out the importance of independent

[1]School of GeoSciences, University of Edinburgh, Edinburgh, UK.
[2]Met Office, Exeter, UK.
[3]National Oceanography Centre, Southampton, UK.
[4]Department of Physics and Astronomy, University of Leicester, Leicester, UK.

Corresponding author: C. Merchant, School of GeoSciences, University of Edinburgh, Crew Building, Kings Buildings, Edinburgh EH9 3JN, UK. (c.merchant@ed.ac.uk)

measurements, based on different measurement principles and accompanied by uncertainties. If two such data sets give the same picture of change, to within their known uncertainties, this provides a high degree of confidence that no major systematic effect has been neglected in either data set and that the uncertainty estimates are realistic. It is then highly unlikely that major "unknown unknowns" have significantly distorted our picture of geophysical change. Later in this paper we show that the ARC SST data set corroborates the variations in global SST given by the in situ-based HadSST3 data set [*Kennedy et al.*, 2011a, 2011b] during the 1990s and 2000s. Further, the satellite data set arguably gives a cleaner, more plausible rendering of specific SST anomaly events and interannual variability.

[6] The main conclusion of this paper is that the ARC SST record provides unequivocal independent evidence that in situ-based understanding of changes in SST over the past two decades is substantially correct. To support this conclusion, the paper progresses as follows. Characteristics of the ARC SST data set are described in section 2. The basis of ARC SST estimates is briefly surveyed. The provision of SST-depth estimates as well as the primary SST-skin retrievals is discussed. Section 3 presents summary validation of the ARC SST data set as a whole. Section 3 covers validation of ARC SST accuracy, precision and stability estimates. Section 4 presents ARC SST results. The climatological annual cycle of SST as observed in ARC is shown. Interannual variability in SST is shown to be cleanly captured in the ARC SST data set compared to an in situ-based assessment. Selected major SST anomalies are characterized using ARC SSTs. Last, the ARC SST time series of global SST anomaly is shown in comparison with HadSST3, and a map of linear trends presented. The paper concludes with a discussion in section 5.

## 2. Overview of ARC SST Data Set

### 2.1. Basic Characteristics

[7] The ARC SST record is based on the measurements of three Along Track Scanning Radiometers (ATSR) [*Smith et al.*, 2012]. Usable data from the first ATSR are available from 2 August 1991, and the last data were obtained on 8 April 2012, when the Envisat platform carrying the third sensor failed. The ARC SST data analyzed in this paper cover the period to the end of 2010.

[8] ATSR radiometers are dual-view sensors that observe the surface both near vertically (nadir) and at about 55° from the vertical (forward along the direction of travel). The underlying measurements comprise infrared and reflectance imagery obtained across a ~500 km swath at spatial resolution between roughly 1 km (nadir) and 3 km (forward). The three infrared channels are centered around wavelengths of approximately 3.7, 11 and 12 $\mu$m, and the reflectance channel in common to all sensors in the series is at 1.6 $\mu$m.

[9] The ARC SST product is available as daily daytime and nighttime averages of any SSTs observed within each 0.1° latitude-longitude cell of the global oceans. Because of the relatively narrow swath of the instruments and as a consequence of cloud cover, a single day's file is relatively sparsely populated with observations. For example, over a 3 day period (after which the orbit track approximately repeats itself) there will be nighttime SSTs available for typically 15% of the global ocean.

[10] Broadly, the channel availability, detector noise characteristics and instrumental stability improve with each instrument in the series, which is reflected in improving data quality in the ARC SST time series. In combination with cooled detectors giving low noise and calibration against two high-quality thermal targets, the extra information content for SST associated with the dual view capability permits very accurate and precise SST retrieval. Dual-view information can deliver (1) consistently small SST biases throughout the high-humidity tropics [*Embury et al.*, 2012b] and (2) a high degree of robustness to atmospheric aerosol [*Merchant and Harris*, 1999; *Merchant et al.*, 1999; *Good et al.*, 2012]. Both of these situations are challenging for single view sensors [*Merchant et al.*, 2009; *Walton*, 1985], especially for daytime observations where shorter-wavelength infrared channels (around 3.7 $\mu$m) are not available for SST retrieval because of solar "contamination" of the observed radiances.

[11] To exploit the information about SST that is available in principle in ATSR brightness temperatures, effective elimination of cloudy observations is required, followed by SST retrieval (using the measured radiances to estimate the SST). For ARC SSTs, both steps are based on simulating the physics of radiative transfer. ARC cloud detection [*Merchant et al.*, 2005] uses approximate radiative transfer informed by parameters appropriate to the time and place of the observation obtained from a numerical weather prediction (NWP) system. ARC SST retrieval relies on full line-by-line layer-by-layer radiative transfer calculations for a selection of representative conditions, used to define the relationship of brightness temperatures and SST for each sensor [*Embury et al.*, 2012a]. These relationships are encapsulated in retrieval coefficients [*Embury and Merchant*, 2012] that combine the measurements such that an SST estimate is obtained.

[12] In general, other satellite SST data sets are empirical, tied to in situ SST to a greater [*Kilpatrick et al.*, 2001] or lesser [*Merchant and Le Borgne*, 2004] degree. Since the ARC SST data set is based on the physics behind the observations, ARC SSTs are not tuned to in situ SST measurements. The importance of having an independent assessment of marine climate changes over recent decades has been described in the introduction.

[13] The degree of independence of ARC SSTs from in situ SSTs is very high, but not total. The NWP parameters used for cloud detection include an SST analysis obtained using in situ (and satellite) measurements, which introduces an indirect link between in situ measurements and the selection of the parts of an image for which SST is estimated. Later in this paper, comparisons will be made to the gridded SST data set HadSST3 [*Kennedy et al.*, 2011a, 2011b], which is based purely on in situ measurements. Some assessment of the random measurement and microbias uncertainties in HadSST3 was estimated with respect to the SSTs in a previous ATSR-based product (an operational version) [*Kennedy et al.*, 2011c]. However, neither of these components is used here, and the HadSST3 ensemble spread used to assess significance is completely independent of ARC SSTs.

[14] The recommended ARC SSTs for climate applications are dual-view two-channel (D2) and dual-view three-channel (D3) SSTs. For a discussion of available channel combinations of the ATSRs, see *Merchant et al.* [2008]. The
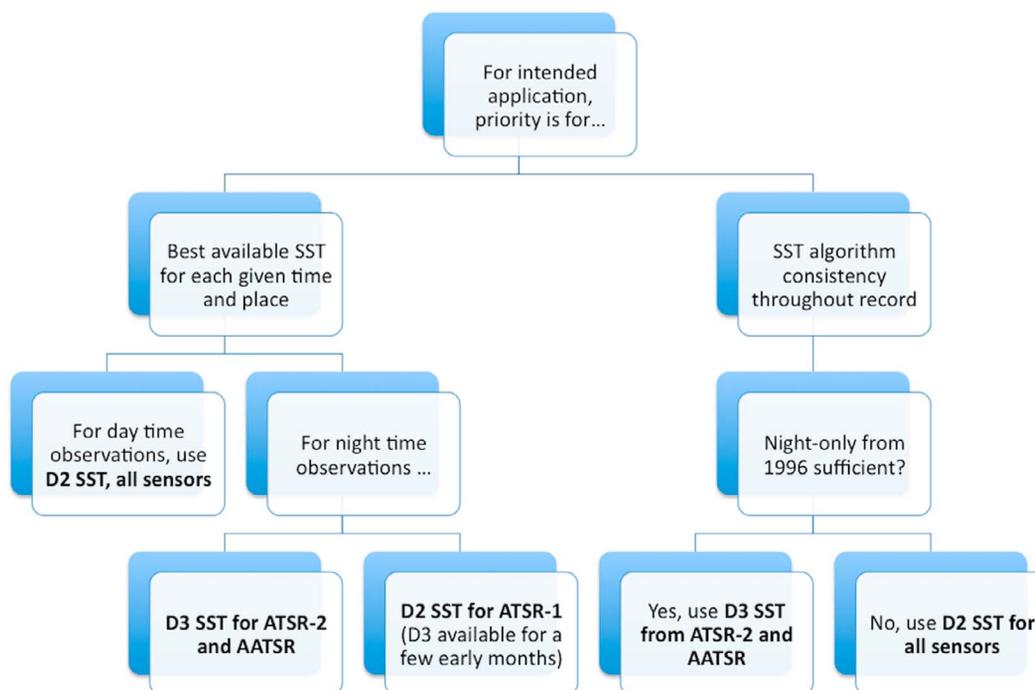
**Figure 1.** Decision tree for choice between SSTs retrieved using dual-view two-channel (D2) and dual-view three-channel (D3) algorithms.

D3 SSTs are only available at night (because the third channel, at 3.7 $\mu$m, is not used for SST for day-lit scenes) and are only available from ATSR-1 between 2 August 1991 and 27 May 1992 because of failure of the 3.7 $\mu$m channel. The D3 SSTs are low-noise, highly robust to atmospheric aerosol, and recommended for climate applications for which consistent nighttime data between July 1996 and March 2012 are sufficient. D2 SSTs are day-and-night SSTs, somewhat more noisy than D3 SSTs, and robust to many atmospheric aerosol conditions. D2 SSTs are recommended where the full time series from 1991 is required with one form of algorithm, or where day-night consistency is paramount. Nonetheless, D3 and D2 SSTs are highly mutually consistent, and we consider it is valid to use D3 when available and D2 otherwise. Comparisons of the D2 and D3 results are available from *Embury et al.* [2012b] in addition to this present paper. Figure 1 illustrates a decision sequence for users to determine the combinations of D2 and D3 most suitable for their application.

[15] A final point is that the periods of overlap between sensors have been used to make the data set as homogeneous as possible. It is not the SST records that are adjusted to create consistency, which would have various disadvantages. Instead, homogenization was done by referencing the brightness temperatures of all the sensors and channels to be consistent with the calibration of the 3.7 and 11 $\mu$m channels of the AATSR. (The reason for selecting this channel combination as reference is presented by *Embury and Merchant* [2012]. A manuscript on method of propagating this reference to the earlier sensors is in preparation.) The net effect, however, is that the SSTs from different sensors are very consistent during overlap periods, and can be used together.

## 2.2. Skin Sea Surface Temperature

[16] A further consequence of being based on physical simulations is that ARC SSTs deliver a true "skin SST" estimate, as explained in this section.

[17] A steep temperature gradient within the upper ∼1 mm [*Katsaros*, 1977] or less [*Hanafin and Minnett*, 2005] of the ocean arises when there is an air-sea flux of heat. The heat flux must be mediated through the surface "skin layer" only by molecular conduction, since turbulent transport is suppressed at the surface by the water-air density contrast. The temperature difference across the skin layer, referred to as the "skin effect," is typically of order −0.1 to −0.2 K [*Donlon et al.*, 2002] but can be as negative as −0.5 K [*Donlon et al.*, 1999]. Under certain conditions, such as when the nonsolar heat flux is, atypically, into the oceans, the difference can be positive. Satellite retrievals of SST are sensitive to the thermal (Planck) emission of the sea surface. At the infrared wavelengths most commonly used for SST remote sensing, the electromagnetic skin depth of seawater is of order 10 $\mu$m, and therefore the retrieved SSTs reflect the temperature near the top of the skin layer. They are therefore referred to as "skin SST."

[18] ARC SST products contain, as their primary observation, the skin SST at the time of observation. Skin SST is the most appropriate SST to use for several purposes. Instantaneous air-sea fluxes are mediated by the skin SST. Skin SST is appropriate to use for calculating the upwelling infrared irradiance (or "longwave flux") from the sea surface. Skin SST is closely coupled to the temperature and specific humidity on the air side of the air-sea interface, and it therefore (with the appropriate parameterization) can be used for calculation of sensible and latent heat fluxes. The skin SST has been stated to be the appropriate SST for

calculating waterside partial pressures of gases in air-sea gas exchange [e.g., *Liu et al.*, 1979; *Fairall et al.*, 2003]. Direct validation of skin SST at an accuracy of order 0.1 K is possible using shipborne radiometers [e.g., *Wimmer et al.*, 2012].

