Remote Sensing of Environment 165 (2015) 27–41

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A decade of sea surface temperature from MODIS

K.A. Kilpatrick *, G. Podestá, S. Walsh, E. Williams, V. Halliwell, M. Szczodrak, O.B. Brown, P.J. Minnett, R. Evans

University of Miami Rosenstiel School for Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, Fl, L 33149, United States

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 16 April 2015
Accepted 22 April 2015
Available online 15 May 2015

Keywords:
MODIS
Sea surface temperature
Validation
Calibration

ABSTRACT

The MODIS (MODerate Resolution Imaging Spectroradiometers) on the *Terra* and *Aqua* satellites, both part of NASA's Earth Observing System, have each been providing high quality global sea-surface temperatures (SSTs) for over a decade. The ability to retrieve accurate SST from satellites is dependent on many factors including spectral response, radiometer noise, pre-launch instrument characterization and in flight behavior, calibration, viewing geometry, cloud screening and correction of the clear-sky atmospheric effects. We present a characterization of both the historical and current MODIS SST products, and we describe the evolution of, and motivation for, improvements in MODIS SST for each Collection released to the public, i.e., each radiance level mission data reprocessing, each driven by improved knowledge of the instrument behavior and better corrections for instrumental artifacts.

1. Introduction

Accurate global measurements of SST are critical to our understanding of past, current and future climate, reliable nowcasts and forecasts of the ocean state, and enhanced numerical weather prediction. Satellite remote sensing provides the best method of deriving accurate ocean-wide SSTs on daily to decadal periods. Methods used to derive SST in the infrared (IR) from space-based measurements of top of the atmosphere (TOA) brightness temperatures (BT) using dual or split window algorithms are mature and the physical processes governing this measurement are well understood. Infrared (IR) satellite SST retrieval is based on measurements taken where the atmosphere is relatively transparent, in so-called "atmospheric windows" in the mid-wave infrared (MWIR, $\lambda = 3.5$–4.1 μm) and long-wave, thermal infrared (LWIR, $\lambda = 10$–12 μm) spectral intervals. The ability to retrieve accurate SST from space is dependent on a number of factors. Some of these factors are related to the radiometer, including its spectral response, noise level, and quality of pre-launch characterization and on-orbit performance, in-flight calibration, and viewing geometry. Other factors are related to the intervening atmosphere: as SST measurements in the IR require clear sky conditions, detection of clouds and atmospheric aerosols are important aspects of the accuracy and precision of SST retrievals; cloud and aerosol detection, however, is not discussed in detail here. Our focus is on correction of the effects on IR propagation through clear atmospheres.

The importance of satellite SST measurements is clearly demonstrated by the succession of IR radiometers that have been operating since the launch of the first Advanced Very High Resolution Radiometer (AVHRR) in 1978. The AVHRR was followed by a spacecraft radiometer specifically designed for accurate SST measurement, the Along-Track Scanning Radiometer, ATSR (Edwards et al., 1990; Minnett, 1995). Since the late 1990s, a suite of spacecraft IR sensors has been introduced by both the USA and European space agencies. These instruments were specified and designed to have the potential to retrieve SST with accuracies approaching those believed at the time to be needed for climate research (Harries et al., 1983). This paper focuses on one of these instruments, the MODIS (MODerate resolution Imaging Spectroradiometer) of NASA's Earth Observing System (EOS) constellation. Two MODIS instruments have been flown and continue to produce data: the first one on *Terra* (launched on 18 December 1999) and the second on *Aqua* (launched on 4 May 2002).

This paper describes the aspects of the MODIS instrument relevant to the retrieval of SST, and the SST products derived from this sensor. First, we introduce the MODIS design and highlight its innovations over heritage sensors. Then, we present a discussion of SST algorithms, the pre- and post-launch characterization of the MODIS instruments, the evolution of calibration procedures, and corrections of instrument artifacts. Finally, we provide data on the precision and accuracy of the currently available MODIS SST products (Collection 5).

2. MODIS instrument design

MODIS passively measures the reflected/emitted earth radiance at the top of the atmosphere at 36 different wavelengths (Salomonson, Barnes, Maymon, Montgomery, & Ostrow, 1989). *Terra* MODIS has a daytime descending orbit with a 10:30 am local equatorial crossing time, whereas *Aqua* MODIS has a daytime ascending orbit and a 1:30 pm local equatorial crossing time. The MODIS sensors were

* Corresponding author at: University of Miami Rosenstiel School for Marine and Atmospheric Science, 4600 Rickenbacker Causeway, Miami, FL 33149, United States.
E-mail address: kkilpatrick@rsmas.miami.edu (K.A. Kilpatrick).

http://dx.doi.org/10.1016/j.rse.2015.04.023
0034-4257/

designed for multi-disciplinary applications and research covering the atmosphere, land surfaces, the cryosphere and the oceans. The selection of spectral bands for SST was based primarily on the heritage instrument: AVHRR. Additional bands were included to augment the potential of MODIS beyond heritage instruments. The spectral characteristics of all 36 MODIS bands are discussed elsewhere (Barnes, Pagano, & Salomonson, 1998; Esaias et al., 1998; Guenther et al., 1996; Salomonson et al., 1989). The five MODIS IR bands and the bandwidths used to provide daily global estimates of SST are listed in Table 1.

MODIS was the first spacecraft radiometer to have multiple infrared bands in the MWIR atmospheric transmission window (bands 20, 22 and 23). As this window is more transparent and less variable than the LWIR region (bands 31 and 32), the MWIR region is very suitable for the derivation of SST. In addition, by being on the short-wavelength side of the peak of Planck's function at ocean temperatures, this spectral interval offers greater sensitivity to changes in the SST than the LWIR, even though the signal itself is smaller. As with all other infrared radiometers on spacecraft that take measurements in the MWIR atmospheric transmission window, the ability to derive SST from the measurements depends on a radiometer design with a suitable signal-to-noise ratio. While some heritage instruments had a single channel in this spectral region and were used in conjunction with the LWIR to retrieve SST (Llewellyn-Jones, Minnett, Saunders, & Zavody, 1984), MODIS was the first and only sensor to date for which SST could be derived from the MWIR window alone. This product is referred to as "SST4", where the "4" indicates the proximity of the bands used to MWIR wavelengths of 4 μm. Unfortunately, the MWIR measurements are contaminated by scattered and reflected solar radiation – "sun glint" – in daylight; therefore only night retrievals are possible using these bands.

Besides the additional spectral bands, several innovative aspects of the MODIS design extended its capabilities beyond those of heritage instruments. Such innovations included a double-sided "paddle-wheel" scan mirror and multiple detectors for each spectral band: ten detectors for each of the 1-km resolution bands, digitized to 12 bits.

MODIS IR measurements are calibrated on-orbit using an internal blackbody target and a view of deep, cold space, in a fashion similar to that used for the AVHRR. Unlike AVHRR's base-plate target, however, the MODIS blackbody calibration target is internal to the instrument and thereby avoids the large temperature excursions experienced in the AVHRR calibration procedure around the orbit (Brown, Brown, & Evans, 1985; Trishchenko, Fedosejevs, Li, & Cihlar, 2002). The advantages and problems of some of the MODIS instruments' design features are discussed in detail in subsequent sections, as they have impacts on the accuracy of SST retrievals.

3. MODIS SST algorithms and coefficient estimation

This paper is authored by members of the NASA MODIS SST science team and focuses primarily on the currently available, Collection 5, products made available from NASA Goddard Ocean Biology Processing Group (OBPG) (www.oceancolor.gsfc.gov). The products of each "Collection" represent a revision of both the instrument calibration model and the SST geophysical retrieval algorithm. Public releases of ocean products prior to 2005, Collections 3 and 4, were produced at the NASA EOS MODIS Adaptive Processing System (MODAPS) and are no longer continued. These earlier Collections are discussed as appropriate to provide historical context. The issues identified in earlier products motivate changes made for Collection 5, and reflect our ongoing efforts to reduce systematic errors in the retrieved MODIS SST when possible.

Table 1
MODIS long- and mid-wave infrared bands used in SST retrieval (MODIS Theoretical Basis Document No. 25, 1999).

Band number	Band center (μm)	Band width (μm)
20	3.750	0.1800
22	3.959	0.0594
23	4.040	0.0608
31	11.030	0.5000
32	12.020	0.5000

Two ocean skin SST products are produced from the MODIS IR Collection 5 data. The first product, commonly referred to as the "SST", is produced from MODIS LWIR bands 31 and 32 ($\lambda = 11$ and 12 μm, respectively). The form of the algorithm is the heritage AVHRR Pathfinder algorithm (Kilpatrick, Podestá, & Evans, 2001), used to produce and extend the 30 year SST climate record from NOAA's AVHRR sensors (Kilpatrick, 2013). The functional form of the Pathfinder SST algorithm is a derivative of the split window NLSST (non-linear SST) algorithm of (Walton, Pichel, Sapper, & May, 1998):

$$SST = c_{1+}c_2 * T_{11} + c_3 * (T_{11} - T_{12}) * T_{sfc} + c_4 * (\sec(\theta) - 1) * (T_{11} - T_{12}) \quad (1)$$

where SST is the retrieved surface temperature, T_{11} and T_{12} are the brightness temperatures (BTs) in the 11 and 12 μm bands, T_{sfc} is an estimate of the surface temperature, and θ is the satellite zenith angle. For MODIS the T_{sfc} is the NOAA/NCEP weekly or daily blended Reynolds 0.25 degree spatial resolution product (Reynolds, Rayner, Smith, Stokes, & Wang, 2002; Reynolds et al., 2007). Parameters c_1 through c_4 are the algorithm coefficients for the MODIS 11–12 μm SST, derived using collocated in situ buoy SSTs and satellite observations.

Coefficients are estimated for (i) low and high atmospheric water vapor regimes, and for (ii) month of observation. For most months from launch until 2004 each month in the satellite record used a weighted 5-month sliding window [0.25, 0.5, 1.0, 0.5, 0.25]. In a few exceptions from early in the TERRA mission prior to 2002, when the instrument experienced several shut-downs and safe-holds, the five month window had to be adjusted or truncated. Starting in 2005 the systematic bi-weekly calibration updates by the MODIS calibration support team (MCST) used in the generation of the top of the atmosphere radiances (level 1b products; L1b) results in BTs with a predictably stable calibration. As a result of this stability in the level-1b products, and no Pinatubo-size volcanic eruptions injecting large quantities of ash into the stratosphere (Reynolds, 1993), the SST algorithm coefficients no longer required updates on a monthly basis. For 2005–2014, Collection 5 products use month-of-year coefficients developed from training data collected in 2004.

The accuracy and limitations of an NLSST algorithm have been discussed by many others (Barnes et al., 1998; Barton, 1995; Barton, Prata, & Llewellyn-Jones, 1993; Llewellyn-Jones et al., 1984; McClain, Pichel, Walton, Ahmad, & Sutton, 1983; Merchant, Harris, Roquet, & Le Borgne, 2009; Minnett, 1986; Walton, McClain, & Sapper, 1990; Walton et al., 1998). Brown and Minnett also provided a detailed discussion in the original NASA MODIS SST Algorithm Theoretical Basis Document (MODIS ATBD no. 25, 1999, available from http://modis.gsfc.nasa.gov/data/atbd/ocean_atbd.php). Three important sources of variability in the accuracy of MODIS SST retrievals are (i) absorption in the 11 and 12 μm bands caused by water vapor, a highly variable component of the atmosphere, (ii) changes in atmospheric path length (sometimes called air mass) due to scan angle, and (iii) the need for an estimate of the surface temperature which acts as a proxy scaling function for the brightness temperature deficits ($SST_{insitu} - T_{11\ or\ }T_{12}$) at different environmental temperatures and water vapor concentrations. An additional source of error in LWIR NLSST algorithms involves episodic variations in aerosol absorption due to volcanic ash or terrigenous dust.

