The SST Quality Monitor (SQUAM)

Prasanjit Dash
NOAA NESDIS, Center for Satellite Applications and Research (STAR), 5200 Auth Road, Camp Springs, Maryland 20746, USA

Cooperative Institute for Research in the Atmospheres (CIRA), Colorado State University, Foothills Campus, Fort Collins, Colorado 80523, USA

Alexander Ignatov
NOAA NESDIS STAR, 5200 Auth Road, Camp Springs, Maryland 20746, USA

Yury Kihai
NOAA NESDIS STAR, 5200 Auth Road, Camp Springs, Maryland 20746, USA

Perot Systems Government Services, 8270 Willow Oaks Corporate Dr, Fairfax, Virginia 22301, USA

John Sapper
NOAA NESDIS, Office of Satellite Data Processing and Distribution (OSDPD), 5200 Auth Road, Camp Springs, Maryland 20746, USA

Corresponding author:
Prasanjit Dash
(CSU/CIRA Research Scientist II)
NOAA NESDIS STAR, WWB, Rm. 601-1,
5200 Auth Rd, Camp Springs, MD, USA 20746
Tel.: +1-301-763-8053x168; Fax: +1-301-763-8572
Email: *prasanjit.dash@noaa.gov*; *prasanjit.dash@colostate.edu*
Web: *http://www.star.nesdis.noaa.gov/star/Dash_P.php*

Washington DC,
Submitted: Nov-13-2009; resubmitted with minor revisions: Apr-30-2010

Abstract

The National Environmental Satellite, Data, and Information Service (NESDIS) has been operationally generating sea surface temperature (SST) products (T_S) from the Advanced Very High Resolution Radiometers (AVHRR) onboard NOAA and MetOp-A satellites since the early 1980s. Customarily, T_S are validated against *in situ* SSTs. However, *in situ* data are sparse and not available globally in near real-time (NRT). This study describes a complementary SST Quality Monitor (SQUAM), which employs global Level 4 (L4) SST fields as a reference standard (T_R) and performs statistical analyses of the differences, $\Delta T_S=T_S-T_R$. The results are posted in NRT at http://www.star.nesdis.noaa.gov/sod/sst/squam. The T_S data analyzed are the heritage NESDIS SST products from NOAA-16, -17, -18, -19, and MetOp-A, from 2001–present. The T_R fields include daily Reynolds, RTG, OSTIA, and ODYSSEA analyses. Using multiple fields facilitates distinguishing artifacts in satellite SSTs from those in the L4 products. Global distributions of ΔT_S are mapped and their histograms are analyzed for proximity to Gaussian shape. Outliers are handled using robust statistics, and the Gaussian parameters are trended in time to monitor SST products for stability and consistency. Additional T_S checks are performed to identify retrieval artifacts by plotting ΔT_S *vs*. observational parameters. Cross-platform T_S biases are evaluated using double differences, and cross-L4 T_R differences are assessed using Hovmöller diagrams. SQUAM results compare well with the customary *in situ* validation. All satellite products show a high degree of self- and cross-platform consistency except NOAA-16, which flies close to the terminator in recent years and its AVHRR is unstable.

1. Introduction

Sea surface temperature (SST) products have been operationally derived at NESDIS from the Advanced Very High Resolution Radiometers (AVHRR) since the early 1980s, employing regression-based multi-channel SST (MCSST) and non-linear SST (NLSST) techniques (McClain *et al.*, 1985; Walton, 1988). Satellite SSTs are best validated against *in situ* radiometers which also measure skin SST (*e.g.*, Suarez *et al.*, 1997; Donlon *et al.*, 1998; Kearns *et al.*, 2000; Minnett *et al.*, 2001; Noyes *et al.*, 2006). However, individual sea campaigns to collect *in situ* radiometry data are limited in space and time, their cost is prohibitive, and long-term routine deployment of radiometers at sea still remains difficult (Donlon *et al.*, 1998, 2002). Therefore, satellite SSTs are customarily validated against *in situ* SSTs from fixed and drifting buoys (*e.g.*, Walton *et al.*, 1998; Kilpatrick *et al*., 2001; Brisson *et al*., 2002; Dong *et al.*, 2006; O'Carroll *et al.*, 2006a,b; Haines *et al*., 2007; Lazarus *et al*., 2007, Merchant *et al*., 2008). However, global distribution of buoys is sparse and non-uniform in space and time (*cf.*, Garraffo *et al*., 2001). Furthermore, they originate from different countries and agencies, which use various measurement protocols thus rendering their quality non-uniform (*cf.*, Emery *et al.*, 2001). Moreover, attaining reliable validation statistics with *in situ* data typically requires up to a month, still leaving large geographical areas underrepresented.