## 2.3. Depth Sea Surface Temperature

[19] In some other contexts, SSTs at shallow depths are more relevant. For the ARC project, with its focus on providing robust climate information, an important consideration is to relate satellite and in situ records of SST in order to make the link with the historical SST record. "Historical SSTs" include thermometer readings from seawater collected in buckets from ships (collected systematically since around 1850) or on engine intakes, through to the telemetered measurements from SST-capable drifting buoys that have become prevalent during the satellite era [e.g., *Kent et al.*, 2010]. The precise depths of such measurements are rarely certain, and have been estimated by *Kent et al.* [2007]. Seawater in buckets was probably drawn from within the upper meter of ocean. Drifting buoys typically have thermistors placed such as to sit 10 to 20 cm below the surface of a calm sea, but, at least while still attached to their drogues, drifting buoys may spend time submerged in well-mixed, heavy sea conditions. The approach in ARC has been to provide modeled adjustments to selected depths of 20 cm (used for comparison with drifting buoys) and 1 m (used for comparison with moored buoys of the global tropical moored buoy array, GTMBA). The adjustments derive from applying NWP fields to a model of the skin effect and of near-surface stratification, as described by *Embury et al.* [2012b]. The skin effect adjustment, based on a modification of the work of *Fairall et al.* [1996], is typically of order +0.2 K, and depends primarily on wind speed and insolation. The stratification between the subskin temperature and a depth of 0.2 or 1 m is usually small (<0.05 K). Significant stratification within the upper meter of ocean tends to occur by mid-morning only under conditions of sustained, extremely low wind (<3 m s$^{-1}$) and strong insolation, so only a small fraction of ARC daytime SSTs are associated with a significant subskin-to-depth adjustment. For ARC nighttime SSTs, the adjustment is generally negligible.

[20] However, as *Embury et al.* [2012b, Figure 7] show, there is a detectable diurnal cycle in the near surface, with SSTs warming at rates up to 0.1 K h$^{-1}$ around 10:00 LT and cooling down by −0.03 K h$^{-1}$ around 22:00 LT. The satellite overpass time for AATSR on Envisat was earlier than for ATSR-1 and ATSR-2 by half an hour (ascending node of 22:00 LT rather than 22:30 LT). For all platforms, the ascending node crossing time was stable, usually within seconds and always within a few minutes. This means that, without correction for this diurnal cycle, a step would be introduced into the ARC SST time series at the changeover between ATSR-2 and AATSR. We therefore use the skin-to-depth adjustments also to minimize this aliasing of the diurnal cycle into the long-term record. The skin-to-depth adjustments for AATSR adjust to half an hour later than the skin-SST observation time, as if Envisat had had the same ascending node time as the two earlier platforms. The SST-depth time series in ARC is, to our knowledge, the only satellite SST record where diurnal-cycle aliasing has been minimized in this way.

## 2.4. Practical Comments

[21] Users of the ARC data set will find it useful to be aware of the following considerations.

[22] Compared to other satellite SST data sets users may be used to (e.g., the Pathfinder SST data set from Advanced Very High Resolution Radiometer (AVHRR) [*Kilpatrick et al.*, 2001]), ATSR spatiotemporal coverage is relatively sparse. Because of the dual-view design, the swath width is ~500 km, or about a sixth of the swath of AVHRR usable for SSTs. For many periods, morning/evening and afternoon/night AVHRR SSTs are available, whereas ATSR SSTs are only morning/evening. ATSR should therefore be considered a complementary instrument to other sensors, offering high accuracy (shown below) at the expense of sparser sampling.

[23] ATSR-1, as mentioned earlier, lost a channel early in life. The lost channel was centered near 3.7 $\mu$m, was usable for nighttime SST retrieval, and helped increase robustness of SSTs to stratospheric aerosol. ATSR-1 displayed interference between the detectors and the detector cooler, which added additional noise to brightness temperatures with a characteristic signature in imagery [*Harris et al.*, 1995]. The radiometric noise worsened during ATSR-1's lifetime as the temperature of its actively cooled detectors was allowed to rise (to preserve mission lifetime). The calibration effects of the change in detector temperature have been accounted for in the ARC time series, yet the SSTs in the last 2 months of the main ATSR-1 mission (from May 1996) show unexpected drift in SST. These months are not recommended for climate applications, and are not included in the results presented in this paper.

[24] ATSR-2 operated for 6 months in parallel with ATSR-1, with ATSR-2 data available from June 1995. The two platforms flew 1 day apart with the same local equator crossing time. ATSR-2's operation was then interrupted on 22 December 1995 by trouble with the scan mechanism. It was restarted successfully in June 1996 and operated until overlap with AATSR in 2002. ATSR-1 full mission operations ceased 1 month before the ATSR-2 restart, and this gives a 3 month gap if the ATSR-1 data from the last 2 months are excluded. The navigation of the platform carrying ATSR-2 degraded due to a failure on 15 January 2001. There is a 6 week data gap followed by 4 months during which the accuracy geolocation and forward nadir colocation were degraded, with a corresponding impact on SST precision (see section 3.1). Improved control of the platform was achieved on 1 July 2001, and the SST quality seems recovered thereafter. Note that the SST uncertainty information in the ARC data set does not yet account for the period of less accurate satellite navigation.

[25] AATSR was the most stable sensor with the best noise performance. It operated extremely well until the Envisat platform failure in April 2012. There is uncertainty about the calibration of the 12 $\mu$m channel of AATSR, but the ARC SST coefficients are defined in such a way as to avoid bias from this calibration issue, as explained by *Embury and Merchant* [2012].

[26] Figure 2 is a time line that summarizes many of the events affecting the ATSR instruments and the context of SST retrieval. It is color-coded to indicate the relative quality of the available ARC SSTs for different periods.
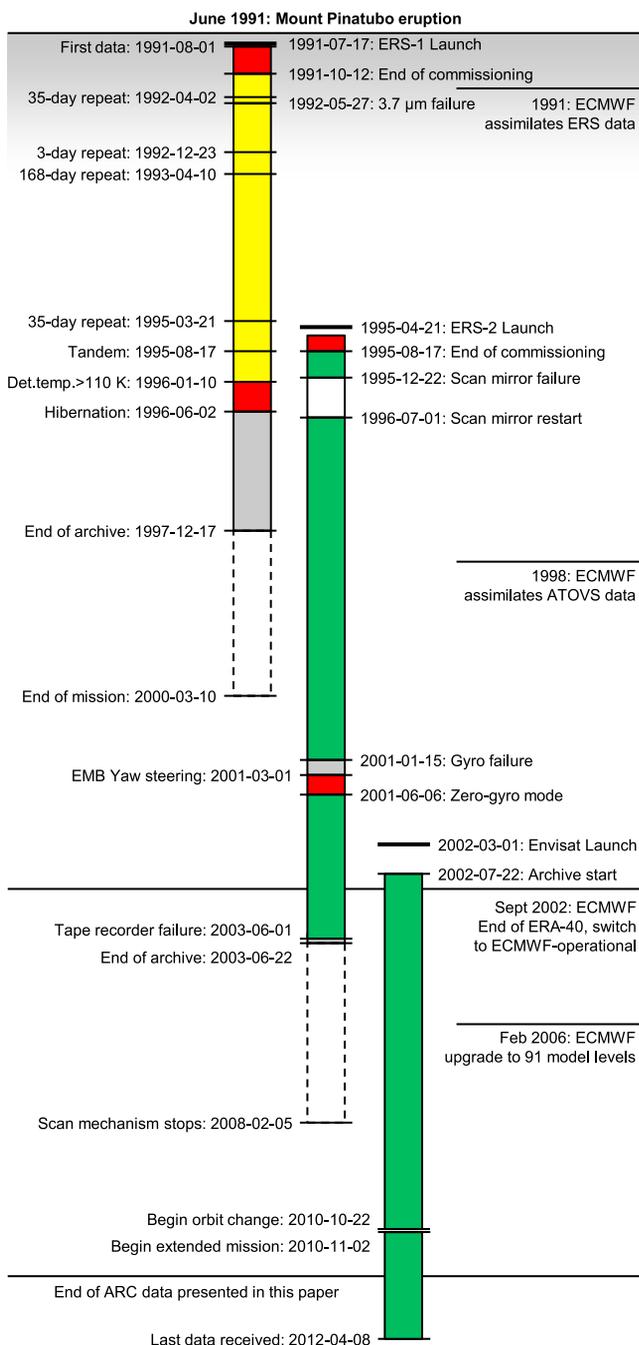
**Figure 2.** Timeline of events affecting ATSR missions and indication of relative ARC SST quality within the ARC data set. Green indicates best SSTs, generally no problems. Yellow indicates affected by instrument or other issues that increase SST noise and/or bias. Red indicates precommissioning data, or SST adversely affected by significant instrument problems. Gray indicates partial SST availability. White indicates no ARC SSTs (instrument data either interrupted or not in an accessible archive). ECMWF is the European Centre for Medium-range Weather Forecasting, whose fields are used within ARC cloud detection, and changes in these input fields are also noted on the right. ATOVS are atmospheric sounding radiances assimilated at ECMWF. "Det. temp." is detector temperature.

[27] There are minor data gaps (a few days) throughout all missions because of orbit maneuvers, platform degassing events, etc. Degassing is most prevalent in the early phase of Envisat (AATSR).

[28] Individual ARC SST records do not have uniform uncertainty. For this reason, each SST is accompanied by an uncertainty, which is an estimate of the standard deviation of the error distribution arising from random and correlated sources of error. This allows a user to select or weight SSTs according to this uncertainty information.

## 3. Summary Validation of ARC SST Data Set

### 3.1. Bias (Accuracy)

[29] One of the targets for the ARC project was that the SST accuracy should be 0.1 K, by which we mean that the mean error (bias) should be within $\pm 0.1$ K for all regions. The assessment of this target is based on comparison of the ARC SSTs with validation data, with ARC SST minus validation SST hereafter referred to as "discrepancy." Compared to the previous assessment of this target [*Embury et al.*, 2012b], the results here (1) use updated, more precise in situ locations extracted from the ICOADSv2.5 IMMA data set [*Woodruff et al.*, 2011], reducing errors associated with identifying spatial coincidence between satellite and in situ, (2) include new matches to Argo profiling floats for which the uppermost measurements at a mean depth of approximately 4 m have been extracted from the Met Office Hadley Centre EN3 data set [*Ingleby and Huddleston*, 2007] from 2004 onward, and (3) cover all three ATSR sensors used in the ARC project.

[30] Accuracy is assessed using the median discrepancy between ARC $SST_{0.2m}$ values and drifting buoy measurements across the whole time series, as shown in Figure 3. Most 5° latitude-longitude cells in Figure 3 are within the target range for D2 and D3 SSTs. Since the calibration accuracy of the drifting buoys to which the ARC SSTs are matched is thought to be of order 0.2 K [*O'Carroll et al.*, 2008], cells with fewer than ∼16 independent drifting buoys are not statistically reliable for determining a 0.1 K bias with high confidence (i.e., at the $2\sigma$ or 95% confidence level). Statistical uncertainty in the validation data account for some of the outlier cells that appear in common between plots.

[31] Global median bias statistics for D2 and D3 retrievals (again, after model-based adjustments of the skin-SST retrievals to $SST_{0.2m}$ or $SST_{1m}$) are given relative to drifting buoys, the Global Tropical Moored Buoy Array (GTMBA) [*McPhaden et al.*, 1998; *McPhaden et al.*, 2009; *Bourlès et al.*, 2008] and the shallowest measurement of Argo profiles, in Table 1. The methodology follows *Embury et al.* [2012b]. The shallowest measurement for most Argo profiles is between 3 and 5 m depth, so there is a 2 to 4 m difference in the nominal depths in this case. However, Argo SSTs are a useful complement to the drifting and the tropical moored buoy networks, being highly accurate and distributed across all latitudes.