The second product, SST4, is unique to MODIS and is derived from measurements in the 4 μm MWIR atmospheric window, bands 22 and 23. As will be shown below, this is a superior product at night compared to the heritage-based NLSST LWIR SST. The MODIS MWIR bands 22 and 23, 3.9 μm and 4.0 μm, are less susceptible to atmospheric absorption (Fig. 1) and are much narrower than MODIS band 20, 3.7 μm

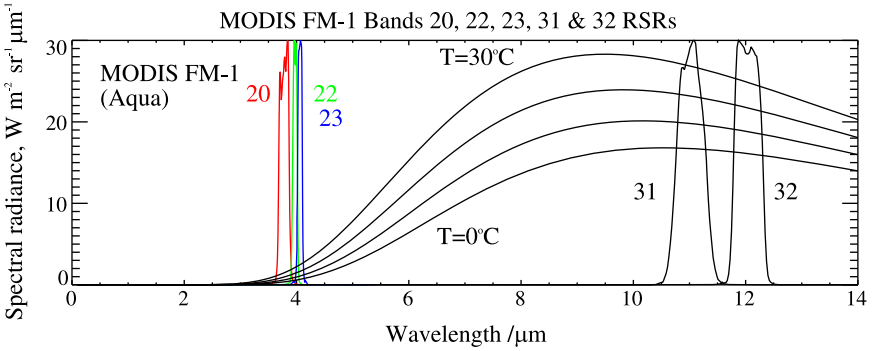

Fig. 1. Atmospheric transmission spectra in the Infrared for Polar, Mid-latitudes and Tropical atmospheres: Top panel. Relative spectral response for the MODIS IR bands: bottom panel. The smoothed lines are the Plank's function between 0 and 30 °C (Minnett & Barton, 2010).

(Table 1). Band 20 is comparable to the AVHRR's channel 3b, used in NOAA's operational nighttime 3-band SST product.

The form of the SST4 algorithm is:

$$SST4 = c_{1+}c_2 * T_{3.9} + c_3 * (T_{3.9} - T_{4.0}) + c_4 * (\sec(\theta) - 1) \quad (2)$$

where the $T_{3.9}$ and $T_{4.0}$ are the brightness temperatures for MODIS bands 22 and 23 in the 4 μm MWIR atmospheric window.

Originally, the NASA Science Team planned to use the radiative transfer model (RTM) of Závody (Závody et al., 1995) to derive algorithm coefficients to produce SST from MODIS. This RTM, which used the HITRAN library of spectral properties of atmospheric gases (Rothman et al., 1987), was the basis of the atmospheric correction algorithms to derive skin SSTs from the ATSRs on ERS-1 and ERS-2. Such use of an RTM requires an accurate and traceable sensor model, as well as a good representation of the atmospheric state and accurate spectral data (Llewellyn-Jones et al., 1984). Once Terra was on orbit, it immediately became clear that the pre-launch instrument characterization had not been adequate and several artifacts were present in the images, making the RTM approach not viable. All official NASA MODIS SST products released to the public are therefore based on empirically derived coefficients for SST.

After launch, until enough collocated in situ observations were collected on orbit to assemble an adequate matchup database (MDB, co-located in situ and satellite observations), the SST Science Team derived the first coefficients using co-located same-day SST retrievals from the AVHRR Pathfinder SST night-time products. This version, known as MODIS Collection 3, was the first MODIS SST dataset made available to the public. To derive the coefficients for the Collection 3 products, several days of co-temporal AVHRR global 4 km level 3 Pathfinder SSTs and MODIS BTs were regressed to produce a single set of coefficients for each of two water vapor regimes; as defined by a threshold of 0.7 K for the difference between bands 31 and 32. The ability to estimate algorithm coefficients using the well-validated AVHRR, against the full MODIS swath and using the synoptic coverage of level 3 images, helped to partially correct for some of the initial at-launch imperfectly characterized instrumental artifacts in the *Terra* MODIS, as these artifacts would masquerade as part of the atmospheric effects. Additional days of independent Pathfinder SST fields were used to validate the Collection 3 products and to quantify the MODIS response versus scan angle (RVS) anomaly in the LWIR band, which will be discussed in more detail below. The Pathfinder SST also aided in identifying an L1b data problem in the MWIR bands, which was subsequently corrected by the MODIS Calibration Support Team (MCST), that resulted in a seasonal and hemispheric bias in the MWIR SST4s in the first on-orbit *beta* MODIS *Terra* fields (Collection 2, not released).

Beginning with Terra MODIS Collection 4, the MODIS SST science team and the MCST responsible for the accuracy of the L1b input products, understood better the instrument artifacts and the limitations of the instrument calibration model. Sufficient numbers of in situ buoy measurements had also been acquired and MODIS SSTs were reprocessed by NASA using coefficients derived from buoys in a match-up database (MDB). Table 2 provides a summary of the methods used for coefficient estimation for each collection. The experience gained and lessons learned from comparing Terra MODIS to AVHRR were transferred directly to Aqua and allowed the initial Aqua products to be declared, validated and released more rapidly after launch.

The MDB used to estimate coefficients includes temporally (±30 min) and spatially (within 10 km) coincident MODIS BTs and other ancillary data (e.g., viewing geometry) and in situ buoy SST. The in situ observations from moored and drifting buoys came from two different sources. Prior to 2011, buoy data were obtained via the US Navy from the GTS network, and were quality-controlled as described by Kilpatrick (Kilpatrick et al., 2001). After 2011, quality-controlled buoy data were obtained from the NOAA/NESDIS/STAR In situ Quality Control Monitor or iQUAM (Xu & Ignatov, 2013).

The MODIS sensor measures IR radiation originating from a layer less than 1 mm thick at the ocean surface; the temperature of this layer is referred to as the surface skin temperature, or skin SST. MODIS SST products are reported as skin SST, not a sub-surface temperature. Although the empirical algorithm coefficients are derived from sub-surface buoy measurements, the c_1 coefficients in Eqs. (1) and (2) are adjusted to derive a skin temperature by subtracting 0.17 K. This value of −0.17 K represents the average skin to sub-surface temperature difference (Donlon, Minnett, Gentemann, & Nightingale, 2002), as measured by ship-borne radiometers. Comparisons of the MODIS skin SST to sub-surface in situ buoy SSTs reported in this paper are therefore expected to have a cold bias (i.e., satellite cooler than buoy) of about −0.2 K. That is, the satellite SST residuals relative to buoy

Table 2
Methods used for coefficient estimation for each MODIS collection.

Product release	Coefficients
Collection 1 & 2 Terra	Coefficients based on a radiative transfer model. Not released to public.
Collection 3 Terra	Single set of coefficients based on regressions against AVHRR skin SST temperature estimates
Collection 4 Terra	Algorithm coefficients estimated for each month/year (e.g., March 2002) using in situ SSTs from buoys. The intercept (c1) bias was adjusted to produce a skin temperature.
Collection 5 2000–2004 Terra and Aqua	Algorithm coefficients estimated for each month/year (e.g., March 2002) using in situ SSTs from buoys. The intercept (c1) bias was adjusted to produce a skin temperature.
Collection 5 2005 to present Terra and Aqua	Algorithm coefficients estimated for each month of the year (e.g., August) using in situ SSTs from buoy data for 2004. The intercept (c1) bias was adjusted to produce a skin temperature.

measurements should not be centered around zero, if the algorithm and coefficients are working properly.

The relationship between skin and sub-surface temperature is, however, not constant. Under low winds, this relationship can be highly variable — vertically, horizontally and temporally (Donlon et al., 2002; Minnett, 2003; Minnett & Ward, 2000; Murray, Allen, Merchant, Harris, & Donlon, 2000; Ward, 2006). In addition to the cool skin effect, the sub-surface SST can differ from the skin SST as a result of diurnal heating. The difference depends on the depth of the in situ measurement, as well as the heating and cooling history of the near-surface layers on a given day, and can result in the skin temperature being much warmer than the sub-surface SST (Minnett, 2003). Differences often exceed 1 K, occasionally reaching >6 K under highly stratified diurnal thermoclines (Gentemann, Minnett, Leborgne, & Merchant, 2008). The use of sub-surface temperature for satellite validation adds the consequence of these near-surface gradients to the error budget of satellite retrievals and leads to an over-estimate of retrieval uncertainties (Kearns, Hanafin, Evans, Minnett, & Brown, 2000; Minnett, 1991). For validation and coefficient estimation, we use only nighttime observations in an attempt to avoid the possibility of highly stratified diurnal thermoclines. Additionally, uncertainties of in situ buoy temperatures have recently been shown to be larger than previously thought, ~0.15 to 0.25 K; this makes a significant contribution to the overall satellite SST error budget (Emery, Sandra, Wick, Peter, & Craig, 2001; O'Carroll, August, Le Borgne, & Marsouin, 2012).

In addition to comparisons to sub-surface buoy observations, MODIS SST products also were validated against skin SST measurements from shipboard radiometers such as the Marine-Atmosphere Emitted Radiance Interferometer (M-AERI; (Minnett, Knuteson, Best, & Osborne, 2001)) and ISARs (Infrared Sea surface temperature Autonomous Radiometer; (Donlon et al., 2008)) deployed during research cruises and on ships of opportunity. These radiometers provide highly accurate measurements of skin SST, and avoid the potentially confounding presence of sub-surface gradients. These radiometers provide a good estimate of the retrieval uncertainty. However, the number of available measurements by shipboard radiometers is much smaller than those from the large network of buoys.

4. On-orbit characterization of sensor artifacts in the MODIS SST products

Following launch, skin SSTs derived from the MODIS calibrated TOA BTs were characterized and validated using in situ SSTs and SSTs from other validated satellite-IR and microwave SST sensors. We found four main problem areas in the early Collections: in order of discussion these are (i) response versus scan angle (RVS) across the swath (caused by an angular dependence of the scan mirror reflectivity), (ii) differences between mirror sides, (iii) differences between detectors, and (iv) changes in instrument electronic configurations.

4.1. Response versus scan angle

As mentioned in the Introduction, the use of a large, two-sided rotating paddle-wheel scan mirror in MODIS was a significant departure from the optical design of the heritage instruments (Barnes & Salomonson, 1993). An advantage of the traditional design is a constant angle of incidence (45°) on the scan mirror of the electromagnetic energy from the earth that reaches the detectors. Thus, the reflectance of the mirror surface is constant and any imperfections, including degradation over time, could be well compensated during the calibration procedure. However, when used with multiple detectors per band, a 45° mirror leads to "image rotation" across the swath. That is, the projection of a row of detectors onto the earth surface is aligned in the flight direction at nadir but rotates as the scan angle departs from nadir. This can be avoided by using a two-sided paddle-wheel mirror in which the axis of rotation is through the center of the mirror and parallel to the reflecting surfaces, being aligned with the vector of spacecraft velocity. On the other hand, a disadvantage of a paddle-wheel mirror is the changing reflectivity for radiation reaching the detectors as the mirror rotates; these changes are a function of the angle of incidence on the mirror surface and wavelength of the radiation.

The angular dependence of the spectral reflectivity of the *Terra* MODIS mirror was determined before launch by measurements of mirror witness samples at three separate laboratories. In contrast, the measurements for the *Aqua* MODIS were done on the actual mirror sides. The at-launch correction for *Terra* MODIS was derived by averaging the *Terra* witness sample measurements and the actual *Aqua* mirror measurements, with the expectation that this approach would require post-launch refinement. Characterization and continued monitoring of the response versus scan angle and mirror side remain to this day critically important to the ability of MODIS to provide accurate TOA BTs and SSTs.

Significant residual reflectivity variations vs. scan angle (RVS) were found on orbit in *Terra* LWIR SST first light retrievals, and were identifiable as unrealistic gradients in the SST across the swaths. Large linear temperature discontinuities were also present in composites or global Level 3 images when adjacent overlapping passes from opposite sides of the swath, with different angles on the mirror surface, were combined (Fig. 2). The along-scan striping seen in the expanded view of the Caribbean Sea near Cuba (Fig. 2D) has been traced to noise in the on board digitizer, very small calibration imbalances between the 16 detectors, and differences between the two mirror sides. The largest and most problematic of the Terra SST issues at-launch was the RVS issue in the LWIR SST, which resulted in almost a 2 K difference between the SST4 and the SST products across the scan (Fig. 2E).

As discussed in the next paragraph, the large RVS artifact was subsequently corrected and discontinuities at swath boundaries were no longer visually apparent in the first public release (Collection 3) but RVS characterization continued to be refined in later releases.

Initially the natural variability of SST fields, combined with the large RVS, prevented the empirical derivation of a consistent set of coefficients for the atmospheric correction algorithm (Collection 2). A couple of months into the mission, a fortuitous and temporary solution to the RVS problem came when the scan mirror was inadvertently left rotating and data were collected while the earth-view aperture door was closed. Assuming that the temperature of the inside of the door was uniform, the measured gradients would be the result of the residual, uncorrected RVS effects. Using the data collected with the closed earth-view door, it was feasible to distinguish the different RVS properties of the two mirror sides; both were more pronounced than indicated by the pre-launch characterization. The RVS correction derived from these measurements produced a marked improvement in the continuity of SST across swath boundaries, and therefore replaced the pre-launch scheme.