This study explores an alternative approach for near real-time (NRT) monitoring of satellite SST products called the SST Quality Monitor (SQUAM). The SQUAM is based on statistical self- and cross-consistency checks, applied to differences between satellite SST, T_S, and global reference SST fields, T_R (Level 4, or L4 products), $\Delta T_S = T_S - T_R$ (Ignatov *et al.*, 2004; Dash *et al.*, 2009). Several different reference fields may be used, from an optimally interpolated blended satellite/*in situ* analysis (*e.g.*, Reynolds *et al.* 2002; 2007; Gemmill *et al.*, 2007; Stark *et al.*, 2007, 2008) to single/multiple satellite SST analyses, or even a climatological SST (*e.g.*, Bauer-Robinson 1985, Casey and Cornillon, 1999). The underlying assumption is that the probability density function of global ΔT_S is close to a Gaussian shape (although the distributions of both T_S and T_R are highly asymmetric). Statistical moments of a Gaussian distribution can thus be used to quality control (QC) the satellite SSTs and monitor them for stability and cross-platform consistency, in NRT.

The major premises of the SQUAM approach are that global reference fields cover the world oceans much more fully and uniformly, and the quality of such "sea truth" is also comparatively more uniform in space and time than that of *in situ* SST. This is because multiple satellite SST data, used in the production of L4 products, have already undergone extensive QC and have been bias-adjusted to match *in situ* SSTs, which were also quality controlled prior to blending (*cf.*, Reynolds *et al.* 2007). As a result, the number of "match-ups" with L4 fields is more than two orders of magnitude larger, and their geographical coverage and quality are much more uniform than (yet anchored to) the *in situ* SSTs. This provides a synoptic global snapshot of satellite SST performance (global maps, histograms, and dependencies of ΔT_S), and allows monitoring of the ΔT_S global statistics on fine time scales approaching NRT.

Ideally, an L4 product should optimally blend multiple satellite and *in situ* SSTs into a "true" SST. However, in reality most global SST analyses produced today use AVHRR data and one might therefore question whether comparison against these L4 products provides an independent assessment of the AVHRR SST. To explore sensitivity to T_R field, SQUAM employs several global L4 SSTs, including Reynolds, RTG, OSTIA, and ODYSSEA. These products are produced by different teams using various blending and optimal interpolation methods, and with different combinations of satellite (polar and geostationary; infrared and microwave) and *in situ* SST data as input. In particular, different L4s use different AVHRR

SST products, derived from different NOAA and MetOp platforms and using different cloud screening and SST algorithms (*cf.*, May *et al.*, 1998; Kilpatrick *et al.*, 2001; Le Borgne *et al.*, 2007; Gemmill *et al.*, 2007).

Currently, SQUAM evaluates NESDIS operational heritage SST products from five AVHRR/3 sensors onboard NOAA-16 (27 February 2001 - present), -17 (15 April 2003 - present), -18 (16 August 2005 - present), -19 (25 May 2009 - present), and MetOp-A (19 September 2007 - present). (Note that NESDIS heritage SST products are not input to any current L4 product employed in SQUAM.) SQUAM functions have been automated and operate with minimum manual intervention. Global processing is performed and statistics are posted to a dedicated website at *http://www.star.nesdis.noaa.gov/sod/sst/squam* within ~24 hours of availability of MUT and L4 products. The main purposes of SQUAM are to identify, in NRT, sensor and algorithm malfunctions, assess cross-platform consistency of products, diagnose artificial dependences, and generate global SST difference maps for highlighting residual clouds.

The paper is organized as follows: Sections 2 and 3 describe the NESDIS heritage Main Unit Task (MUT) SST data and the reference SST fields used. Section 3 also shows a brief comparison between different T_R fields. Section 4 details the global QC concept, with emphasis on handling outliers. Monitoring of SST for stability and self- and cross-platform consistency is described in Section 5. Section 6 concludes the paper and provides an outlook for the future.

2. NESDIS Main Unit Task (MUT) heritage SST product

2.1. MUT SSTOBS data

The NESDIS heritage Main Unit Task (MUT) system (McClain *et al.*, 1985; Walton, 1988; McClain, 1989) has been in operational use since the 1980s. It first ingests AVHRR 4 km global area coverage (GAC) level 1b data (Goodrum *et al.*, 2003), performs navigation and calibration of raw counts to radiances, and converts top-of-atmosphere radiances to brightness temperatures (BT). Further processing is performed on 2×2 GAC pixel arrays (referred to as unit arrays), resulting in an effective spatial resolution of ~8 km and the retrievals are restricted to ±53° view zenith angle. The MUT does not attempt to process every 8km unit array because of processing constraints since the heritage software was designed in an era when processing every pixel was not feasible. In the nighttime algorithm, only one unit array is chosen from a larger 11×11 target array sized to map to the ~20km High Resolution Infrared Sounder (HIRS) footprint. This unit array is chosen such that one of the four GAC pixels in the unit array is the warmest pixel in the target (*i.e.*, has the maximum BT in AVHRR channel 4). The warmest pixel is chosen to maximize the chance of a cloud free retrieval. In the daytime algorithm, retrieval density varies according to an input table which specifies, for each 10°×10° latitude longitude box, how many unit arrays to attempt to process within each 11×11 target array. SSTs are calculated by applying the regression equations listed in Table 1 using the unit array's averaged brightness temperatures. Retrievals are saved in platform-specific rotating files (SSTOBS), which are archived once a week. At any point in time, SSTOBS files contain ~8 days of global data. Along with SST, they also report time, location, view zenith angle (VZA), solar zenith angle (SZA), day/night flag (based on threshold SZA=75°), relative azimuth angle, reflectances in visible and BTs in thermal infrared channels, and nearest 100km analyzed field SST derived from last 24 hours of satellite SSTs. More details about MUT products are found in Ignatov *et al.* (2004). These archived SSTOBS "weekly" files are analyzed in this study.