[32] For AATSR and ATSR-2, the median discrepancy between ARC depth SSTs and drifting buoy or GTMBA SSTs is within $\pm 0.05$ K of zero, for all sensor combinations, type of algorithm and day or night. For AATSR, the comparison with the somewhat deeper Argo gives results that are relatively more positive during the day, with the D2 daytime
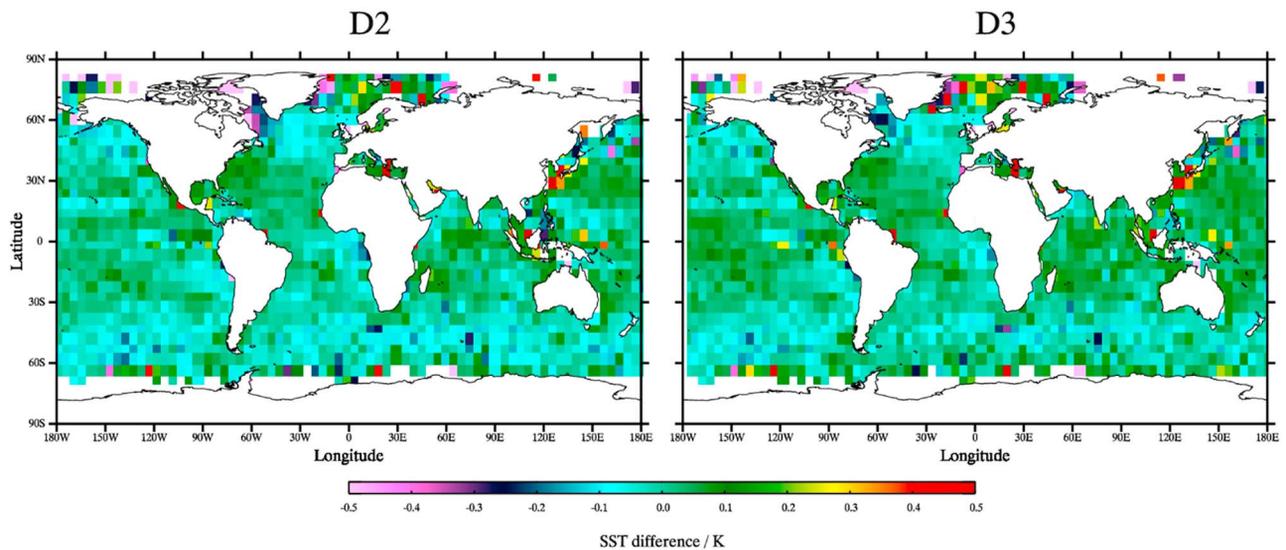
**Figure 3.** Median discrepancy of nighttime ARC SSTs (depth SST estimates for a depth of 20 cm) minus matched drifting buoys, averaged on 5° latitude-longitude cells. This is averaged across all sensors, so AATSR is heavily weighted because of the increase in drifting buoy numbers since the year 2000. Results are shown for dual-view retrievals using (left) two (D2) and (right) three (D3) channels.

mean discrepancy being warmer than the D2 nighttime discrepancy by 0.06 K, and warmer than the D3 nighttime discrepancy by 0.05 K. This may partly reflect mean thermal stratification between the ARC $SST_{1m}$ and the Argo SST depth. Nonetheless, the results suggest that the 0.1 K target accuracy is met for ARC SSTs from AATSR and ATSR-2 and also that the SSTs from these two sensors are consistent with each other to well within 0.1 K. The intersatellite consistency has been achieved in ARC by exploiting the overlap period of ATSR-2 and AATSR to obtain consistency in the calibration and simulation of their brightness temperatures.

[33] The overlap of ATSR-2 and ATSR-1 was similarly exploited to tie ATSR-1 at the end of its life to ATSR-2 at the start of its life. However, ATSR-1 presents additional challenges. The detector temperature of ATSR-1 was not stable, which affects the calibration of at least the 12 $\mu$m detector temperature. The calibration impact of the detector temperature trend has been modeled using the best available information of the impact on the sensor calibration, but it is not clear how to tie the start of life of ATSR-1 to ATSR-2. In ARC, we elected to tie the D2 SST at the detector temperatures prevalent at the start of the ATSR-1 mission to the SSTs obtained using in addition the 3.7 $\mu$m channel, which was available for the first ~8 months of the mission. However, this is an area where more investigation should be done.

[34] ATSR-1 is also problematic because of the stratospheric aerosol present from May 1991 and diminishing through to roughly the end of 1993, arising from the massive eruption of Mount Pinatubo in the Philippines. The ARC coefficients, following the techniques of *Merchant et al.* [1999], are designed to be "aerosol robust" [*Embury and Merchant*, 2012], i.e., insensitive to the presence of this mode of aerosol. Robustness depends on accurate forward modeling of the brightness temperature impact of the aerosol relative to aerosol-free sky. A residual sensitivity to the presence of aerosol is therefore possible. (This is examined in section 4.3.1.)

[35] There are thus two reasons why early ATSR-1 SSTs could be biased relative to later ATSR-1 SSTs: uncertainty in the effect of changing detector temperatures, and residual sensitivity to stratospheric aerosol. It appears from Table 1 that ATSR-1 SSTs over the full lifetime are negatively biased by between −0.05 and −0.1 K relative to the later sensors, depending on the in situ measurements used for the comparison.

[36] The time evolution of the monthly, global, median discrepancy relative to drifting buoys is shown in Figure 4 (left). For comparison, we show the equivalent plot for Argo matches from 2004 onward (derived from <1% as many matches as the drifting buoy plot, but nonetheless giving a consistent picture). During the overlap of ATSR-1 and ATSR-2 (roughly the last 6 months of 1995), the overlap

**Table 1.** Global Median Discrepancy (K) Between ARC SSTs and Different Types of In Situ Measurements[a]

| | D2 Daytime | D2 Nighttime | D3 Nighttime | $SST_{skin}$-$SST_{depth}$ Day, Night |
|---|---|---|---|---|
| AATSR–Drifters | 0.040 | 0.013 | 0.021 | −0.129, −0.166 |
| AATSR–GTMBA | 0.030 | −0.004 | 0.018 | −0.094, −0.164 |
| AATSR–Argo | 0.072 | 0.009 | 0.021 | −0.117, −0.158 |
| ATSR2–Drifters | 0.036 | 0.034 | 0.034 | −0.124, −0.173 |
| ATSR2–GTMBA | 0.044 | 0.021 | 0.026 | −0.111, −0.162 |
| ATSR1–Drifters | −0.051 | −0.081 | −0.112 | −0.107, −0.184 |
| ATSR1–GTMBA | −0.058 | −0.104 | −0.117 | −0.108, −0.158 |

[a]The median discrepancies quoted are all matches between the indicated ATSR sensor and type of in situ observation. ATSR-1 measurements predate the Argo system, and for ATSR-2 only a few tens of matches are found (not shown). Note that the D2 results for ATSR-1 are quoted for SSTs using cloud detection that does not depend on the 3.7 $\mu$m channel which failed; the D3 results for ATSR-1 come from a restricted number of matches from the ~8 month period before the 3.7 $\mu$m channel failed. ARC $SST_{0.2m}$ data are used for the statistics with respect to drifters, and ARC $SST_{1m}$ data are used for Argo and GTMBA statistics. $SST_{skin}$-$SST_{depth}$ is the mean skin-to-depth adjustment (K) calculated in the satellite-based SSTs, where the depth is 0.2 m (drifters) or 1 m (GTMBA, Argo).
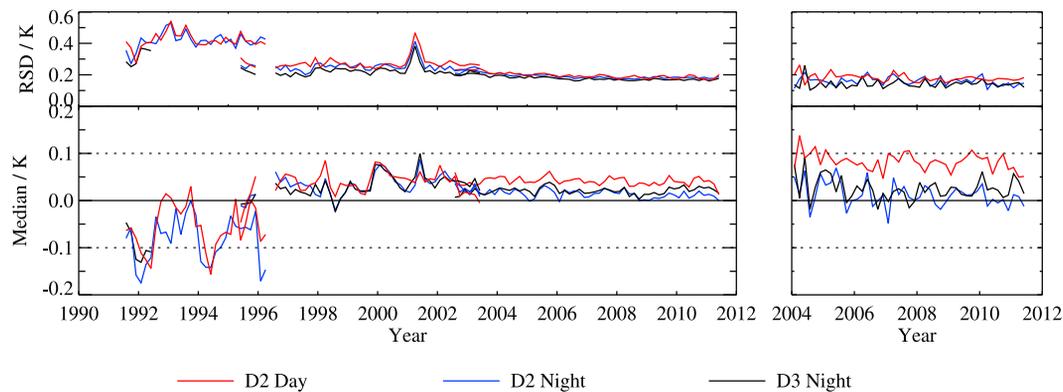
**Figure 4.** (left) Time series of (bottom) median discrepancy and (top) robust standard deviation (RSD) for the ATSR mission compared to drifting buoys. Results are shown for D2 daytime (red), D2 nighttime (blue) and D3 (black) retrievals. (right) The equivalent time series for AATSR compared to Argo. RSD is calculated using the median absolute deviation from the median, scaled by a factor such that for a Gaussian distribution, the RSD equals the conventional standard deviation.

analysis has brought the median discrepancy of the two sensors into alignment to within 0.05 K, which gives us an estimate of how closely the two sensors agree. The agreement between ATSR-2 and AATSR during their overlap (late 2002, early 2003) is closer. The median discrepancy relative to drifting buoys is relatively constant and generally between 0.00 and 0.05 K throughout the ATSR-2 and AATSR period. The robust standard deviation (RSD) of differences relative to drifting buoys for ATSR-2 is slightly larger than for AATSR, and again is stable, except for a few months in early 2001. This was a period when the attitude control of the ERS-2 platform was degraded, and thus the satellite geolocation and nadir-forward colocation have larger errors. This means the satellite and in situ SSTs are less precisely matched, and the errors from mismatch in location are significantly greater, adding to the RSD of discrepancy. However, there is no obvious effect on the mean discrepancy, so ARC SSTs from this period are not biased relative to before or after the event. The largest biases are associated with ATSR-1, as noted earlier, and here we see that there is an evolution of these biases in time, with a bias that is more negative than −0.1 K around the start of 1992 dissipating by mid-1993, followed by another negative excursion around the start of 1994, and a much smaller difference during late 1995. Given its timing, it is tempting to interpret the first negative excursion to residual sensitivity to the Pinatubo aerosol, which is considered further in section 4.3.1. In mid-1994, there was a rapid rise in the ATSR-1 detector temperatures (which are actively cooled) from around 92 K to around 98 K, and thereafter the detector temperatures rose to about 105 K by the end of the main ATSR-1 mission (mid-1996). This instrumental factor seems likely to play a role in the second negative excursion in the data. However, it is difficult to be definitive because the stability of the drifting buoy ensemble is not controlled or guaranteed during this period. Compared to the 2000s, the number of drifting buoys at that time was low, and the geographical coverage was uneven. These are both factors that render the stability of the validation values here open to question. A more formal analysis of the stability of ARC SSTs is therefore presented in section 3.3.

## 3.2. Standard Deviation (Precision)

[37] We estimate the precision of the ARC data by three-way analysis, which allows the simultaneous estimation of the precision (standard deviation of the errors) of each of three observation types. Here, version 5 SSTs from the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E) available from mid-2002 [*Wentz and Weissner*, 2000; *Wentz et al.*, 2003] were used as the third observation type with ARC SSTs and drifter SSTs. The AMSR-E data are gridded on a 0.25° latitude-longitude grid. The ARC-drifting buoy collocations were provided by the Met Office processing system used for the near real time monitoring of ATSR (details outlined by K. Lean and R. W. Saunders (Validation of the ATSR Re-processing for Climate (ARC) dataset using data from drifting buoys and a three-way error analysis, submitted to *Journal of Climate*, 2012)).

[38] *O'Carroll et al.* [2008] previously undertook a three-way analysis on earlier data sets of AATSR, AMSR-E and drifting buoy SSTs. They derived and discussed the applicability of an expression for the error variance ($\sigma_x^2$) of a set of observations $x$:

$$\sigma_x^2 = 0.5\left(V_{xy} + V_{zx} - V_{yz}\right) \tag{1}$$

where $V_{xy}$ is the variance in the difference between two observation types, $x$ and $y$, etc. In equation (1), the term $-0.5V_{yz}$ deducts the variance contribution from $y$ and $z$ from the mean of the variances of $x$ relative to $y$ and $z$ (which is $0.5(V_{xy} + V_{zx})$) to yield an estimate of the variance of $x$ itself. The three data sources must be closely collocated in time and space. Here a tolerance of different observation times of 180 min is used, and the buoy location must lie within the ARC grid cell, which in turn must lie within the AMSR-E grid cell. Differences in observation time and the different nature and spatial scales of the three types of observation mean that some true geophysical variability will be folded in unknown proportions into the precision estimates.

[39] The new three-way precision analysis was carried out for the years 2003–2009 in order to assess whether there are any trends in the precision of any of the types of observation.

**Table 2.** Standard Deviation of Error for 2003–2009 for ARC D3 $SST_{1m}$, AMSR-E SST and Drifting Buoy SST

| Instrument | Standard Deviation of Error for Each Year (K) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
| ARC $SST_{1m}$ | 0.137 | 0.129 | 0.139 | 0.137 | 0.138 | 0.136 | 0.134 |
| AMSR-E SST | 0.468 | 0.462 | 0.462 | 0.466 | 0.482 | 0.489 | 0.500 |
| Buoy SST | 0.189 | 0.174 | 0.155 | 0.152 | 0.149 | 0.149 | 0.153 |

Since we closely reproduce the approach of *O'Carroll et al.* [2008], there is good comparability with the earlier results.