A better approach to characterizing the RVS is to measure the emission from the scan mirror during a "deep space maneuver" (DSM), which entails increasing the pitch rotation rate of the spacecraft while in eclipse, so that for part of this time the MODIS (and other sensors on Terra) are directed towards deep space, directly away from the sun. Three years after *Terra's* launch, the first DSM was conducted on 26 March 2003. The maneuver resulted in an improved and more traceable RVS correction. Again, the correction was different for the two mirror sides. Moreover, the DSM correction was greater than the previously implemented, Collection 3, closed-door correction. Fig. 3 shows the normalized percent reflectivity of the scan mirror for each of the three RVS characterizations: pre-launch, closed door, and DSM, for LWIR bands 31 and 32. The LWIR bands had a much larger RVS signal than the MWIR (not shown).

The difference in the RVS correction based on the DSM relative to the RVS correction from the closed door is shown in Fig. 4, as functions of

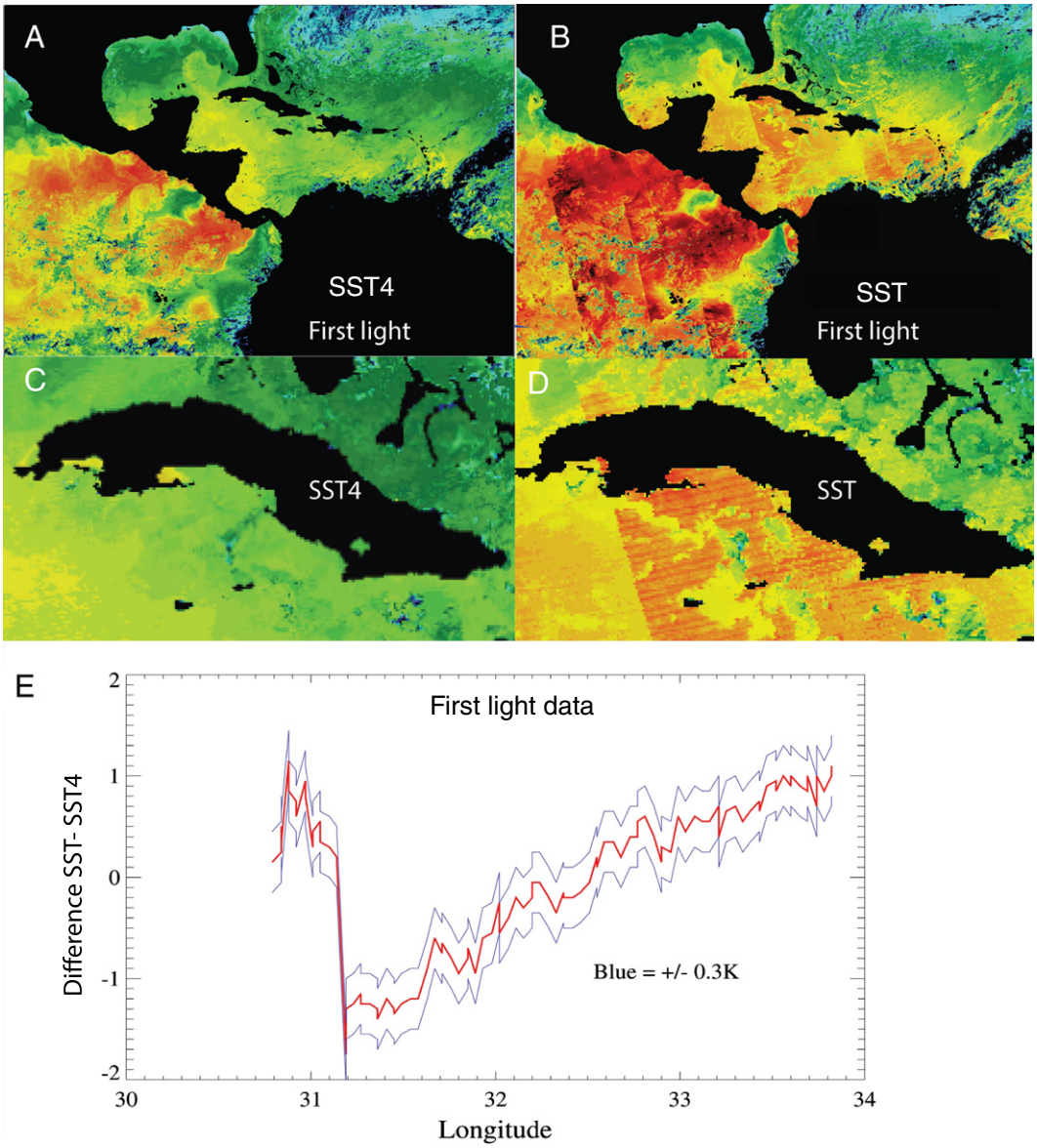

Fig. 2. First light image using a "warmest pixel" composite of TERRA SST's of the Gulf of Mexico region. These composites are from 20 to 27 March 2000 using the at-launch instrument characterization. Selecting the warmest pixel in a grid cell reduces the effects of clouds. Where no cloud-free pixels exist in a cell, the cells will appear cloudy. Panel B is representative of the at-launch impact on SST of uncorrected RVS in the LWIR first light level 1b data. Panel A: MWIR SST4, B: LWIR SST. Magnified region around Cuba C: SST4, D: SST, E: Average difference in SST4 − SST (red line) from transects across the swath from multiple granules located between 31 to 35 degrees W longitude showing a 2 K difference across the scan. The blue lines are the envelope of the ±estimate of the pre-launch SST RMS . All images have the same color scale.

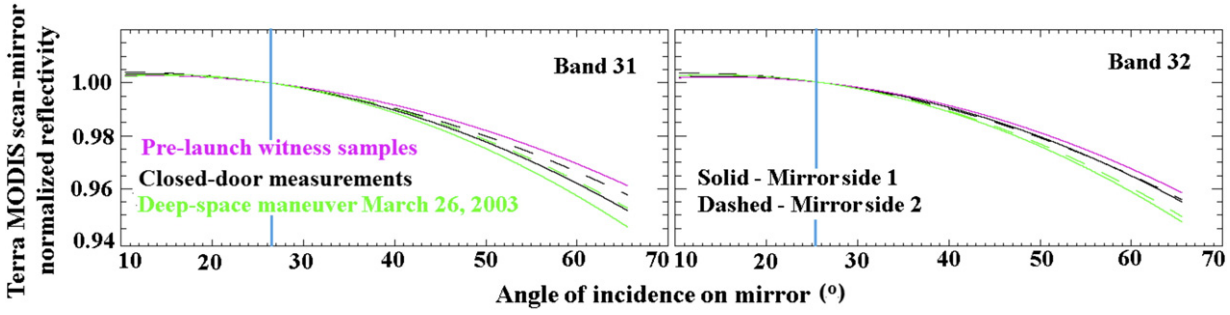

Fig. 3. Terra normalized percent mirror reflectivity as a function of angle of incidence derived from pre-launch witness samples of the Terra MODIS mirror and actual Aqua mirror measurements (magenta), closed door measurements (black), and the Terra Deep-space maneuver of 2003 (green). The values for the two mirror sides are shown as solid and dashed lines for the measurements taken on-orbit. The normalization is set to unity at the angle on the mirror (26.3°) of the measurement of the internal blackbody calibration target location shown by the blue vertical line.

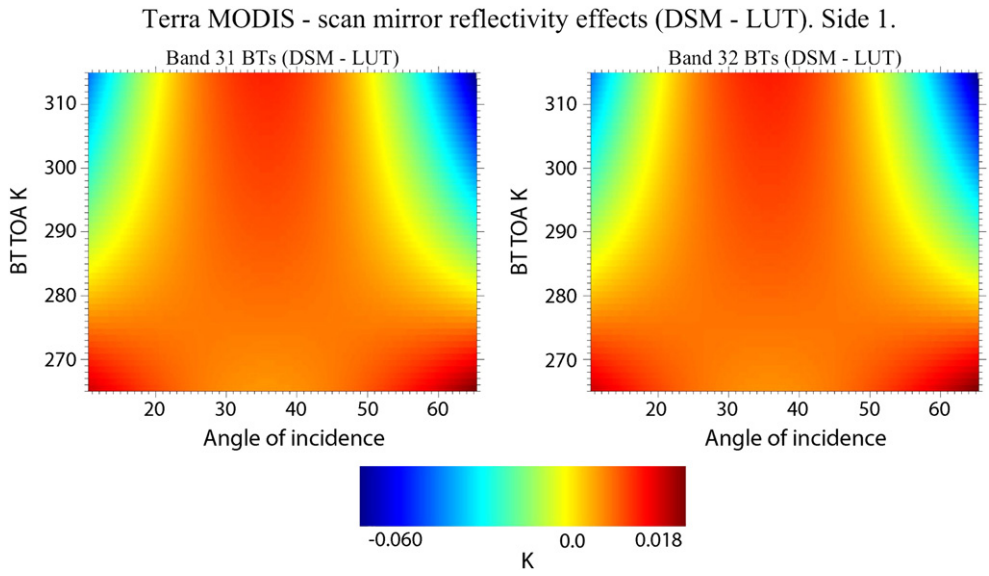

Fig. 4. Difference in magnitude of the Terra RVS correction developed from the closed earth-view door and used for Collection 3, and that from the Deep Space Maneuver, on March 26, 2003 used for Collections 4 and 5. The effects are greatest at the extreme angles of incidence on the scan mirror and for scene temperatures that depart from the mirror temperature (taken to be 275 K). The green to blue in the upper corners and the very dark red regions in lower corners of the image are where the absolute differences are >0.01 K. The values of the RVS correction are normalized to be unity at the angle of incidence of the on-board blackbody calibration target (26.3°). The LUT (Look-Up Table) values are those from the "Closed Door" measurements.

(i) angle of incidence on the scan mirror and (ii) the scene brightness temperature. The RVS correction, labeled LUT (Look-Up Table) in Fig. 4, is that derived from the closed-door measurements. The corrections are each normalized to an angle of incidence on the scan mirror of 26.3°. The figure is for mirror side 1; the corresponding figure for mirror side 2 (not shown) is similar, but not identical. The medium orange and yellow regions in each panel are for conditions where discrepancy is <abs (0.01 K). The effects of the improved RVS correction from the DSM are greatest at the more extreme scan angles, i.e., at the edges of the swath. The changes are also largest when the scene temperature deviates from the scan mirror temperature — when the two are the same, the RVS effects are zero. The plots are limited to the brightness temperature range typical of SST measurements and do not encompass the full dynamic range of the bands. The RVS correction based on the DSM was used to revise the TOA BTs for Collections 4 and 5

LWIR SST derived from Collection 5 BTs using the DSM RVS is shown in Fig. 5. This image of the Gulf of Mexico region is for the same week, beginning 20 March 2000, as the at-launch image in Fig. 2, and demonstrates the improvement in the cross scan behavior of the derived SST using the DSM RVS characterization.

A summary table is provided (Table 3) listing the type of RVS correction applied to the L1b data for each collection.

After RVS corrections derived from the deep space maneuver are applied to the L1b data, small residual uncorrected RVS artifacts still remain in the retrieved LWIR SSTs of ~0.1 K in the currently-available Collection 5 SSTs from both MODIS instruments (Fig. 6). Trends in the central tendency of both Terra and Aqua SST residuals are clearly apparent. For the LWIR SST on Terra, a pattern of linearly increasing residuals for satellite zenith angles (SZA) from −40° to 40° is observable, with steeper roll-offs at angles > 40°. An NLSST-type algorithm tends to produce a colder bias at higher scan angles in atmospheres of low to moderate water vapor (<35 kg m^2) (Merchant et al., 2009; Petrenko, Ignatov, Kihai, Stroup, & Dash, 2014). This is due to the combination of higher IR absorption along a longer atmospheric path and lower surface emissivity. However, the change in bias as a function of path length alone should be symmetrical about nadir, which would be the case if the RVS correction of the mirror reflectivity were correct. Since the patterns in these figures for the LWIR are not symmetrical, and clearly different on opposite sides of the scan line, whereas the atmospheric path-length is the same, we concluded that the RVS correction is still incomplete in Collection 5, and adds a systematic error to the retrievals up to 0.10 K

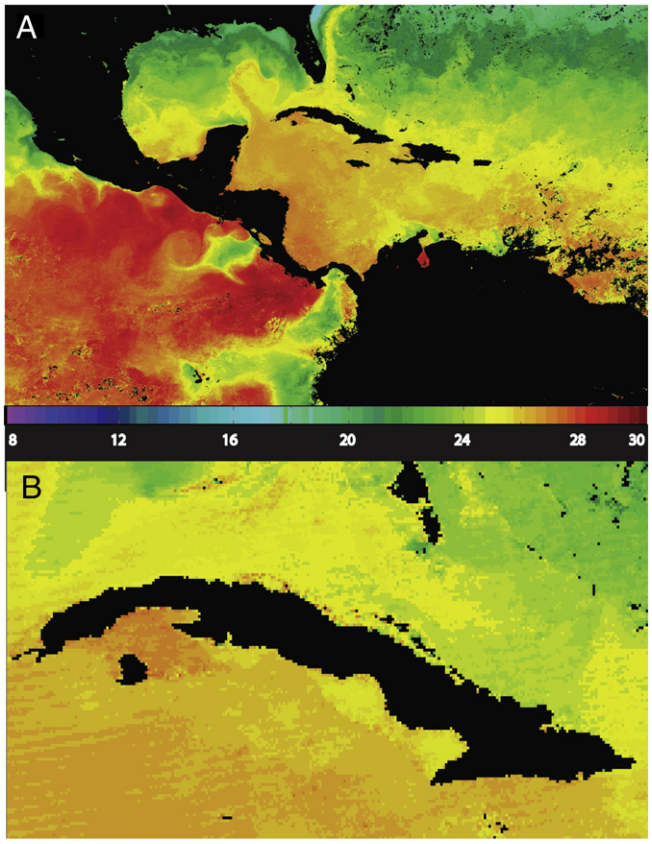

Fig. 5. Collection 5 Weekly image of LWIR SST (°C) of Gulf of Mexico region from the week of March 20th 2000 derived from L1b data after the RVS correction was improved from Deep Space Maneuver.