[40] The inferred precision values are shown in Table 2. The same precision is found for the ARC data in 2003 as was found for the AATSR data in the study of *O'Carroll et al.* [2008]. Similar precision is also found for the AMSR-E SST in the two studies, while the precision of the buoy SSTs in this analysis is slightly lower. The precision estimates for the ARC SSTs have the smallest range over the seven years and do not have any obvious trend. For the AMSR-E SSTs there seems to be a deterioration in precision over time, which could be due to the instrument degrading with age and/or increasing radiofrequency interference. The buoy SST precision improves slightly in the early years and thereafter is stable. The Data Buoy Cooperation Panel (DBCP) indicates an increase in the number of drifting buoys reporting on the GTS between 2003 and 2005. The introduction of the regime with increased numbers of drifting buoys coincides closely with the improvement in precision. However, we have not found documented evidence that the "extra" buoys deployed are of different quality.

[41] The analysis above is possible only for AATSR (because of the availability of AMSR-E, an example of the utility of observing SST by several independent means). It is clear from Figure 4, and knowledge of the instruments involved that ATSR-2 SST precision is likely to be comparable to that of AATSR (except during the period of degraded geolocation accuracy), while ATSR-1 SSTs are markedly less precise.

### 3.3. Stability

[42] We assessed the temporal stability of ARC SST estimates at 1 m depth through comparison with SST measurements from GTMBA moorings in the tropical Pacific. The components of the GTMBA outside of the tropical Pacific have existed for too short a time for this application. Outside the tropics (e.g., Gulf Stream, Gulf of Mexico, UK/Western European Shelf), the operational moorings managed by the National Meteorological and Hydrological Services (NMHS; e.g., NOAA, UK Met Office) were assessed, but typically too few passed the selection criteria in each region. Thus, the stability of the ARC $SST_{1m}$ outside of the tropics is subject to ongoing investigation.

[43] Buoy measurements were extracted from the International Comprehensive Ocean-Atmosphere Data Set (ICOADSv2.5 [*Woodruff et al.*, 2011]) and collocated with the ARC SSTs. The ICOADS measurements were quality controlled using the ICOADS trimming flags to discard any observation more than 4.5 standard deviations from the climatological median. This should exclude any gross outliers in the buoy data but maintain the extreme values associated

with the El Niño Southern Oscillation (ENSO) [*Wolter*, 1997]. The $SST_{1m}$ values collocated with ICOADS moored buoy measurements are the value for a clear-sky 0.1° grid cell containing the buoy observation. The difference between the $SST_{1m}$ and the buoy SST was then calculated for each collocated pair of observations. The differences were deseasonalized (DSST hereafter) by subtracting the mean annual cycle in difference in each time series. Any DSST more than 3 standard deviations from the climatological monthly mean for a given buoy was discarded. This limit has been chosen to exclude any DSSTs that may be cloud contaminated or contain errors that have not been otherwise detected. Deseasonalizing was done to minimize the risk that aliasing of any annual cycle in difference would cause step changes to be falsely detected. This was not expected to be a problem for the tropics (since annual cycles are small), but was done in this study nevertheless.

[44] Each buoy used in the stability assessment had a minimum of 120 months with 5 or more DSST values over the period 1991–2009. Separate monthly mean time series of DSST were constructed for daytime and nighttime data from each buoy meeting this requirement. These were then assessed for step changes using a penalized maximal $t$ test (PMT) [*Wang et al.*, 2007]. Step changes were assumed to indicate spurious changes in the buoy data, unless they were clustered in time across multiple buoys, which would be more consistent with a step change in the ARC $SST_{1m}$ data. No such clustering in time was found, therefore any buoy with a step change statistically identified in either time series was discarded. The individual difference time series are noisy and the sensitivity of the PMT tests is therefore low. Noise in the individual time series will also affect any trend analysis.

[45] To increase the sensitivity of the PMT and reduce the impact of the noise in the differenced time series, the DSST values have been averaged across the retained buoys to give monthly mean composite time series for the daytime and nighttime ARC $SST_{1m}$ (Figure 5). When the PMT is applied to these combined time series, step changes are identified in both the daytime and nighttime series during 1993, with ARC $SST_{1m}$ ~0.1 K warmer after the change. The timing is consistent with the reduction of stratospheric aerosols after the 1991 eruption of Pinatubo. The step detection technique characterizes this as a step, but in reality it appears more gradual. The excursion of bias seen against drifting buoys in 1994 in Figure 4 is not evident here.

[46] In order to assess the stability of the ARC SSTs, a linear trend model with AR(1) errors (which allows some correlation between any given month and the previous month) has been fitted to the two difference time series from 1994 onward. No significant trends in the differences are found and the confidence intervals are smaller than the ARC target stability of 0.005 K $yr^{-1}$. The 95% confidence intervals for the trends are −0.0026 to 0.0015 K $yr^{-1}$ (daytime) and −0.0018 to 0.0019 K $yr^{-1}$ (nighttime). These results suggest that the ARC SSTs meet the target stability in the tropics from 1994 onward. As mentioned in section 3.1, sensitivity to the Pinatubo aerosols and sensor instability are candidate explanations for the ~0.1 K negative shift of the early ~2 years of the ATSR-1 SSTs. Both the PMT and trend analysis assume that the error characteristics of the monthly mean values tested remains constant over time.
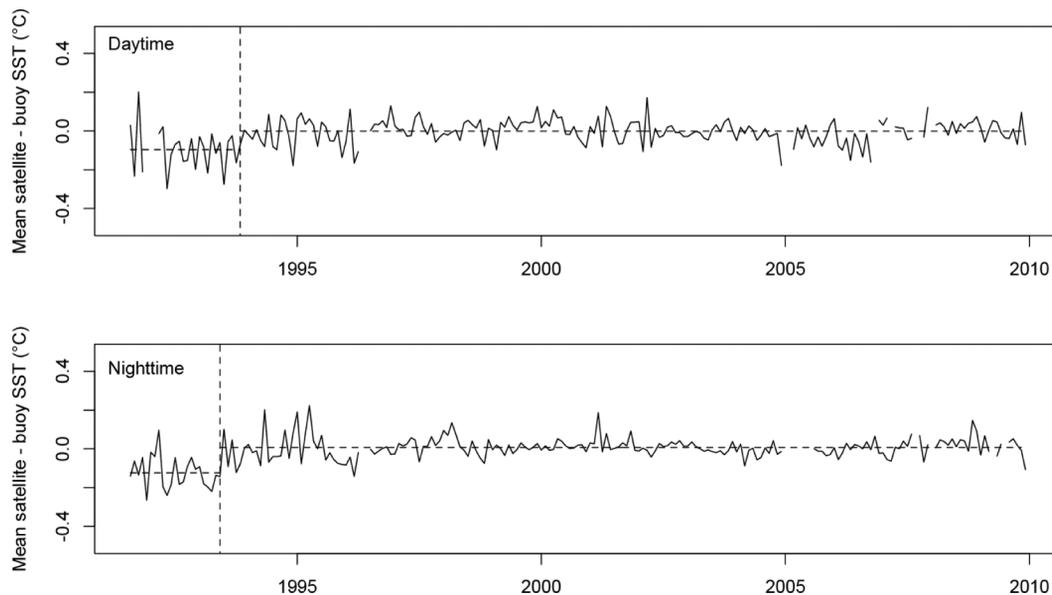
**Figure 5.** Time series of the deseasonalized composite monthly mean differences (DSST). The dashed lines indicate the identified break points and mean values for each segment.

However, the standard deviation of errors for the period of the ATSR-1 satellite are approximately double that for the ATSR-2 and AATSR periods due to the larger errors in ATSR-1 retrievals (as previously shown in Figure 4 (top)). Nevertheless, when the PMT and trend analysis are performed using only the ATSR-2 and AATSR data similar results are found.

[47] The lack of long-term stable reference sites, especially in the extratropics, has been problematic for this research, as has lack of accessible metadata on the existing in situ buoy SST measurements. It is not clear that the current in situ observing system is adequate for assessing the stability of the satellite SST record to the required accuracy outside of the tropics.

## 4. ARC SST Results

### 4.1. Mean and Seasonal SST

[48] The ability of an SST data set to represent the mean and seasonal distributions of SST is a basic requirement. Figure 6 shows the average SST over the period 1991 to 2010, for each of 4 months (January, April, July and October), as found in ARC SST$_{0.2m}$.

[49] The areas of highest average SST in Figure 6 are seen in the Gulf of Carpentaria to the north of Australia in January, in the Indian Ocean in April and in the Red Sea and Persian Gulf in July and October. SST is seen to fall below 0°C over large areas of the high northern latitudes in January, April and October. There is reduced coverage in high northern latitudes in July because nighttime data only are used for Figure 6. In the Southern Ocean, close to the ice edge, the SST remains below 0° in all months.

[50] The very sharp SST gradients along the northern boundary of the Gulf Stream are clearly seen, particularly in January and April. Areas of upwelling around the coasts and in the sharp eastern equatorial cold tongue are also evident.

[51] In both the Arctic and Antarctic, gaps are seen in data coverage due to seasonal sea ice cover.

### 4.2. Interannual Variability

[52] Although the mean seasonal behavior presented in section 4.1 is a feature that a satellite SST data set must represent well, arguably the interannual variability of SST anomaly is of more intrinsic interest because this is less well known from in situ measurements in many sparsely sampled regions of the ocean.

[53] The interannual variability in the ARC retrievals is shown via maps (Figure 7) for four example months (January, April, July and October) of the standard deviation across the time series of the SST monthly anomaly calculated over cells of a 5° latitude by 5° longitude grid. Areas of high variability associated with ENSO events are seen in the eastern and central tropical Pacific, particularly in January. Relatively quiescent regions are also seen, e.g., the Indian Ocean and the tropical warm pool region.

[54] When compared to fields of standard deviation for HadSST3 in the same months (Figure 8), the general patterns and magnitudes of variability seen in the ARC and HadSST3 data sets agree well, although the variability in the ARC data set is smoother. The ARC maps have large-scale patterns that are more coherent with fewer localized maxima of variability. This is likely due to the generally smaller errors of the ARC SSTs, and, for at least some years and grid cells, more adequate sampling. There are some isolated high-variability cells in the ARC standard deviation field. This can result from a combination of relatively small samples (e.g., because of high levels of cloud cover and/or sea ice) and some retrievals within those samples having large (outlier) errors because they were adversely affected by undetected cloud or sea ice. The ARC standard deviation field in the northwest Pacific hints at possible residual cloud contamination there in July. The pattern is different from that seen in the map for October, associated with the Kuroshio extension, and also from that seen in HadSST3. It also echoes a pattern (not shown) of much higher (>2 K), spurious variability previously identified in older ATSR SST
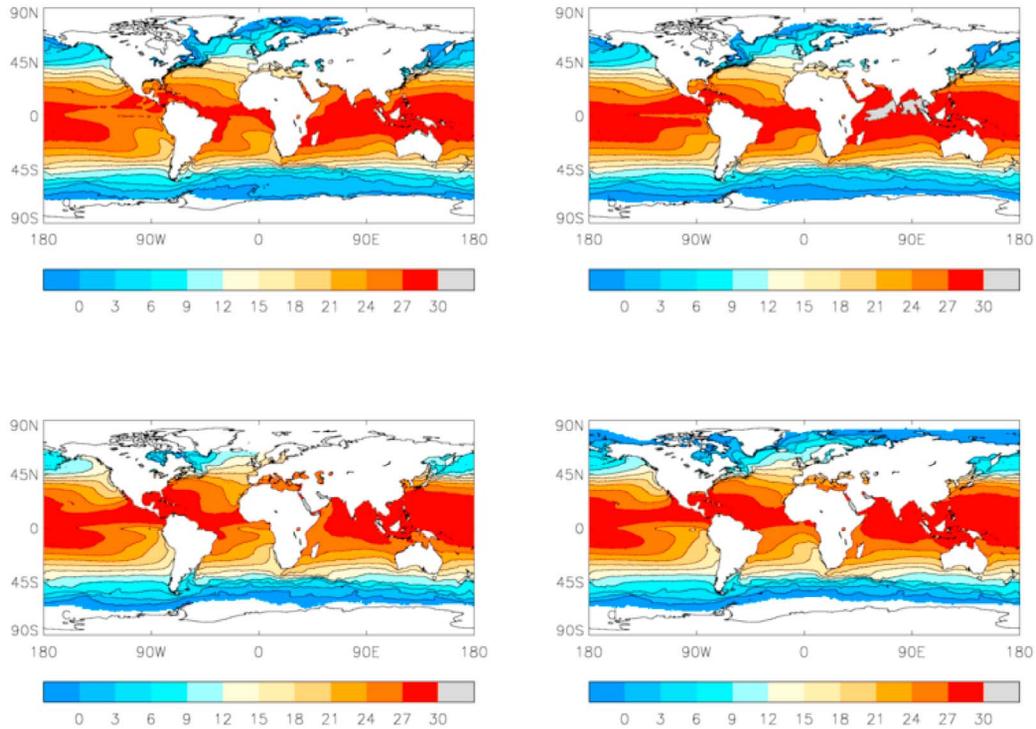
**Figure 6.** Average $SST_{0.2m}$ (°C), 1991–2010, on a 1° latitude by 1° longitude grid in (top left) January, (top right) April, (bottom left) July and (bottom right) October. Retrievals used here are D3 (nighttime only), except that D2 night is used during the ATSR-1 period with no D3. In summer, there may be no nighttime measurements at high latitudes because of persistent sunlight. The average SST value for a month was calculated in grid cells where SST was retrieved in that month during at least 1 year between 1991 and 2010.
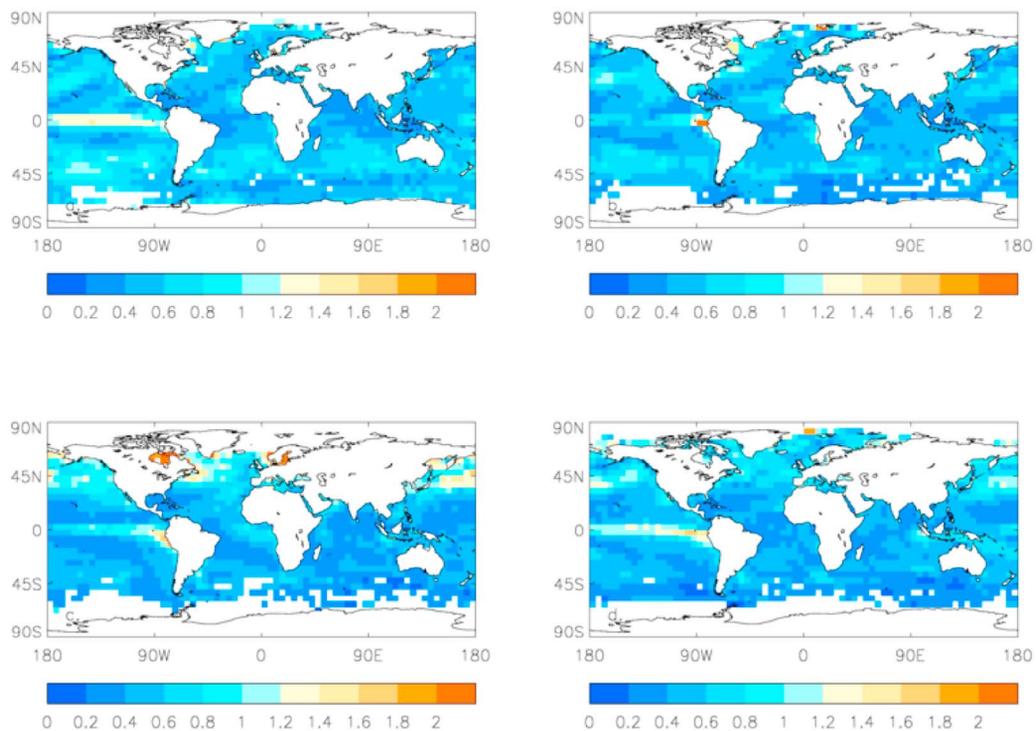


**Figure 7.** ARC $SST_{0.2m}$ standard deviation (°C), 1991–2010, on a 5° latitude by 5° longitude grid in (top left) January, (top right) April, (bottom left) July and (bottom right) October. Retrievals used were D3 when available and D2 night during ATSR-1 period with no D3.
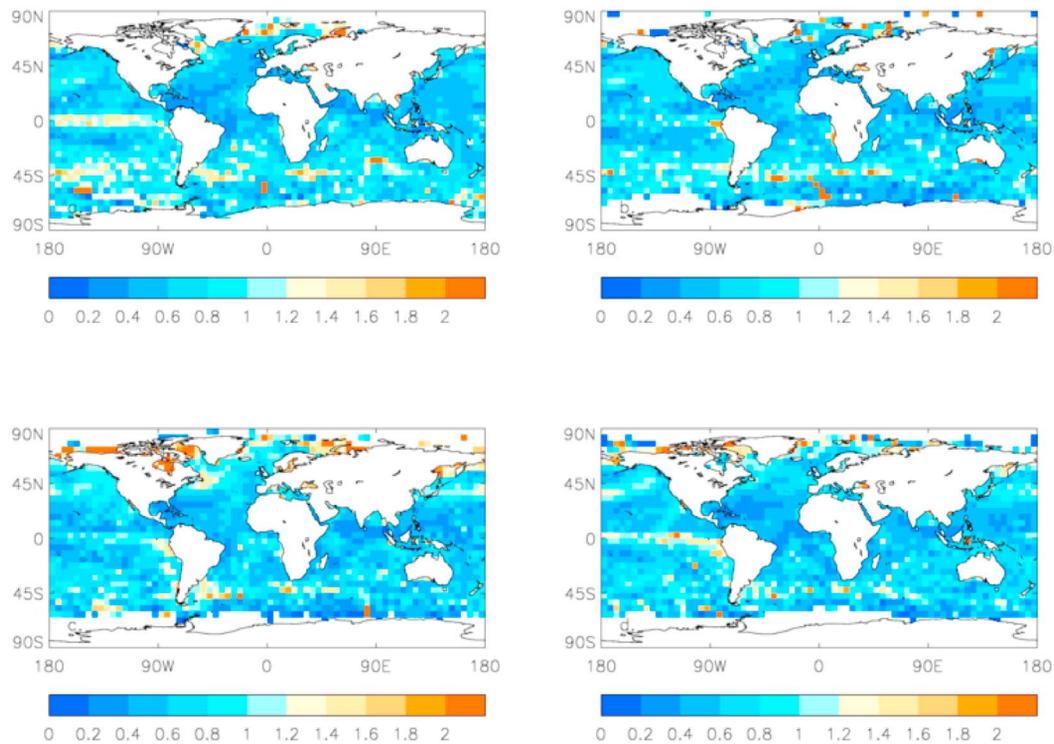
**Figure 8.** HadSST3 ensemble median SST standard deviation (°C), 1991–2010, on a 5° latitude by 5° longitude grid in (top left) January, (top right) April, (bottom left) July and (bottom right) October.

products in the northwest Pacific when the operational cloud clearing scheme was used; this was attributable to the inadequate screening of clouds.

[55] There is notably less variability in the ARC SSTs south of South Africa (i.e., the region of the Aghulas Retroflection and Return Current) and in the vicinity of the Falklands Current than is seen in the standard deviation maps for HadSST3. These are active ocean areas with strong thermal gradients. This is likely due to there being more spatially complete sampling and averaging over the 5° grid cells in the satellite data set than in the in situ data set; with fewer measurements, there is more scope in the in situ measurements for subsampling of the true spatial variability within the cell to lead to errors in the cell mean that appear as interannual variability in HadSST3. This interpretation is supported by the particularly close correspondence between ARC and HadSST3 variability in the North Atlantic, including the western boundary current region, where the in situ measurement coverage is very good.

### 4.3. Representation of Major Anomalies and Events

[56] Sections 4.1 and 4.2 demonstrate that the climatological mean and the variability of SST are convincingly represented by the ARC SST data set. In this section, we consider the representation of major events and anomalies. Because the ATSRs are narrow swath sensors with, mostly, one sensor in orbit, the sampling of SST anomaly events is sparser than is available from AVHRR instruments, which inevitably increases the sampling uncertainty in tracking an anomaly of interest. On the other hand, the accuracy and precision of each SST obtained is better, and the dual-view mode of retrieval should be more robust against atmospheric

anomalies including aerosol events. Looking at particular events allows us to assess the trade-off between more limited sampling and improved SST quality in the representation of important SST anomalies in the ARC data set.

### 4.3.1. Retrieval of SST After the Eruption of Mount Pinatubo in 1991

[57] In June 1991 the eruption of Mount Pinatubo in the Philippines created stratospheric aerosol that spread to all latitudes in the course of half a year and caused stratospheric aerosol levels well above background levels for ~2 years. Stratospheric aerosol generally cools the climate and such events are of interest in that they can provide insight into climate sensitivity to top-of-atmosphere forcing over interannual timescales. This particular aerosol period also coincided with strong El Niño activity in the tropical Pacific. Although this latter circumstance somewhat confounds the interpretation of the aerosol impact on climate, it is a period for which accurate knowledge of SST is of great interest. The ATSR-1 record commences 2 months after the main Pinatubo eruption. Dual view ARC SSTs (D3 and D2 retrievals) are designed to be robust to stratospheric aerosol, to a greater degree than is possible with a single-view sensor. So, it is useful to compare ARC SSTs with SSTs obtained from AVHRRs within the Pathfinder project over this period.

[58] The latitudinal average differences between the Pathfinder SSTs and the HadSST3 median show strong relative cooling in the tropics of more negative than −1 K in the Pathfinder data set between June 1991 and early 1992. By contrast, the ARC SSTs are relatively close to HadSST3 when first available in August 1991 and thereafter. This indicates that the dual-view retrievals that are by design "aerosol robust" are indeed relatively insensitive to the
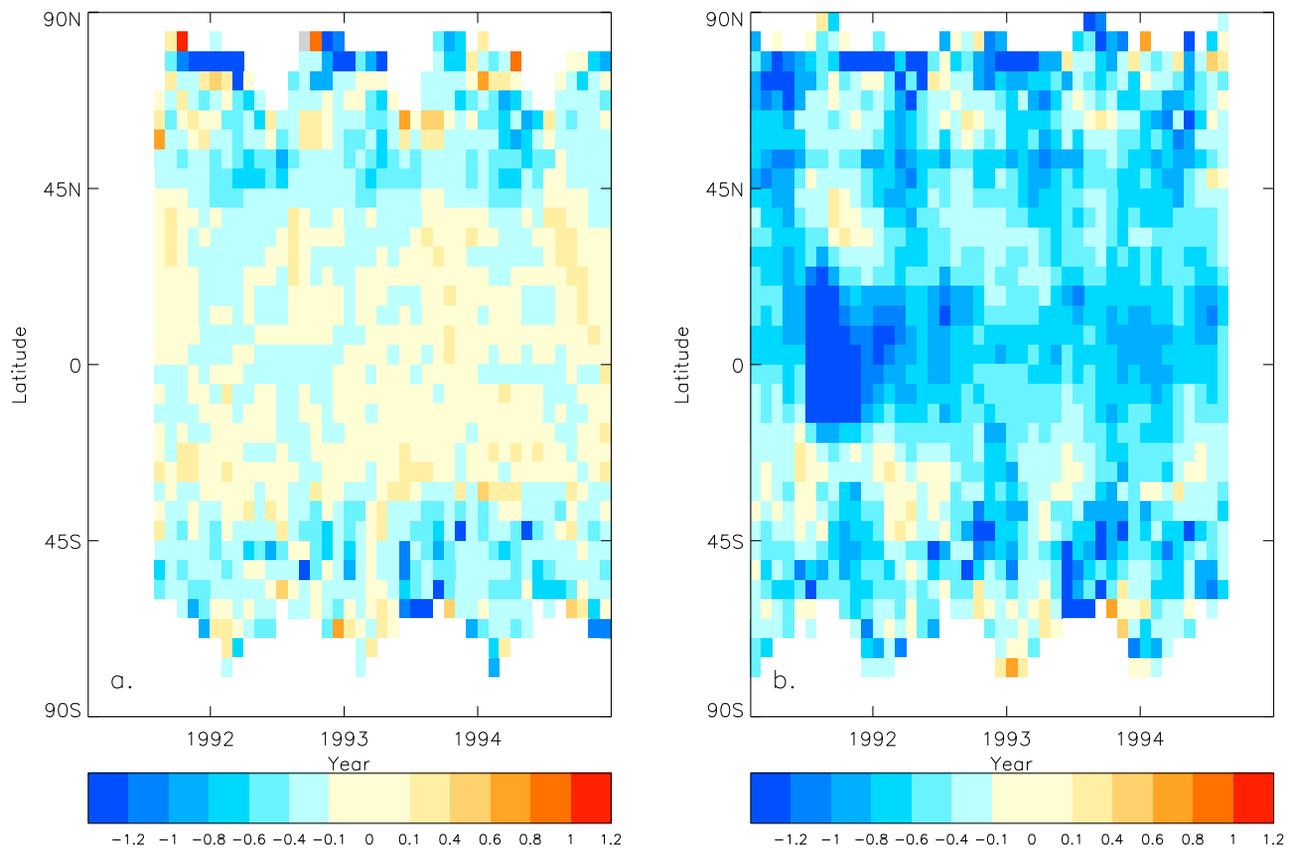
**Figure 9.** Zonal mean SST difference (°C), 1991–1994: (left) ARC SST$_{0.2m}$ minus HadSST3 median and (right) Pathfinder v5.2 minus HadSST3 median.

effects of the stratospheric aerosol (albeit with a residual cool bias of order a tenth of a degree, section 3.3).