Table 3
Summary of the type of LWIR RVS correction applied for each released collection of sea surface temperature products.

Collection/release	Type of RVS characterization
Collections 1 & 2	Terra pre-launch witness samples combined with actual pre-launch Aqua mirror measurements
Collection 3	Terra closed door measurements
Collection 4	Terra deep-space maneuver
Collection 5	Terra-deep-space maneuver Aqua-pre-launch actual mirror measurements

across the scan between −45 and 45° SZA for pixels classified as being of the highest quality. The cold bias approaches −0.2 to −0.3 K relative to a skin temperature at SZA > 50°.

While the SST residuals in the LWIR show a systematic and asymmetric effect as a function of satellite zenith angle, the MWIR SST4 residuals for *Terra* and *Aqua* (Fig. 6 Panel B) are, as expected, relatively symmetrical about nadir, with a maximum excursion at intermediate angles (around ±40°). The ranges of fitted values in the MWIR from −45 to 45° SZA is relatively small, about 1/3 of the magnitude of that seen in the LWIR.

4.2. Mirror side reflectivity

Significant differences were found on orbit between the SST residuals for the two mirror sides on Terra MODIS, but not on Aqua (Fig. 7). The residuals associated with each of the two Terra mirror sides are different, and the magnitude of the difference between the mirror sides changes temporally, particularly in the first two years after Terra launch. Mirror side 1 has been relatively stable, whereas mirror side 2 appears to have degraded more quickly in the early years. Since 2002, the difference between the median of residuals from each mirror side has remained constant relative to each other.

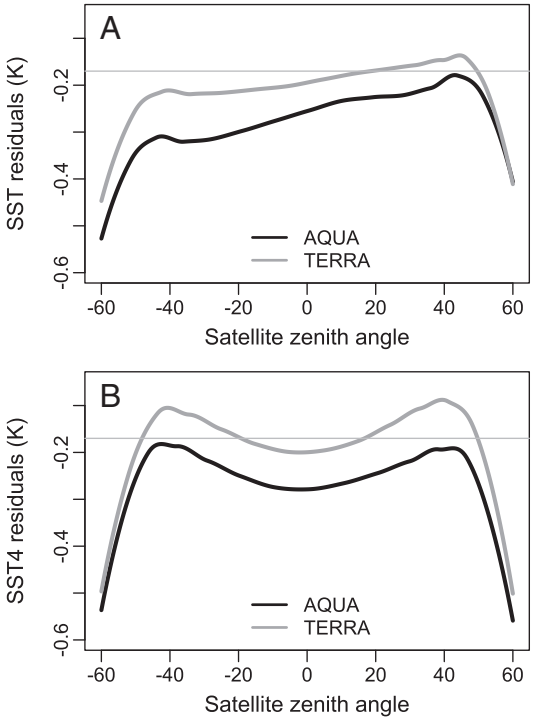

Fig. 6. Residual RVS trends in Collection 5 SST products Loess fits of (MODIS skin SST − subsurface buoy SST) from the matchup database as a function of satellite zenith angle. Aqua: Black Aqua line, Terra: Gray line after DSM RVS correction. Upper panel A: LWIR SST products; lower panel B: MWIR SST4. The thin horizontal line is the expected difference of −0.17 K caused by the thermal skin effect.

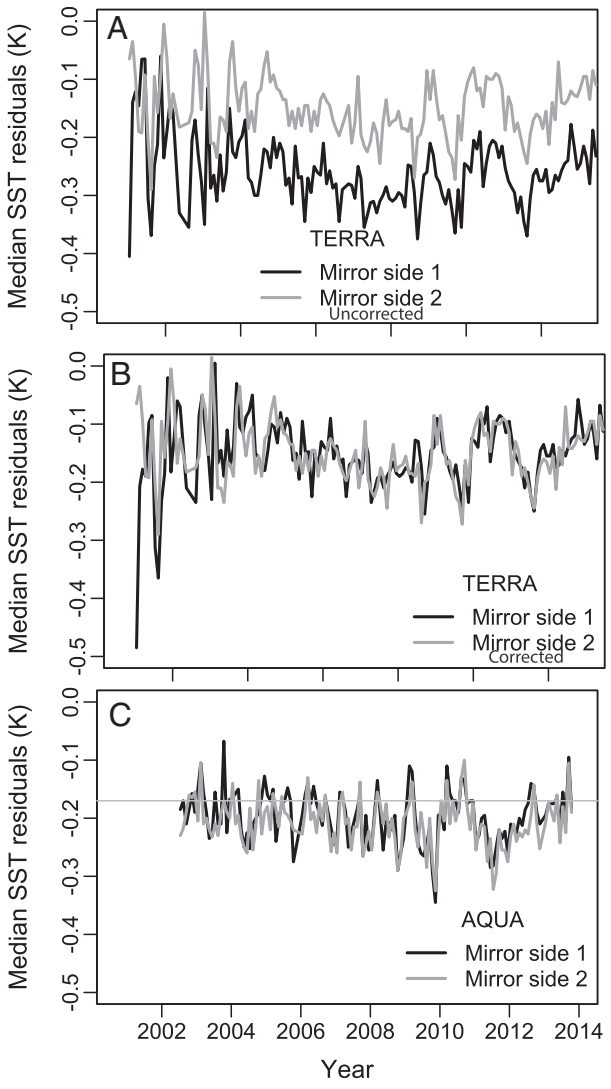

Fig. 7. Mirror side characterization time series of median residuals, MODIS-buoy SST for the LWIR SST. Black: mirror side 1, Gray: mirror side 2; Panel A: Terra LWIR uncorrected mirror side; Panel B: Terra Collection 5 after mirror side 2 normalization. Panel: Aqua LWIR. No correction is needed for AQUA.

The temporal correlation between SST residuals for both mirror sides is 0.748. Medians of SST residuals were −0.232 K for mirror side 1 and −0.124 K for mirror side 2, that is, a difference of ~0.112 K.

While being tested before launch, the *Terra* instrument suffered an incident that may have resulted in gaseous compounds of unknown type being deposited on one or both of the mirror sides. These compounds degassing on orbit may explain the variable degradation rate. It should be noted that during the early part of the *Terra* mission – prior to 2002 – the instrument also experienced several shutdowns, and upon restart the assignment of mirror side number would on occasion be uncertain. In contrast to *Terra*, the *Aqua* LWIR median mirror side differences are small, of the order of 0.01 K (Fig. 7 lower panel) and very stable in time; the MWIR SST4 for both sensors shows little difference between mirror sides (figure not shown).

As a result of the post-launch mirror characterizations, the Collection 4 and 5 atmospheric correction algorithms for *Terra* included an ad hoc correction to the LWIR SST values. The corrections sought to adjust measurements made with mirror side 2 to those of mirror side 1, based on the monthly median behavior of the global residuals from 2000 through 2002. Differences between mirror sides were estimated separately for each event that caused the instrument to be restarted. For *Terra*, variable time-dependent offsets between mirrors were applied for the first two

years. A constant offset was used past 2002 (the Look-Up Table of these mirror correction factors can be found at http://oceancolor.gsfc.nasa.gov/DOCS/modis_sst/). This mirror normalization decreased the dependencies of SST residuals on mirror side, but did not completely eliminate a mirror-side banding pattern in 1 km data. This banding, is visually randomly in some granules each day, and is apparent in groups of 10 scan lines corresponding to measurements from each of the 10 detectors.

4.3. Effects of multiple detectors

The calibration of each of the detectors in the IR MODIS bands is accomplished using an on-board blackbody calibration target and measurements of cold-space, away from the Sun. This approach is analogous to that used on the AVHRR, with the significant improvement that the black body is mounted within the cavity of the instrument and was designed to have high emissivity. The blackbody and space-view measurements provide a scan-by-scan, two-point calibration procedure to determine the gain and offset for each of the ten detectors in each band.

The bands used for SST estimation are located within MODIS on two separate focal plane assemblies (FPAs): bands 20, 22 and 23 are on the short-wave-infrared and mid-infrared (SWIR/MWIR) FPA, whereas bands 31 and 32 are on the long-wave-infrared (LWIR) FPA. The ten detectors for each band are aligned normal to the scan direction, so that each detector measures radiance from adjacent strips nearly orthogonal to the sub-satellite track. The black-body calibration target contains heaters, and episodically these are used to raise the temperature of the target, which permits on-orbit detection of changes in non-linear detector responses compared to those measured prior to launch. A more detailed description of the MODIS FPAs and blackbody is available elsewhere (Guenther, Xiong, Salomonson, Barnes, & Young, 2002; Xiong, Chiang, Esposito, Guenther, & Barnes, 2003; Xiong, Wenny, & Barnes, 2009). While each of the 10 detectors for each band is calibrated on-orbit separately, the noise inherent to the calibration procedure results in uncertainties in the offset and gain for each detector. Some detectors are much noisier than others; the consequences will be discussed in more detail in the following section. The uncertainty in both the calibration and the digitizer noise results in "stripes" along the scan direction, nearly orthogonal to the sub-satellite track in the images of retrieved SST (Figs. 2d and 5b), for both MODIS sensors. The stripes in SST fields are more pronounced than in the L1b BTs for each band. The increased striping is because the SST atmospheric correction algorithm involves terms with band differences, and therefore any noise or error in the individual bands is magnified in the derived SST. For the *Terra* mission, we developed empirical corrections to remove some of the detector striping by "balancing the detectors", for Aqua the inter-detector differences were smaller and no corrections are applied. The approach taken for the Terra inter-detector normalization procedure consisted of defining a "reference" detector, with which all other detectors would be made consistent. The reference detector chosen was the fifth in the series, as it was near the center of the detector array (though the fourth or sixth detectors would have been equally good choices).

A few hundred granules of 5-minute Earth-view data from cloud-free regions with small spatial gradients in SST and TOA BTs were selected as the normalization dataset. Although granules were selected for their low spatial variability, real geophysical pixel-to-pixel variability is still present. To perform the relative normalization, the geophysical variation in the BT fields needed to be further reduced to achieve the required accuracy. This was accomplished by comparing BT values between adjacent detectors. First, the mode of differences between the fifth detector and the adjacent fourth detector was computed and expressed as a percentage of the value at the reference detector. This step allowed the correction of the fourth detector relative to the fifth. Then, the mode of the differences between the fourth and third detectors was computed, allowing the relative correction of the third detector with respect to the fourth — and thus by association to the fifth. This procedure was repeated between all pairs of adjacent detectors. By summing the differences outward away from the fifth detector and applying the resulting offsets to each detector, all detectors are then in relative calibration to the fifth detector. Using this procedure causes the error in the estimate of the mode to accumulate with increasing distance from the fifth detector. However, this error was found to be less than the error introduced by the natural geophysical variability, as was inherent in a fifth to ninth detector difference, for example. These single offset values were computed for each detector and each band was applied equally across the entire scan. The BT normalization factors for each detector and band are shown in Table 4.