[59] There are zonal differences of Pathfinder relative to HadSST3 that are warmer than +0.1 K during the period of the negative equatorial bias, around 40°N (late 1992) and 40°S (first half of 1993). These do not feature so prominently at these latitudes in other years and are probably also a consequence of the Pinatubo aerosol. The coefficients for Pathfinder are empirically derived over a 3 month rolling window, and give zero mean relative to the matches to drifting buoys used for the empirical regression. Since there is negative bias in the equatorial region arising from the direct effect of the stratospheric aerosol, it is reasonable that there is a positive bias induced in other zones.

[60] The aerosol-related large relative cold bias in the Pathfinder data is seen against a general background level of relative bias of a few tenths of a degree in that data set, relative to HadSST3.

[61] Other features are present in Figure 9 that do not relate to the Pinatubo event. Both satellite data sets show a seasonal cycle relative to HadSST3 at midlatitudes (roughly 40° to 60° in both hemispheres). In both satellite data sets, the midlatitude SSTs obtained in the winter are relatively cooler, while in summer, they are relatively warmer. The effect appears larger in the Pathfinder SSTs. This is not fully understood at present. Sampling effects could play a part, since the satellite observations should represent only clear-sky conditions, whereas the in situ measurements are all-weather. However, artifacts in the SST records arising from

cloud detection (differential levels of residual cloud contamination in different seasons) or SST retrieval may also contribute. There are some time-latitude cells with large relative biases in common between the two panels. These appear mostly at higher latitudes (for example, the negative bias between June and August 1993 at 60°S to 65°S). In these cases, the discrepancy apparently arises in HadSST3, either because of differential sampling within the zone (likely to be less spatially complete in HadSST3) or the influence of some biased in situ measurements.

### 4.3.2. Representation of the El Niño Event of 1997/1998

[62] As ENSO events have wide-ranging impacts on the climate system, it is important that any SST data set represents them well. Here we explore the large El Niño event of 1997/1998 and the subsequent La Niña. Figure 10 shows the average SST anomaly for the Niño 3.4 region for 1996–1999 and the evolution of the SST anomaly across much of the Pacific between 5°N and 5°S in the three data sets: ARC (satellite-only average), HadSST3 (in situ-only average) and the Daily Optimal Interpolation (OI) [*Reynolds et al.*, 2007] (a blend of in situ and satellite measurements).

[63] Overall, the agreement between ARC and HadSST3 in their representation of these events is good. However, the peak of the El Niño is stronger in ARC and the evolution of the subsequent La Niña event is more coherent than in HadSST3. The ARC SST anomaly is more negative before and after the El Niño event, consistent with the more coherent field seen here than in HadSST3. The Daily OI SST anomaly also appears more coherent than that of HadSST3.
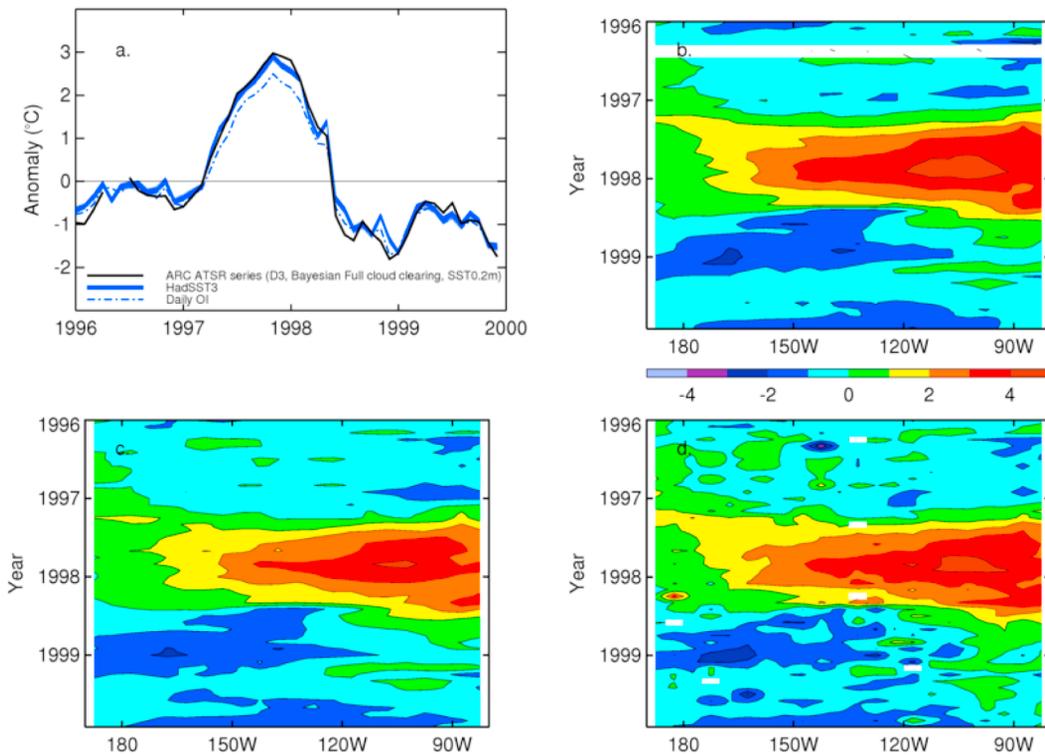
**Figure 10.** (a) Average SST anomaly (°C, relative to 1961–90) for Niño 3.4 region [170°W–120°W, 5°N–5°S], 1996–1999, from ARC $SST_{0.2m}$ (solid black), HadSST3 100-member ensemble (solid blue) and Daily OI (dot-dashed blue). (b–d) SST anomaly as above for longitudes 170°E–80°W, averaged over latitudes 5°N–5°S, for 1996–1999, as represented in different data sets: ARC (Figure 10b), daily OI (Figure 10c) and HadSST3 ensemble median (Figure 10d).

However, relative to ARC, Daily OI has a lower value for the peak of the El Niño by 0.4°C and underestimates the westward extent of anomalies above 1°C in spring of 1997. ARC and HadSST3 agree well on these features.

### 4.3.3. SST Anomalies in the Arctic Ocean

[64] The extent of Arctic sea ice cover has been decreasing observably during the past few decades [*Comiso*, 2002; *Comiso et al.*, 2008; *Serreze et al.*, 2007]. During the summer of 2007, there was a record-breaking retreat of Arctic sea ice cover, leading to much concern about the future stability of the Arctic environment. The loss of ice attracted much attention and explanations have been put forward [e.g., *Maslanik et al.*, 1996; *Comiso et al.*, 2008; *Rösel and Kaleschke*, 2012]. The SSTs recorded in the resultant clear water during the summer of 2007 were also remarkable [*Steele et al.*, 2008]. Figure 11 shows monthly SST anomalies from the ARC SST record, for the month of August in each year from 2004 to 2009. In 2007, the region of ice clear water in the Pacific sector of the Arctic Ocean was greater than in previous years. The size and the extent of the region of warm SST anomaly in the clear water is substantial. For much of the area the positive anomaly is around 4°C, and in places it exceeds values of about 6°C.

[65] Figure 12 shows in more detail the extent and intensity of this anomalous behavior in the summer of 2007. Figure 12 shows data for August 2007 at the full 0.1° resolution, and illustrates (white areas in N Pacific) the sampling limitations of the data set in persistently cloud areas, as well

as the good coverage obtained for this particular Arctic anomaly.

[66] The exceptional nature of the SST anomalies in the summer of 2007 is clear when compared to the anomaly time series over the preceding two decades. Summer monthly mean temperature anomalies in the box identified in Figure 12 are plotted in Figure 13 between 1991 and 2012. The August 2007 anomaly exceeds the mean for 1991–2009 by more than 3 standard deviations. Model studies [*Steele et al.*, 2010] suggest that this warm anomaly was caused primarily by elevated inward surface heat flux, allowed by an unusually early retreat of the ice sheet; the lack of ice cover also promoted increased wind forcing and consequent advection of warm water through the Bering Strait into the Chukchi, Laptev and Beaufort Seas.

### 4.4. Global and Regional Trends

[67] Most analyses of global sea surface temperature change have, so far, not made much use of satellite retrievals of SST, instead being based on SST measurements made in situ. Here we assess the global and regional changes seen in the ARC SST anomaly and, in the case of the global average SST anomaly, compare it to that from the HadSST3 ensemble.

[68] Figure 14 shows the global average SST anomaly, relative to the long-term average over the 1961–1990 period, from ARC and the HadSST3 ensemble. The HadSST3 ensemble represents the uncertainty in adjustments made to the in situ measurements to account for the effects of changes in measured SST arising from the evolution of the in situ
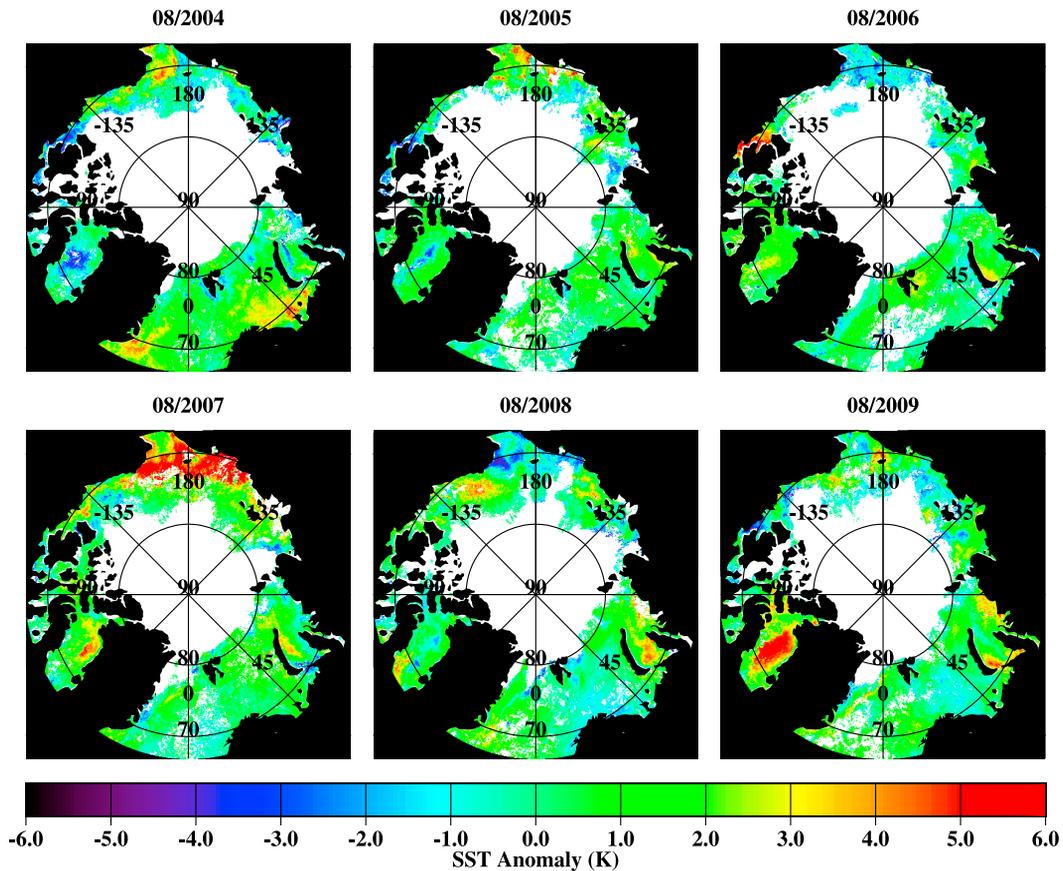
**08/2004**                            **08/2005**                            **08/2006**

**08/2007**                            **08/2008**                            **08/2009**

**Figure 11.** Monthly SST anomalies over the Arctic Ocean for the month of August are shown for each year from 2004 to 2009. The anomalies are the differences between the AATSR/ARC SST observations and a monthly ARC climatology, based on the three-sensor ATSR SST data record between 1992 and 2009. The daytime D2 SST data are shown, with cloud- and ice-masked areas in white.

observing system (see *Kennedy et al.* [2011b] for details). The global average SST anomaly from ARC sits within the spread of the HadSST3 ensemble for much of the record. Over the ATSR-2 and AATSR records, the global mean from ARC falls infrequently outside the envelope of the HadSST3 ensemble. However, the two sets of time series do differ apparently significantly during the ATSR-1 period, for reasons discussed in sections 2.5 and 3.1.