Examination of SST images derived from bands 31 and 32 of Terra MODIS in areas where the atmosphere appears to be clear and the SST to be relatively uniform, suggests that the detector-to-detector banding before normalization had an amplitude of approximately 0.3 K. Balancing the detectors at the TOA by the factors in Table 4 provides a qualitative improvement and reduces the gross striping in SST in some regions and viewing geometry, but does not eliminate the stripes completely (Fig. 5B). A number of alternative methods for statistical smoothing or noise reduction techniques have been presented over the years; some recent approaches for MODIS L1b (Gumley et al., 2009) and VIIRS (Bouali & Ignatov, 2014) have had some success. However, for MODIS Collections 3–5 none of the proposed algorithms had matured to the point of being implemented in NASA's L1b operational processing for SST bands 20, 22, 23, 31, and 32. Unfortunately a two-point, on-orbit calibration procedure designed to capture nonlinearity, is simply not sufficient to capture these small changes in detector responses from those measured in the pre-launch characterization (Guenther et al., 1996). The median error in SST that can be associated with the differences between detectors in Collection 5, again in select transects across the focal plane where the geophysical SST field appears to have minimal gradients, is of the order of ~0.1 ± 0.15. In some situations, this imbalance is still great enough to be visible to the eye, but not at all times and locations.

4.4. Impact of the electronics on SST retrievals

The MODIS instruments have two complete sets of redundant signal electronics — named Side-A and Side-B, one of which is active at any time. The duplicated electronics are intended to help ensure longevity of the instruments by providing redundancy. The primary electronics set is Side-A, which unfortunately turned out to have higher noise levels on orbit than Side-B for the Terra MODIS (Xiong et al., 2009). Terra MODIS began collecting data using Side-A electronics on 24 February 2000. As a result of a number of anomalous situations, data useful for scientific applications became available only on 30 October 2000, when the instrument was switched to Side-B electronics. The B-side electronics were used until 15 June 2001. Terra MODIS experienced a Side-B power supply shutdown anomaly and did not take data from 15 June 2001 to 2 July 2001. When Terra MODIS was restarted, it was commanded to take science data using the Side-A power supply but

Table 4
Percent change in L1b LWIR BTs to normalize the detectors across the focal plane to the detector 5.

Detector	Band 31	Band 32
1	0.0423	0.0433
2	0.0242	0.0246
3	0.0330	0.0222
4	0.0065	0.0148
5	0.0000	0.0000
6	0.0006	−0.068
7	−0.0064	−0.0221
8	−0.0398	−0.0221
9	−0.0362	−0.0437
10	−0.0322	−0.0303

Side-B electronics; this has been the Terra default configuration since. In contrast, the Side-B electronics and power supply have been used for the Aqua MODIS throughout its mission.

Since the two electronic sides are not identical, switches between Terra electronic configurations can potentially impact retrievals when month/year coefficients are not used, particularly on the MWIR product. For Terra, there are at least three "epochs" in the early years of the mission that can be traced to the electronic configuration in use (Fig. 8). The shifts in the median difference of residuals with respect to buoy temperatures are made particularly apparent in Fig. 8, because SST4 residuals were determined using coefficients derived for each month of the year, i.e., using matchups for a given month for all years (e.g., all Octobers from 2000 to 2013). The vertical lines in Fig. 8 indicate the dates of switches in the electronics configuration of Terra MODIS. These lines coincide clearly with large discontinuities in the median value of SST4 residuals. For Collections 3 through 5, the effects of change between electronic configurations are compensated by estimating coefficients separately for the first two epochs. However, it is important to remember that the start and the end months at each discontinuity may not have a complete 5 month window of matchups; after the last Terra configuration change in 2002 and through 2005, coefficients were estimated for a month/year combination.

Two further electronics issues prevented the MODIS SSTs from achieving the original expected on-orbit accuracy. These are noise in the digitization, and the radiometric dynamic range. For both MODIS sensors the detector outputs of either electronic side, are nominally digitized to a 12-bit resolution (i.e., values 0-4095), and should have been an improvement in resolution over the 10-bits for the AVHRRs. However, the least significant bit of the MODIS analog-to-digital converters was found on orbit to be unreliable, resulting in "non-uniform bin-fill" (Guenther et al., 2002). The effects of digitizer imperfections are manifest in high variability at the count-to-count level in histograms of the digital counts in the earth-view data (Fig. 9).

In Fig. 9 we see that noise is dependent on the detector where spikes/losses of counts are repeated for a given detector for both bands. In further analysis of the bit frequency for this granule we see that the frequency distribution of counts is detector dependent (Fig. 10). If the digitizer were operating properly we would expect histograms of digital counts to be smooth and the occupancy levels of the lower order bits to be 50%. Neither is the case. The histograms of bit occupancy show that, in general, all low order bits are below the expected 50% level. The problem is not simply a least significant bit issue (i.e., alternate bins being "narrow"), which would manifest itself with bit 0 showing low (or high) occupancy with all others showing 50%. The 50% level is not anticipated for high order bits as the environmental signal is not expected to fill the dynamic range of each detector uniformly. These results indicate problems up to the 4th or 5th bit. This type of plot is a sensitive indicator of systematic problems with the digitizers, but this effect cannot be corrected and is only shown here to explain why the expected benefit of 12-bit digitization was not realized. The values for bands 31 and 32 are comparable to those of the 10-bit digitized values from the corresponding channels of AVHRR. The histogram of the digital values in Fig. 9 shows most of the environmental signal being represented in a range of about 500 digital numbers out of a total range of 4096.

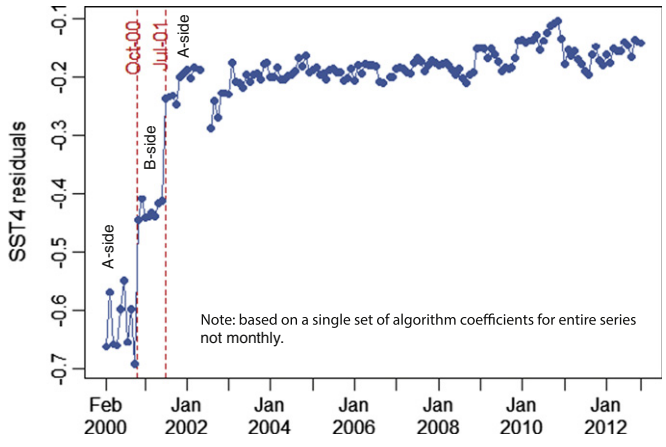

Fig. 8. MWIR SST4 residuals for different Terra electronic configurations prior to correction. Algorithm coefficients are not time-dependent for this plot. Abrupt changes in residual errors with respect to buoy time series are coincident with dates of changes in electronic configurations from A-side to B-side. Dashed red lines indicate the dates of switches between electronic configurations. Collections 3–5 used monthly coefficients derived separately for each of the electronic configuration time periods (epochs), and therefore the jump in median residuals at electronic transitions is compensated.

It is very difficult to quantify the consequences of this behavior on SST retrieval, as it is not simply a case of noise being imposed on every pixel. It is clear, however, that some pixel values are erroneously digitized, up to and including the 5th bit. Using the information provided by Jack Xiong (MCST) we note that the digitization increment at T = 290 K corresponds to a brightness temperature differences of ~0.026 K for the MWIR and 0.05 K for the LWIR bands on Terra. The brightness temperature differences are, of course, increased by 2, 4, 8, 16, … for increasing bit number. Both MODIS share the same digitizer design and so this issue is expected to be present on Aqua. However a detailed study was not done for Aqua as the problem cannot be fixed on orbit.

Another issue that contributes to MODIS SST not meeting the planned accuracy is the extended dynamic range of bands 31 and 32, saturating at ~400 K for *Terra* MODIS. This was a result of a compromise whereby the requirements for measuring the temperature of forest fires was folded into the requirements to measure SST. Originally it was planned to separate these two needs using dual-gain detectors, but the change to a single-gain band resulted in about a factor of two loss in the radiometric resolution of the digital signal at BTs typical of the SST. As a result the anticipated improvement in the radiometric resolution of the LWIR bands was not achieved in the *Terra* MODIS. The saturation temperature for bands 31 and 32 in the *Aqua* MODIS was reset to 340 K, which is still higher than needed for SST retrievals, but accommodates the requirements of the land-surface community in many situations.

5. Validation of collection 5 SST products

Validation requires knowledge of the total error budget in SST retrievals. This section is intended to provide error statistics of the retrievals by comparisons to both in situ sub-surface temperature and skin SST measurements from buoys and ship board radiometers, respectively. The in situ measurements used for validation were independent of the training data set for coefficient estimation. The validation statistics we present include the contributions from four potential error sources: (i) atmospheric correction, (ii) MODIS instrumental artifacts, (iii) in situ measurements, and (iv) the observation time and space collocation window between the satellite and in situ observations in the MDB.

We first provide aggregate global statistics for the entire mission. Aggregate statistics may not always give a true sense of SST retrieval accuracy at different times and places. For example, the signs and magnitudes of the error can be different in geographic space or time and may cancel one another, producing an overly optimistic aggregated assessment. Time series for different latitude bands allow us to evaluate the error in different geographic regions and atmospheric regimes.

Aggregate global statistics for the currently available MODIS SST Collection 5 products relative to sub-surface in situ buoy temperatures are presented in Table 5. SST statistics relative to the M-AERI skin measurements are shown in Table 6. All statistics presented are for night-time best quality pixels. For the MODIS, with a maximum viewing

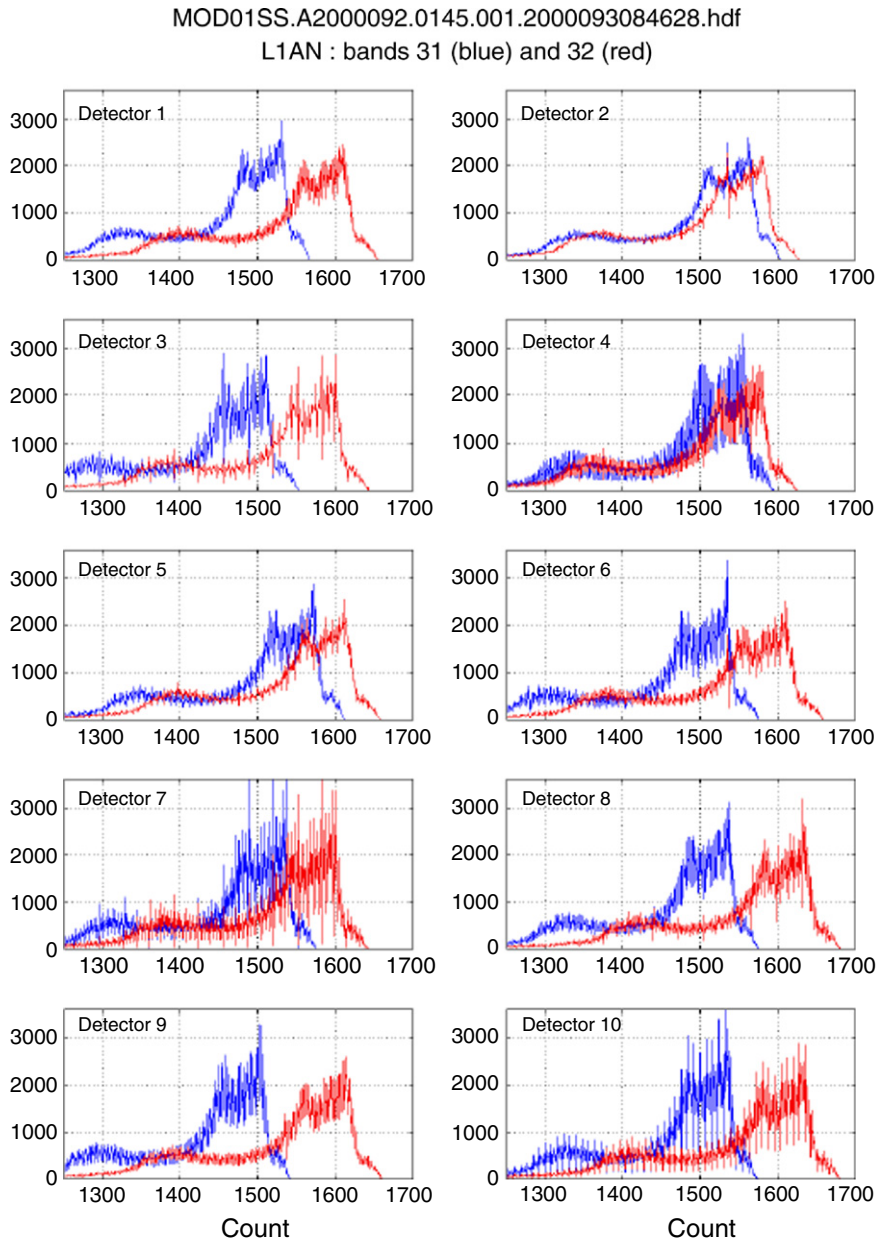

Fig. 9. Detector noise and digitizer reliability histograms of digital counts in each of the 10 detectors for bands 31 and 32 for a single Terra L1a earth view granule. These histograms indicate that the least significant of the 12 bits is unreliable and certain detectors have increased noise, for example detector 4. Blue: band 31, red: band 32.

geometry of 60°, these are pixels determined to be both cloud free and are at scan angles < 55°. NASA produced SST products available from the Ocean Biology Processing Group (OBPG) use the heritage NASA EOS quality scale of 0–4, where best = 0, bad = 3 and not processed = 4. These products are also available as part of the international Group High Resolution Sea Surface Temperature project (GHRSST) (Donlon et al., 2007). For GHRSST L2p files the NASA EOS scale is inverted to conform to the newer consensus GHRSST definition where best = 5.