[69] As we have particular confidence in the stability of the ARC record from 1996 onward, we examine trends in
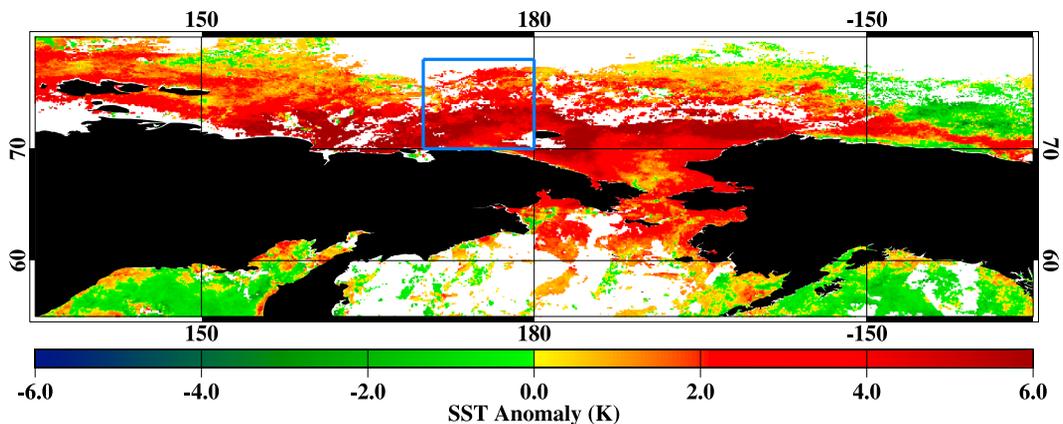
**Figure 12.** SST anomaly data from Figure 11 for August 2007 in cylindrical map projection, over the regions north of Siberia and Alaska. The sample box identified in the Chukchi Sea is the subject of Figure 13.
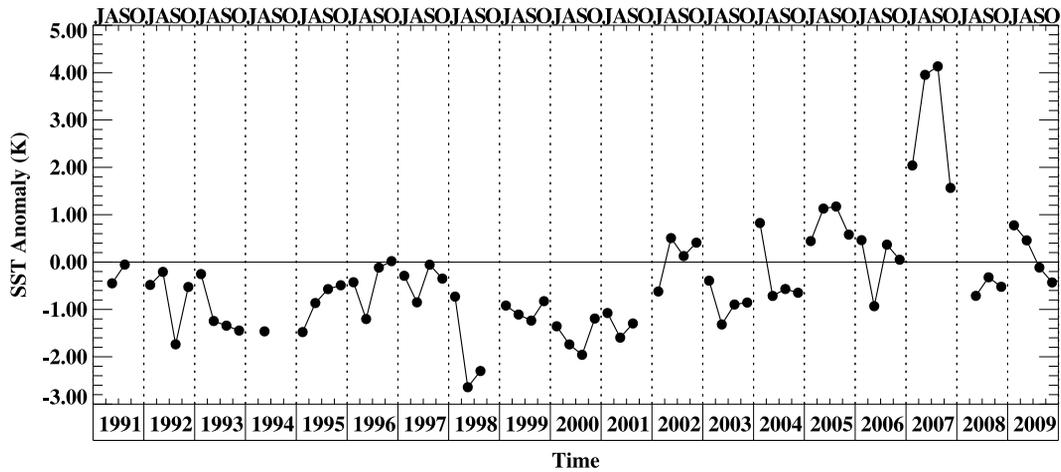
**Figure 13.** The monthly mean SST anomalies in the sample box identified in Figure 12 are shown for the ARC ATSR SST record up to 2009.

SST anomaly on the regional scale for that subperiod. Figure 15 shows a map of linear trends in individual monthly 1° latitude by 1° longitude grid boxes.

[70] The SST trends are not uniform in space. There are substantial areas of relatively little change (within ± 0.25°C/decade) and some areas of negative trend (particularly in the eastern Pacific and in the Southern Ocean, where trends are locally more negative than 0.5°C/decade. There are also areas of positive change, and in the high northern latitudes, local linear trends exceed 2°C/decade in some places.

## 5. Discussion

[71] Our aim in this paper is to present an overview of the ARC SST data set. Here, we summarize the points made and then comment on future related developments.

[72] The ARC SST data set is an observational data set, obtained by applying the physics of radiative transfer to cloud detection and SST estimation. The overlap of sensors in the series has been exploited to make the record homogeneous in time, and the SSTs for all sensors and types of retrieval have in effect been tied to the calibration of the 3.7 and 11 $\mu$m channel measurements of AATSR. The primary observed variable is the skin SST at the time of observation. The skin effect, near-surface stratification and diurnal heating rates are modeled to adjust the skin SST to SST at depth, standardizing for the half hour difference in satellite overpass time between Envisat and the earlier missions. Thus, $SST_{0.2m}$ estimates are also available in the ARC products, and are more directly comparable to in situ data.

[73] Validation results show that the ARC SSTs meet the 0.1 K accuracy that was a target for the project, a statement we justify by showing that the mean discrepancy between ARC $SST_{0.2m}$ and matched drifting buoys is generally within ±0.1 K for most regions (Figure 3) and for most of the record (Figures 4 and 5). The precision (standard deviation of errors) is better than 0.14 K for AATSR D3 SST (Table 2) and is similar for ATSR-2, and for D2 SST for both sensors (Figure 4) but is greater for ATSR-1. As befits a time series intended for use for climate, the record appears to be stable between 1994 and 2010 to better than the target 0.005 K $yr^{-1}$

with 95% confidence. However, we note that, so far, this has been confirmed only for tropical regions because of the difficulty of identifying adequately stable points of reference outside of the Global Tropical Moored Buoy Array. The ATSR-1 sensor is, over its life, biased cold by −0.05 to −0.1 K relative to the later sensors, which may partly be due to some sensitivity to post-Pinatubo volcanic aerosol in the stratosphere during 1991 to 1993 (Figure 5).

[74] The ARC data show the expected climatological behavior for SST (Figure 6) and appear to give useful, spatially coherent information on interannual variability (Figure 7 compared with Figure 8). While the sensitivity of the ARC
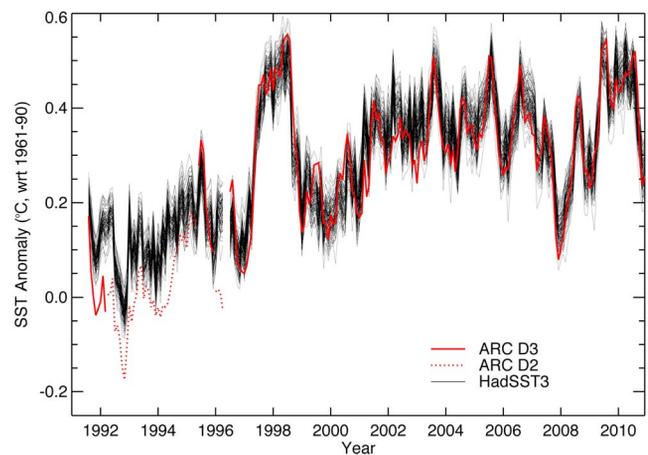


**Figure 14.** Global mean SST anomaly (°C, relative to 1961–1990). Red lines indicate ARC nighttime $SST_{0.2m}$ time series: solid red indicates D3 retrievals, used when available, and dotted red indicates D2 retrievals used during the ATSR-1 period with no D3. Black lines indicate the 100-member HadSST3 ensemble. Data were first averaged on to a 5° latitude by 5° longitude monthly grid, according to the method used by *Kennedy et al.* [2011a]. Only grid boxes where measurements/retrievals were available in both data sets were used.
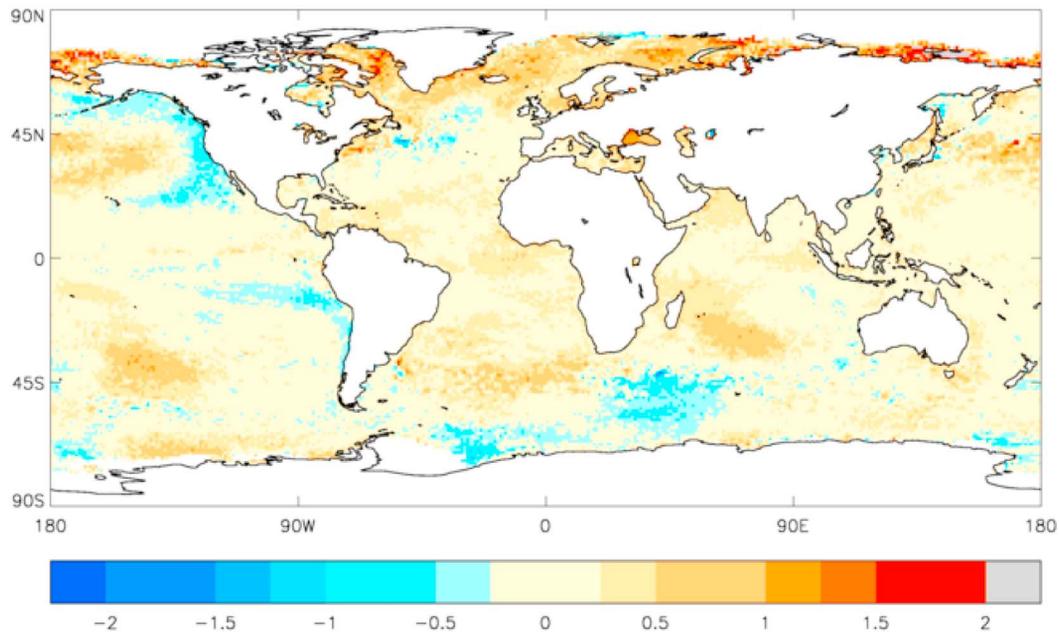
**Figure 15.** Linear trends in ARC SST$_{0.2m}$ anomaly (°C per decade) for 1° latitude by 1° longitude grid box averages, 1996–2010. Retrievals used were D3, when available, plus D2 night during ATSR-1 period. Linear trends were calculated using the method of median of pairwise slopes [*Lanzante*, 1996].

data set to the post-Pinatubo stratospheric aerosol has been mentioned, its ∼0.1°C magnitude is much less than for other satellite SST data sets (∼1°C) (Figure 9).

[75] ARC SSTs are particularly accurate and precise but, compared to other satellite data sets, are more sparsely sampled. Nevertheless, regional SST anomaly events are well quantified in the data, as shown with examples of El Niño and recent Arctic anomalies (e.g., Figures 10 and 13).

[76] On the basis of the above, we argue that the ARC SST data set gives a useful independent corroboration of data sets of SST change over recent decades. More is to be done, and we present here a first comparison of the global SST anomaly from ARC with an in situ only ensemble (HadSST3). There is good agreement in both the general trend and variability (Figure 14). The ∼0.1 K discrepancy in first few years is thought to arise from the satellite record, as discussed above. That two analyses based on independent measurements using different measurement techniques agree so closely suggests that no major unknown source of error at the global scale exists in either data set over the period 1996 to 2010.

[77] What of the future for independent, climate quality SST from space?

[78] Operations of the Envisat platform carrying AATSR terminated abruptly in April 2012, with no apparent prospect of recovery. There will be a gap of about 2 years or more before the next dual-view radiometer mission is operational, which will be the first Sea and Land Surface Temperature Radiometer (SLSTR) [*Donlon et al.*, 2012]. Being convinced of the value to climate research of an independent SST time series from space, concepts for bridging this gap between AATSR and SLSTR are under development. The SST retrieval scheme for SLSTR will be based on and consistent with the methods developed for the ATSRs in ARC. In the meantime, the ARC SST time series will soon be extended to April 2012 within the context of the European

Space Agency Climate Change Initiative (ESA CCI) project for SST.