The global median night time difference relative to buoys is very close to the theoretical sub-surface skin difference of −0.17 K (Donlon et al., 2002; Kearns et al., 2000), and <0.1 K relative to MAERI as required by many SST users. The presence of long tails in the distribution of SST residuals, as a result of outliers tied to unidentified cloud contamination or anomalous atmospheres (Szczodrak, Minnett, & Evans, 2014), can influence the common metrics of central tendency and dispersion i.e. the mean and standard deviation. For this reason we provide more resistant metrics (also called robust) of the median and robust standard deviation (interquartile range (IQR) divided by 1.348) as used by Merchant & Harris (1999). Standard deviations are in the range of 0.4 to 0.5 K, but the more robust estimate of the standard deviation of night residuals is in the range of 0.2 to 0.4 K, for Terra and Aqua SST4 and SST respectively. These numbers also include contributions from uncertainties in the validating measurements and those introduced by the method of comparison. Since the launch of Terra, the M-AERIs have taken measurements over the latitude range from the ice-edges in the Arctic and Antarctic, and in all of the worlds oceans (see Figure 36 of Donlon et al., 2014). Of course, the sampling is not uniform, with a preponderance of measurements taken in the low- to mid-latitudes of the Atlantic Ocean. And while a spatially and seasonally uniform sampling regime would be highly desirable, it is not feasible to achieve. The consequences of non-uniform sampling on the error estimates of errors and uncertainties in satellite-derived SSTs are the subject of a current study. The global estimates of the MODIS

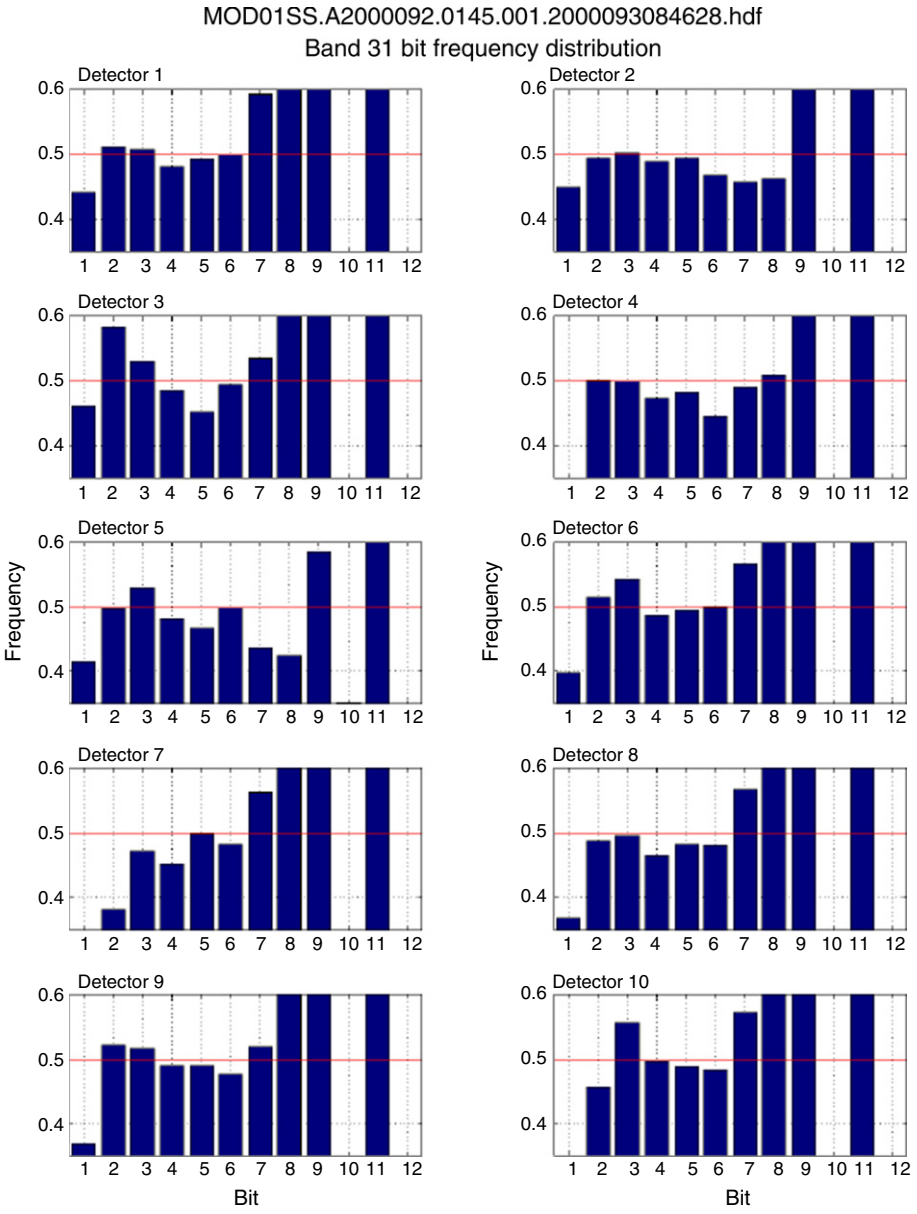

Fig. 10. Frequency of bit occupancy for the 10 detectors for Band 32 from 15 granules of data.

uncertainties are an improvement over heritage IR products which are in the 0.5 to 0.7 K for Pathfinder SST, depending on the specific AVHRR sensor considered and its age (Kilpatrick, 2013). The MODIS statistics reported here are also similar to those reported previously (Minnett, 2010) for 1 km MODIS data and more recently by Gentemann (2014) for MODIS Aqua. Nevertheless, time series of retrieval bias relative to buoys for both MODIS sensors at times suggest a repeating seasonal pattern in some latitude bands, particularly at high latitudes (Fig. 11). Gentemann (2014) used a triple collocation validation using buoys, NASA MODIS products, and the microwave

Table 5
Collection 5 — Night-time LWIR skin SST and MWIR SST4 error characterization relative to in situ buoy subsurface temperatures for NASA EOS quality level 0 (best) and quality 1 (degraded). Residual MODIS skin minus buoy temperatures (K), as mean, median, standard deviation (SD), robust standard deviation (RSD = interquartile range/1.386 Merchant & Harris, 1999), and count.

	Product	Mean	Median	SD	RSD	Count
Terra	SST Q0	−0.191	−0.145	0.510	0.402	140,436
	SST Q1	−0.602	−0.465	0.94	0.623	71,314
Aqua	SST Q0	−0.260	−0.205	0.509	0.359	113,197
	SST Q1	−0.649	−0.75	0.97	0.616	63,697
Terra	SST4 Q0	−0.095	−0.155	0.422	0.245	148,700
	SST4 Q1	−0.592	−0.400	0.833	0.476	90,744
Aqua	SST4 Q0	−0.236	−0.180	0.423	0.239	112,758
	SST4 Q1	−0.659	−0.455	0.839	0.478	85,662

Table 6
Collection 5 — Night-time LWIR SST and MWIR SST4 error characterization relative to skin SST from M-AERI. Residual statistics of satellite minus M-AERI skin temperatures (K) for NASA EOS quality level 0 (best) and quality 1 (degraded) for mean, median, standard deviation (SD), and robust standard deviation (RSD = interquartile range/1.346 Merchant & Harris, 1999), count.

	Product	Mean	Median	SD	RSD	Count
Terra	SST Q0	0.014	−0.030	0.505	0.430	2538
	SST Q1	−0.370	−0.33	0.746	0.667	1613
Aqua	SST Q0	−0.046	−0.027	0.503	0.434	1183
	SST Q1	−0.418	−0.344	0.709	0.640	776
Terra	SST4 Q0	−0.024	−0.020	0.401	0.289	2782
	SST4 Q1	−0.415	−0.36	0.616	0.461	1846
Aqua	SST4 Q0	−0.067	−0.029	0.436	0.330	1314
	SST4 Q1	−0.488	−0.439	0.634	0.522	1012

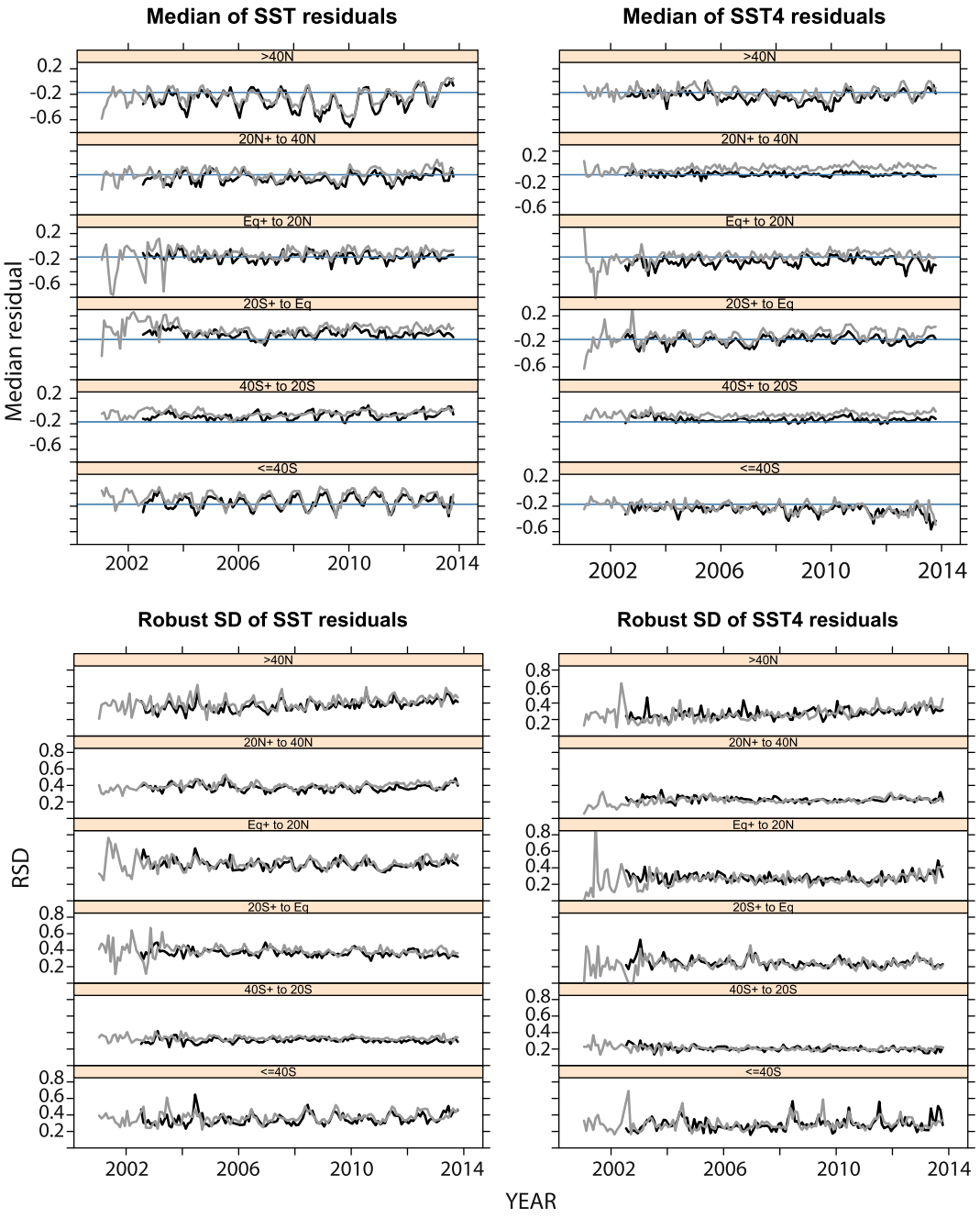

Fig. 11. Time series plots of SST and SST4 error characteristics relative to in situ buoys for each latitude band. Gray line: Terra; Black line: Aqua. Top row: monthly median MODIS SST and SST4–buoy SST, Bottom row: monthly robust standard deviation SST and SST4 residuals. The LWIR SST median differences have a strong seasonal signal at high latitudes and are phase shifted between the Northern and Southern Hemispheres.

radiometer AMSR-E SSTs both instruments are aboard Aqua, and found regional problems in both AMSR-E and MODIS SST data products. A cool bias was found in the analysis of 25 km averages of the night MODIS Aqua SST at high latitudes. This cool latitudinal bias was hypothesized to be the result of both less efficient cloud flagging in the MODIS product at cold temperatures, and the sampling error that would be introduced into the analysis by the presence of clouds in the 25 km average used during the collocation to the AMSR-E resolution. Small regional and seasonal trends in SST buoy residuals are not unique to MODIS and have been reported for other SST datasets using an NLSST formulation, for example AVHRR (Kilpatrick et al., 2001). These trends are likely due to the fact that the NLSST is based on a global atmosphere averaged in space and time, while in reality the local atmosphere can have both systematic and random deviations from a global average.

Based on the results of a simple bootstrap sampling experiment, we determined that the stability and confidence in the statistical results for months and latitude bands where the number of available buoys is <100 should be interpreted with caution (Fig. 12). Moreover, the 2000–2002 period – for which there are much fewer matchups – also involved numerous instrument shutdowns and changes in the electronic configurations of *Terra* MODIS, making it even more difficult to accurately evaluate the SST uncertainty. After 2004, the buoy network had broader geographic coverage and the *Terra* instrument was generally more stable. These two factors resulted in more reliable validation statistics.

To arrive at this threshold of a minimum of 100 matchups, first we selected 534 records from the Terra MDB to serve as a reference population of matchups on which to evaluate the impact of sample

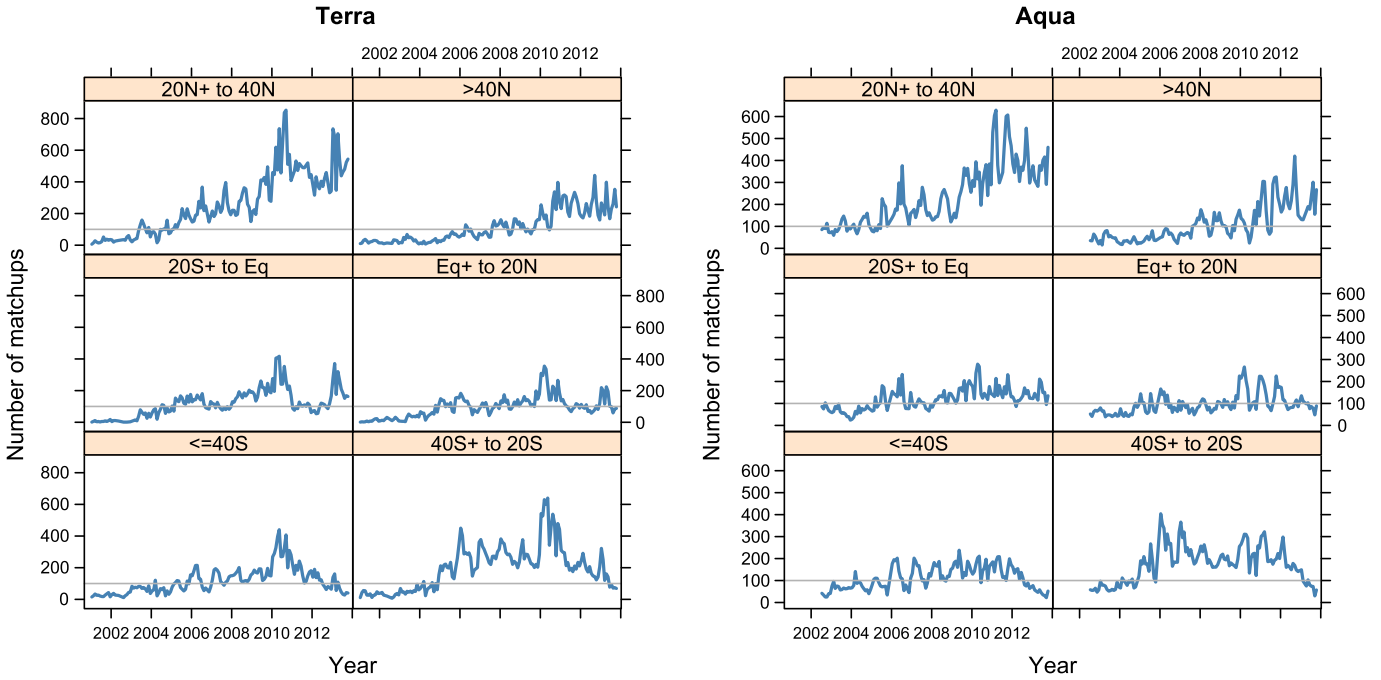

Fig. 12. Monthly counts of buoy matchup observations in each latitude band for each sensor. Left: Terra, Right: Aqua. Gray horizontal line in each panel is at 100 counts. Error characterization may not be stable or statistically significant for months in latitude bands where the count is less than 100 observations.

size on the validity residual statistics. The benchmark records where chosen from a sub-population of all the matchups with characteristics identified as often being associated with higher standard deviation of SST residuals. Using a reference population with higher variability should produce a "pessimistic" estimate of the minimum number of matchups required to accurately characterize the retrievals. The selected benchmark population included matchups between 20 and 40°N, for the boreal spring (April–June), showing relatively warm buoy SSTs between 23 and 29 °C, and had differences in brightness temperatures of bands 31 and 32 of 0.7 to 2 K. The subsurface buoy temperatures in the benchmark pool were corrected to be pseudo skin temperatures, so satellite SST residuals would be centered about zero (mean = 0.00552 K, SD = 0.4322 K). Next, we sampled the benchmark population at 50 different sample sizes (10, 20, 30… 490, 500); for each sample size, 500 samples of that size were taken with replacement and the median SST residual was calculated for each sample. The distribution of median residuals for the 500 samples taken for each sample size is shown in Fig. 13. The black bars in each box represent the grand median of individual medians for a sample size. The width of the box represents the quartiles of these values (i.e., the boxes encompass the central 50% of simulated sample median of residuals). Outliers are shown as open circles. The two orange lines are the 95% confidence intervals of the benchmark population and the red line is the mean. These statistics were estimated by taking 10,000 bootstrapped samples of size 534 from the benchmark population. At a sample size of ~100 matchups the boxes for each sample size begin to narrow and are encompassed within the 95% confidence interval of the original benchmark population.

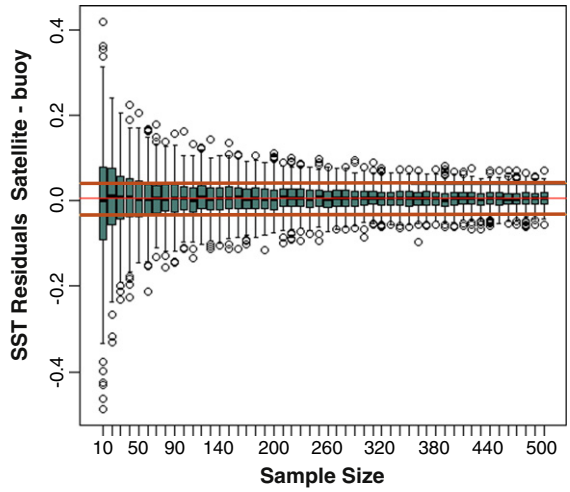

Fig. 13. Impact of the number of available matchups on the reliability of statistical results. The box and whiskers represent the distributions of medians for all samples of a given size. The black line in each box is the median of all samples of a given size. The red line is the median of the benchmark population (N = 534) from which all samples of all sizes were generated. The pair of orange lines is the boundaries of the 95% confidence interval for the benchmark population.

6. Summary and discussion

The MODIS instruments, the first IR radiometers launched by the U.S. with spectral bands specifically designed to measure SST, have been providing accurate daily global measurements of skin SST at a 1 km resolution for more than a decade. The suite of MODIS SST estimates has been reprocessed three times to improve the accuracy and precision of the products as our understanding of the sensor behavior, the atmospheric correction, and the characteristics of SST algorithms improve. Here, we have presented the evolution of SST products from the first publicly available version, Collection 3, to the current Collection 5. We have reviewed the methodology used to characterize and correct for instrument artifacts found on orbit stemming from: RVS, mirror side, and heterogeneity in detector output. This paper provides statistics on the accuracy of the MODIS Collection 5 skin SSTs relative to subsurface temperatures from the global network of in situ buoys, and to skin temperatures from M-AERI radiometers mounted on ships.

The Collection 5 nighttime LWIR skin SST products compared to subsurface buoys, presented in Tables 5 and 6, are nearly identical for both MODIS sensors, for *T*erra the median residual is −0.145 K, standard deviation (SD) is 0.510 K, and robust deviation (RSD) is 0.380 K. For

Aqua the median residual is −0.205 K, SD is 0.509 K, and RSD is 0.0.388 K. The mean differences are close to what we expect to result from the thermal skin effect. These values are similar to those reported for other LWIR IR SST satellite data products: AVHRR V5.3 Pathfinder (Kilpatrick, 2013), and MetOp AVHRR (O'Carroll et al., 2012). Both MODIS instruments have a near zero SST median difference, 0.030 K and −0.028 K, when compared to the highly accurate but more limited numbers of skin temperatures from M-AERI radiometers.

The MWIR MODIS SST4 product is superior to the LWIR SST, since the atmosphere is more transparent and less variable in this spectral domain, with a lower uncertainty: SD of 0.42 K and RSD of 0.24 K. Time series of LWIR Collection 5 and regional analyses indicate that while global aggregate statistics are good, seasonal patterns in the differences with respective to buoy temperatures are evident in some regions, particularly at high latitudes (Fig. 11). Regional seasonal signatures in LWIR SSTs have been previously reported for AVHRR (Gentemann, 2014; Kilpatrick et al., 2001; O'Carroll et al., 2012) and likely result from spatial and temporal variability in water vapor content of the atmosphere. In the late 1990s beginning with V4 of the NOAA/NASA Pathfinder SST project, the standard NLSST algorithm, which previously used a single set of coefficients, was modified to use two sets of coefficients based on the brightness temperature difference between the 11 and 12 μm bands. This change aimed to separate the atmospheric correction into "wet" and "dry" atmospheres. While this technique generally reduced the seasonal signature on a global basis, it was not as successful in removing the seasonal signal at high latitudes. This high latitude artifact may be partially due to the locations of buoy observations used to train coefficients, being heavily concentrated in the mid-Atlantic. North Atlantic mid-latitude atmospheres and the seasonality of Northern hemisphere then dominate the derivation of coefficients for the global atmospheric correction. Furthermore at high latitudes cloud screening is likely less efficient at the cold temperatures, due to the reliance on uniformity indexes to discern cold clouds from clear surface views.

The MODIS design involved several innovative elements. While some of these innovations introduced artifacts of unexpected severity in the initial data, the artifacts have been reduced to a large extent in subsequent generations of SST products, to the point of making current SST retrievals useful for a wide range of applications.

Of prime importance are the multiple spectral bands in the MWIR atmospheric transmission window, which permit retrieval of night-time skin SSTs with better accuracy than possible with the heritage bands. The dual-sided, paddle-wheel scene mirror and using multiple detectors for each spectral band, although the source of problems in the early part of the Terra MODIS mission, permitted longer integration times for each pixel, that resulted in narrower spectral response functions for each spectral band than in the heritage sensor and in the subsequent VIIRS instrument; this in turn led to more effective corrections for the effects of the intervening atmosphere. The mirror design removed the need for a "half-angle-mirror" that is required to compensate for "image rotation" across the swath that occurs when multiple detectors are used with conventional scan mirrors or rotating telescopes, as used in SeaWiFS and VIIRS (Schueler et al., 2002). The increased number of bits in the digital data stream was supposed to provide finer radiometric resolution, which unfortunately was not fully realized.

The multi-spectral nature of MODIS supplies useful data for a very wide user community, supporting research into many components of the earth's climate system. Indeed the number of papers in the reviewed literature using MODIS data now exceeds those based on the Hubble Space Telescope (V. Salomonson, MODIS Science Team Meeting, Columbia, MD, April, 2014). Regarding SST, MODIS SST products are the most requested and downloaded from the GHRSST (Group for High Resolution SST); (Donlon et al., 2007) GDAC (Global Data Assembly Center) at the NASA JPL PO.DAAC (E. Armstrong, pers. comm., 2013).

7. Future directions

With the improved on orbit characterizations and bi-weekly updates to the instrument calibration producing stable calibration and accuracy of L1b fields, we have had success in reducing the seasonal trends at high latitudes by use of coefficients derived as a function of latitude band, rather than on a threshold of BT of difference as a proxy for high and low water vapor. The available in situ data at high latitudes is of course more limited, particularly in winter months and especially during the early years of the *Terra* mission (Fig. 12). To derive monthly coefficients for separate latitude bands requires that the accuracy of the L1b be stable from year to year so matchups from different years can be aggregated during coefficient estimation. Continuous assessment of sensor data and SST estimates suggests that the instruments are stable enough year-to-year to use latitude based coefficients and will use a latitude band approach to coefficient estimation in the next MODIS reprocessing (Collection 6).