[79] Work on maximizing the quality of the ATSR-based record will continue. The archive of calibrated, geolocated brightness temperatures (the "level 1b" imagery products used as inputs to ARC) will be reprocessed in the latter half of 2012, delivering improvements in geolocation and colocation between forward and nadir views, as well as improved visible channel calibration. It is intended to generate a version 2 of the ARC SST time series at some point thereafter, to take advantage of the improved level 1b archive.

[80] To summarize, the 20 year ARC SST data set is the first record of SST from space that combines independence from in situ measurements, consistency between different sensors, and consistency between retrievals based on different channel combinations. The data set provides both skin and depth SST estimates, with specific uncertainty estimates. The diurnal cycle has been accounted for to ensure the long-term stability of the SST-depth record is not affected by changing sampling within the diurnal cycle. In these several respects, ARC SST retrieval methods go beyond previous approaches to satellite-based SST. The result is a record of relatively accurate and precise SST observations with good representation of interannual variability and major SST anomaly events. High stability has been demonstrated, allowing the ARC SST time series to be used as an independent verification of the trends and changes in SST present in in situ-based data sets. We have presented the close agreement since 1996 between the ARC and HadSST3 global SST anomaly time series, which is an important, essentially independent corroboration of changes in global mean SST over the period seen in measurements made in situ. We expect ARC SSTs to have many more applications within climate science.

# References

Bourlès, B., et al. (2008), The PIRATA Program: History, accomplishments, and future directions, *Bull. Am. Meteorol. Soc.*, *89*, 1111–1125, doi:10.1175/2008BAMS2462.1.

Comiso, J. C. (2002), A rapidly declining perennial sea ice cover in the Arctic, *Geophys. Res. Lett.*, *29*(20), 1956, doi:10.1029/2002GL015650.

Comiso, J. C., C. L. Parkinson, R. Gersten, and L. Stock (2008), Accelerated decline in the Arctic Sea ice cover, *Geophys. Res. Lett.*, *35*, L01703, doi:10.1029/2007GL031972.

Donlon, C. J., T. J. Nightingale, T. Sheasby, J. Turner, I. S. Robinson, and W. J. Emery (1999), Implications of the oceanic thermal skin temperature deviation at high wind speed, *Geophys. Res. Lett.*, *26*, 2505–2508, doi:10.1029/1999GL900547.

Donlon, C. J., P. J. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray (2002), Toward improved validation of satellite sea surface skin temperature measurements for climate research, *J. Clim.*, *15*(4), 353–369, doi:10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2.

Donlon, C., et al. (2012), The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission, *Remote Sens. Environ.*, *120*, 37–57, doi:10.1016/j.rse.2011.07.024.

Embury, O., and C. J. Merchant (2012), A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme, *Remote Sens. Environ.*, *116*, 47–61, doi:10.1016/j.rse.2010.11.020.

Embury, O., C. J. Merchant, and M. J. Filipiak (2012a), A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer, *Remote Sens. Environ.*, *116*, 32–46, doi:10.1016/j.rse.2010.10.016.

Embury, O., C. J. Merchant, and G. K. Corlett (2012b), A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability, *Remote Sens. Environ.*, *116*, 62–78, doi:10.1016/j.rse.2011.02.028.

Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young (1996), Bulk parameterization of air-sea fluxes for TOGA COARE, *J. Geophys. Res.*, *101*, 3747–3764, doi:10.1029/95JC03205.

Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson (2003), Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm, *J. Clim.*, *16*, 571–591, doi:10.1175/1520-0442.

Good, E. J., X. Kong, O. Embury, and C. J. Merchant (2012), An infrared desert dust index for the along-track scanning radiometers, *Remote Sens. Environ.*, *116*, 159–176, doi:10.1016/j.rse.2010.06.016.

Hanafin, J. A., and P. J. Minnett (2005), Measurements of the infrared emissivity of a wind-roughened sea surface, *Appl. Opt.*, *44*(3), 398–411, doi:10.1364/AO.44.000398.

Harris, A. R., M. A. Saunders, J. S. Foot, K. F. Smith, and C. T. Mutlow (1995), Improved sea surface temperature measurements from space, *Geophys. Res. Lett.*, *22*, 2159–2162, doi:10.1029/95GL02019.

Immler, F. J., J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vomel (2010), Reference quality upper-air measurements: Guidance for developing GRUAN data products, *Atmos. Meas. Tech.*, *3*, 1217–1231, doi:10.5194/amt-3-1217-2010.

Ingleby, B., and M. Huddleston (2007), Quality control of ocean temperature and salinity profiles–Historical and real time data, *J. Mar. Syst.*, *65*, 158–175, doi:10.1016/j.jmarsys.2005.11.019.

Katsaros, K. B. (1977), Sea-surface temperature deviation at very low wind speeds; is there a limit?, *Tellus, Ser. A*, *29*(3), 229–239, doi:10.1111/j.2153-3490.1977.tb00729.x.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011a), Reassessing biases and other uncertainties in sea surface

temperature observations since 1850: 1. Measurement and sampling errors, *J. Geophys. Res.*, *116*, D14103, doi:10.1029/2010JD015218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011b), Reassessing biases and other uncertainties in sea surface temperature observations since 1850: 2. Biases and homogenization, *J. Geophys. Res.*, *116*, D14104, doi:10.1029/2010JD015220.

Kennedy, J. J., R. O. Smith, and N. A. Rayner (2011c), Using AATSR data to assess the quality of in situ SST observations for climate studies, *Remote Sens. Environ.*, *116*, 79–92, doi:10.1016/j.rse.2010.11.021.

Kent, E. C., S. D. Woodruff, and D. I. Berry (2007), Metadata from WMO publication No. 47 and an assessment of voluntary observing ship observation heights in ICOADS, *J. Atmos. Oceanic Technol.*, *24*(2), 214–234, doi:10.1175/JTECH1949.1.

Kent, E. C., J. J. Kennedy, D. I. Berry, and R. O. Smith (2010), Effects of instrumentation changes on ocean surface temperature measured in situ, *Wiley Interdiscip. Rev. Clim. Change*, *1*(5), 718–728, doi:10.1002/wcc.55.

Kilpatrick, K. A., G. P. Podestá, and R. Evans (2001), Overview of the NOAA/NASA advanced very high resolution radiometer pathfinder algorithm for sea surface temperature and associated matchup database, *J. Geophys. Res.*, *106*, 9179–9197, doi:10.1029/1999JC000065.

Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, *16*(11), 1197–1226, doi:10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L.

Liu, W. T., K. B. Katsaros, and J. A. Businger (1979), Bulk parameterization of air-sea exchanges of heat and water vapor including the molecular constraints at the interface, *J. Atmos. Sci.*, *36*, 1722–1735, doi:10.1175/1520-0469(1979)036<1722:BPOASE>2.0.CO;2.

Maslanik, J. A., M. C. Serreze, and R. G. Barry (1996), Recent decreases in Arctic summer ice cover and linkages to atmospheric circulation anomalies, *Geophys. Res. Lett.*, *23*(13), 1677–1680, doi:10.1029/96GL01426.

McPhaden, M. J., et al. (1998), The Tropical Ocean-Global Atmosphere (TOGA) observing system: A decade of progress, *J. Geophys. Res.*, *103*, 14,169–14,240, doi:10.1029/97JC02906.

McPhaden, M. J., G. Meyers, K. Ando, Y. Masumoto, V. S. N. Murty, M. Ravichandran, F. Syamsudin, J. Vialard, L. Yu, and W. Yu (2009), RAMA: The Research Moored Array for African-Asian-Australian monsoon analysis and prediction, *Bull. Am. Meteorol. Soc.*, *90*, 459–480, doi:10.1175/2008BAMS2608.1.

Merchant, C. J., and A. R. Harris (1999), Toward the elimination of bias in satellite retrievals of sea surface temperature: 2. Comparison with in situ measurements, *J. Geophys. Res.*, *104*, 23,579–23,590, doi:10.1029/1999JC900106.

Merchant, C. J., and P. Le Borgne (2004), Retrieval of sea surface temperature from space, based on modeling of infrared radiative transfer: Capabilities and limitations, *J. Atmos. Oceanic Technol.*, *21*(11), 1734–1746, doi:10.1175/JTECH1667.1.

Merchant, C. J., A. R. Harris, M. J. Murray, and A. M. Závody (1999), Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison, *J. Geophys. Res.*, *104*, 23,565–23,578, doi:10.1029/1999JC900105.

Merchant, C. J., A. R. Harris, E. Maturi, and S. Maccallum (2005), Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval, *Q. J. R. Meteorol. Soc.*, *131*(611), 2735–2755, doi:10.1256/qj.05.15.

Merchant, C. J., et al. (2008), Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers, *Adv. Space Res.*, *41*(1), 1–11, doi:10.1016/j.asr.2007.07.041.

Merchant, C. J., A. R. Harris, H. Roquet, and P. Le Borgne (2009), Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer, *Geophys. Res. Lett.*, *36*, L17604, doi:10.1029/2009GL039843.

O'Carroll, A. G., J. R. Eyre, and R. W. Saunders (2008), Three-way error analysis between AATSR, AMSR-E and in situ sea surface temperature observations, *J. Atmos. Oceanic Technol.*, *25*(7), 1197–1207, doi:10.1175/2007JTECHO542.1.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution blended analyses for sea surface temperature, *J. Clim.*, *20*, 5473–5496, doi:10.1175/2007JCLI1824.1.

Rösel, A., and L. Kaleschke (2012), Exceptional melt pond occurrence in the years 2007 and 2011 on the Arctic sea ice revealed from MODIS satellite data, *J. Geophys. Res.*, *117*, C05018, doi:10.1029/2011JC007869.

Serreze, M. C., M. M. Holland, and J. Stroeve (2007), Perspectives on the Arctic's shrinking sea-ice cover, *Science*, *315*(5818), 1533–1536, doi:10.1126/science.1139426.

Smith, D., C. Mutlow, J. Delderfield, B. Watkins, and G. Mason (2012), ATSR infrared radiometric calibration and in-orbit performance, *Remote Sens. Environ.*, *116*, 4–16, doi:10.1016/j.rse.2011.01.027.

Steele, M., W. Ermold, and J. Zhang (2008), Arctic Ocean surface warming trends over the past 100 years, *Geophys. Res. Lett.*, *35*, L02614, doi:10.1029/2007GL031651.

Steele, M., J. Zhang, and W. Ermold (2010), Mechanisms of summertime upper Arctic Ocean warming and the effect on sea ice melt, *J. Geophys. Res.*, *115*, C11004, doi:10.1029/2009JC005849.

Walton, C. (1985), Satellite measurement of sea surface temperature in the presence of volcanic aerosols, *J. Clim. Appl. Meteorol.*, *24*(6), 501–507, doi:10.1175/1520-0450(1985)024<0501:SMOSST>2.0.CO;2.

Wang, X. L., Q. H. Wen, and Y. Wu (2007), Penalized Maximal *t* Test for detecting undocumented mean change in climate data series, *J. Appl. Meteorol. Climatol.*, *46*, 916–931, doi:10.1175/JAM2504.1.

Wentz, F., and T. Weissner (2000), Algorithm Theoretical Basis Document (ATBD), Version 2. AMSR Ocean Algorithm, *RSS Tech. Proposal 121599A-1*, 66 pp., NASA, Washington, D. C.

Wentz, F., C. Gentemann, and P. Ashcroft (2003), On-orbit calibration of AMSR-E and the retrieval of ocean products, paper presented at 12th Conference on Satellite Meterology and Oceanography, Am. Meteorol. Soc., Long Beach, Calif. [Available at http://ams.confex.com/ams/pdfpapers/56760.pdf.]

Wimmer, W., I. S. Robinson, and C. J. Donlon (2012), Long-term validation of AATSR SST data products using shipborne radiometry in the Bay of Biscay and English channel, *Remote Sens. Environ.*, *116*, 17–31, doi:10.1016/j.rse.2011.03.022.

Woodruff, S. D., et al. (2011), ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive, *Int. J. Climatol.*, *31*, 951–967, doi:10.1002/joc.2103.

Wolter, K. (1997), Trimming problems and remedies in COADS, *J. Clim.*, *10*, 1980–1997, doi:10.1175/1520-0442(1997)010<1980:TPARIC>2.0.CO;2.