Currently the highest quality retrievals from MODIS are limited to scan angles < 55° due to a very rapid non-linear change in the top-of-atmosphere brightness temperature deficit at scan angles > 55°, global statistics can rapidly exceed a bias of −0.5° as angles greater than 60 and have much greater dispersion (Fig. 6). The magnitude of the cold bias at these higher zenith angles is not a simple geometrical function of the path length but actually reflects the overall atmospheric attenuation along the path. Recent studies for the VIIRS sensor, which has a wider swath width and 16 detectors, suggests that quality retrievals at zenith angle > 55° may be possible in geographic domains known to have typically lower atmospheric attenuation (Petrenko et al., 2014). While MODIS is perceived to be the heritage sensor for VIIRS, it is clear that the analysis of VIIRS data can provide useful insights to improve MODIS SSTs, including the possibility of developing a high-accuracy full-swath algorithm for MODIS and improved de-striping methods.

7.1. Data availability

MODIS SSTs are produced and made available to the public by the NASA GFSC Ocean Biology Processing Group (OBPG), and available in near real time in the GHRSST L2P format from the GHRSST GDAC. MODIS SSTs are also archived in perpetuity at the GHRSST Long Term Stewardship and Reprocessing Facility at NOAA National Oceanographic Data Center (LTSRF) for retrospective data.

Acknowledgments

The authors acknowledge the grants from NASA for supporting these studies over the last decade and more: NX11AF26G, NNX11AF03G, NNG04HZ32C, and NNG04HZ33C. We would also like to thank the reviewers for useful comments and suggestions that have helped clarify the text.

References

Barnes, W.L., Pagano, T.S., & Salomonson, V.V. (1998). Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1088–1110.

Barnes, W.L., & Salomonson, V.V. (1993). Modis: A global imaging spectroradiometer for the earth observing system. critical reviews of optical. *Science and Technology*, *CR47*, 285–307.

Barton, I.J. (1995). Satellite-derived sea surface temperatures: Current status. *Journal of Geophysical Research*, *100*, 8777–8790.

Barton, I.J., Prata, A.J., & Llewellyn-Jones, D.T. (1993). The along track scanning radiometer — An analysis of coincident ship and satellite measurements. *Advances in Space Research*, *13*, 69–74.

Bouali, M., & Ignatov, A. (2014). Adaptive reduction of striping for improved sea surface temperature imagery from Suomi National Polar-Orbiting Partnership (S-NPP) visible infrared imaging radiometer suite (VIIRS). *Journal of Atmospheric and Oceanic Technology*, *31*, 150–163.

Brown, O.B., Brown, J.W., & Evans, R.H. (1985). Calibration of advanced very high resolution radiometer infrared observations. *Journal of Geophysical Research*, *90*, 11667–11677.

Donlon, C.J., Minnett, P.J., Gentemann, C., & Nightingale, T.J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*, 353–369.

Donlon, C.J., Minnett, P.J., Jessup, A., Barton, I., Emery, W., Hook, S., et al. (2014). Chapter 3. 2 — Ship-borne thermal infrared radiometer systems. In C.J.D. Giuseppe Zibordi, & C.P. Albert (Eds.), *Experimental methods in the physical sciences* (pp. 305–404). Academic Press.

Donlon, C.J., Robinson, I., Casey, K.S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., et al. (2007). The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society*, *88*, 1197–1213.

Donlon, C.J., Robinson, I., Reynolds, M., Wimmer, W., Fisher, G., Edwards, R., et al. (2008). An infrared sea surface temperature autonomous radiometer (ISAR) for deployment aboard volunteer observing ships (VOS). *Journal of Atmospheric & Oceanic Technology*, *25*, 93–113.

Edwards, T., Browning, R., Delderfield, J., Lee, D.J., Lidiard, K.A., Milborrow, R.S., et al. (1990). The along track scanning radiometer — Measurement of sea-surface temperature from ERS-1. *Journal of the British Interplanetary Society*, *43*, 160–190.

Emery, W.J., Sandra, C., Wick, G.A., Peter, S., & Craig, D. (2001). Estimating sea surface temperature from infrared satellite and in situ temperature data. *Bulletin of the American Meteorological Society*, *82*, 2773–2785.

Esaias, W.E., Abbott, M.R., Barton, I., Brown, O.B., Campbell, J.W., Carder, K.L., et al. (1998). An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1250–1265.

Gentemann, C.L. (2014). Three way validation of MODIS and AMSR-E sea surface temperatures. *Journal of Geophysical Research: Oceans*, *119*, 2583–2598.

Gentemann, C.L., Minnett, P.J., Leborgne, P., & Merchant, C.J. (2008). Multi-satellite measurements of large diurnal warming events. *Geophysical Research Letters*, *35*, L22602.

Guenther, B., Barnes, W., Knight, E., Barker, J., Harnden, J., Weber, R., et al. (1996). MODIS calibration: A brief review of the strategy for the at-launch calibration approach. *Journal of Atmospheric and Ocean Science*, *13*, 274–285.

Guenther, B., Xiong, X., Salomonson, V.V., Barnes, W.L., & Young, J. (2002). On-orbit performance of the earth observing system moderate resolution imaging spectroradiometer; first year of data. *Remote Sensing of Environment*, *83*, 16–30.

Gumley, L., Strabala, K., & Menzel, W.P. (2009). Improved destriping for Terra and Aqua MODIS data: Algorithm description and quantitative radiometric assessment. *IGARRS, Proceedings July 12–17 South Africa*.

Harries, J.E., Llewellyn-Jones, D.T., Minnett, P.J., Saunders, R.W., Zavody, A.M., Wadhams, P., et al. (1983). Observations of sea-surface temperature for climate research [and discussion]. *Philosophical Transactions of the Royal Society of London. Series A Mathematical and Physical Sciences*, *309*, 381–395.

Kearns, E.J., Hanafin, J.A., Evans, R.H., Minnett, P.J., & Brown, O.B. (2000). An independent assessment of pathfinder AVHRR sea surface temperature accuracy using the marine atmosphere emitted radiance interferometer (MAERI). *Bulletin of the American Meteorological Society*, *81*, 1525–1536.

Kilpatrick, K.A. (2013). *Climate algorithm theoretical basis document (C-ATBD) pathfinder sea surface temperature.* NOAA CDR Program Document, CDRP-ATBD-0099.

Kilpatrick, K.A., Podestá, G.P., & Evans, R.H. (2001). Overview of the NOAA/NASA pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, *106*, 9179–9198.

Llewellyn-Jones, D.T., Minnett, P.J., Saunders, R.W., & Zavody, A.M. (1984). Satellite multichannel infrared measurements of sea surface temperature of the N.E. Atlantic Ocean using AVHRR/2. *Quarterly Journal of the Royal Meteorological Society*, *110*, 613–631.

McClain, E.P., Pichel, W.G., Walton, C.C., Ahmad, Z., & Sutton, J. (1983). Multi-channel improvements satellite derived global sea surface temperatures. *Advances in Space Research*, *2*, 43–47.

Merchant, C. J., & Harris, A. R. (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature: 2. Comparison with in situ measurements. *Journal of Geophysical Research: Oceans*, *C10*, 23579–23590.

Merchant, C.J., Harris, A.R., Roquet, H., & Le Borgne, P. (2009). Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer. *Geophysical Research Letters*, *36*, L17604.

Minnett, P.J. (1986). A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space. *Journal of Geophysical Research*, *91*, 8509–8521.

Minnett, P.J. (1991). Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *Journal of Geophysical Research*, *96*, 18475–18489.

Minnett, P.J. (1995). The along-track scanning radiometer: Instrument details. In M. Ikeda, & F. Dobson (Eds.), *Oceanographic applications of remote sensing* (pp. 461–472). Boca Raton, Florida, USA: CRC Press Inc.

Minnett, P.J. (2003). Radiometric measurements of the sea-surface skin temperature: The competing roles of the diurnal thermocline and the cool skin. *International Journal of Remote Sensing*, *24*, 5033–5047.

Minnett, P.J. (2010). The validation of sea surface temperature retrievals from spaceborne infrared radiometers. In V. Barale (Eds.), *Oceanography From Space, Revisited* (pp. 273–295). Heidelberg, Germany: Springer.

Minnett, P.J., & Barton, I.J. (2010). Remote sensing of the earth's surface temperature. In Z.M. Zhang, B.K. Tsai, & G. Machin (Eds.), *Radiometric temperature measurements and applications* (pp. 333–391). Academic Press/Elsevier.

Minnett, P.J., Knuteson, R.O., Best, F.A., & Osborne, B.J. (2001). The marine-atmospheric emitted radiance interferometer: A high-accuracy, seagoing infrared spectroradiometer. *Journal of Atmospheric and Oceanic Technology*, *18*, 994–1013.

Minnett, P.J., & Ward, B. (2000). Measurements of near-surface ocean temperature variability — Consequences on the validation of AATSR on ENVISAT. *ERS-ENVISAT Symposium "Looking Down to Earth in the New Millennium"*. (pp. 10PP)Gothenburg, Sweden: European Space Agency.

Murray, M.J., Allen, M.R., Merchant, C.J., Harris, A.R., & Donlon, C.J. (2000). Direct observations of skin-bulk SST variability. *Geophysical Research Letters*, *27*, 1171–1174.

O'Carroll, A.G., August, T., Le Borgne, P., & Marsouin, A. (2012). The accuracy of SST retrievals from METOP-A IASI and AVHRR using the EUMETSAT OSI-SAF matchup dataset. *Remote Sensing of Environment*, *126*, 184–194.

Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., & Dash, P. (2014). Evaluation and selection of SST regression algorithms for JPSS VIIRS. *Journal of Geophysical Research*, *119*, 4580–4599.

Reynolds, R.W. (1993). Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperatures. *Journal of Climate*, *6*, 768–774.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, *15*, 1609–1625.

Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., & Schlax, M.G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*, 5473–5496.

Rothman, L.S., Gamache, R.R., Goldman, A., Brown, L.R., Toth, R.A., Pickett, H.M., et al. (1987). The Hitran database: 1986 edition. *Applied Optics*, *26*, 4058–4097.

Salomonson, V.V., Barnes, W.L., Maymon, P.W., Montgomery, H.E., & Ostrow, H. (1989). MODIS: Advanced facility instrument for studies of the earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, *27*, 145–153.

Schueler, C.F., Clement, J.E., Ardanuy, P.E., Welsch, C., Deluccia, F., & Swenson, H. (2002). NPOESS VIIRS sensor design overview. *International Symposium on Optical Science and Technology. International Society for Optics and Photonics*. (pp. 11–23).

Szczodrak, M., Minnett, P.J., & Evans, R.H. (2014). The effects of anomalous atmospheres on the accuracy of infrared sea-surface temperature retrievals: Dry air layer intrusions over the tropical ocean. *Remote Sensing of Environment*, *140*, 450–465.

Trishchenko, A.P., Fedosejevs, G., Li, Z., & Cihlar, J. (2002). Trends and uncertainties in thermal calibration of AVHRR radiometers onboard NOAA-9 TO NOAA-16. *Journal of Geophysical Research: Atmospheres*, *107*, 4778–4791.

Walton, C.C., McClain, E.P., & Sapper, J.F. (1990). Recent changes in satellite-based multichannel sea surface temperature algorithms. *J Marine Tech. Soc. Proceedings, Washington, D.C. Convention Center, September 26–28, 1990*, *90*, .

Walton, C.C., Pichel, W.G., Sapper, J.F., & May, D.A. (1998). The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research: Oceans*, *103*, 27999–28012.

Ward, B. (2006). Near-surface ocean temperature. *Journal of Geophysical Research*, *111*, C02005.

Xiong, X., Chiang, K., Esposito, J., Guenther, B., & Barnes, W. (2003). MODIS on-orbit calibration and characterization. *Metrologia*, *40*, S89–S92.

Xiong, X., Wenny, B. N., & Barnes, W. L. (2009). Overview of NASA earth observing systems Terra and Aqua moderate resolution imaging spectroradiometer instrument calibration algorithms and on-orbit performance. *Journal of Applied Remote Sensing*, *3*, 69–79.

Xu, F., & Ignatov, A. (2013). IN situ SST quality monitor (IQUAM). *Journal of Atmospheric and Oceanic Technology*, *31*, 164–180.

Závody, A.M., Mutlow, C.T., & Llewellyn-Jones, D.T. (1995). A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. *Journal of Geophysical Research: Oceans*, *100*, 937–952.