INSERT TABLE 1 ABOUT HERE

Fig. 1 (top panels) shows time series of the number of nighttime and daytime observations, NOBS, from all five platforms. Initially, more AVHRR pixels are classified as "nighttime" by the MUT system, due to the use of SZA=75° day/night threshold (which also contains the twilight zone). However, due to a much heavier sub-sampling in MUT at night, each weekly SSTOBS file contains only 400,000-500,000 nighttime NOBS, whereas during the daytime, NOBS range from 400,000 to 800,000. (Note that these numbers are somewhat "inflated" as each "weekly" SSTOBS file actually contains ~8 days of data, with 1 day overlap. No attempt was made in this study to remove these overlapping days and thus each point in the time series plots corresponds to one entire 8-day SSTOBS file.) Examples of global nighttime and daytime SST maps from NOAA-18 are shown in Fig. 1 (bottom panels). At night, the spatial coverage is globally more uniform and complete, despite generally smaller NOBS. Another observation is that the high latitudes are predominantly observed at night, due to low Sun in these areas.

INSERT FIG. 1 ABOUT HERE

The nighttime NOBS are relatively stable in time, whereas daytime NOBS show large variations partly due to the continuous monthly updates of the AVHRR calibration in the visible bands and periodic revisions of the associated cloud thresholds. Also during daytime, MUT retrieval density can be increased regionally, using a table as previously noted, based on users' requirements. These changes can be further modulated by the seasonal variations in illumination, sun glint, and clouds in the high-density areas.

Note that the NOAA-16 processing in MUT was suspended in mid-2005, shortly after the NOAA-18 became operational, but was then resumed to facilitate multi-sensor consistency analyses and to better quantify the effects of NOAA-16 orbital evolution and sensor anomalies on the SST product. Also, note that in late 2008, NOAA-16 daytime NOBS decreased and nighttime NOBS increased, due to orbital drift. The operational production from NOAA-17 was succeeded by MetOp-A in April 2007 and from NOAA-18 by NOAA-19 in June 2009. Nevertheless, processing of NOAA-17 and -18 continues and their SST products are analyzed in SQUAM. Monitoring SST from multiple platforms in SQUAM enables cross-platform analyses, and may facilitate their future potential blending for improved SST products.

2.2. Modifications to the original MUT SSTOBS data for SQUAM analyses

SST values are saved in the original heritage Main Unit Task (MUT) SSTOBS files to only one decimal place, whereas VZA and all BTs are available to two decimal places. Also, in the original SSTOBS files, only one climatological reference SST field (Bauer & Robinson, 1985) is available. Prior to the SQUAM analyses, two modifications are made to the original SSTOBS data. First, the SSTs are recalculated using the corresponding regression equations given in Table 1 and stored with two decimal places. Second, seven other reference SST fields listed in Section 3 are appended.

3. Global reference SST fields used in SQUAM

This section summarizes seven SST analyses fields appended to MUT data. The Group for High Resolution SST Pilot Project (GHRSST-PP, Donlon *et al.*, 2007; http://www.ghrsst.org) established a concerted effort towards generation and reconciliation of high-quality Level 4 (L4) SST fields. The L4 products employed in SQUAM have been either developed within the GHRSST framework or comply with its standards and specifications. The Reynolds and RTG SSTs are normalized to *in situ* SST and therefore are considered bulk SSTs. The OSTIA product is referred to as a "foundation SST" (Donlon *et al.*, 2007). It minimizes the effect of diurnal thermocline by using only nighttime satellite data, and daytime data with wind speed above 6 m s^{-1}. An empirical correction of 0.17 K is applied to convert satellite skin SST to the foundation. The ODYSSEA is solely based on nighttime satellite

SSTs (no *in situ*), which are subsequently corrected by 0.17 K similarly to OSTIA, and termed a "sub-skin" product. The input data to all L4 products are listed in Table 2.

INSERT TABLE 2 ABOUT HERE

3.1. Reynolds Optimal Interpolation SSTs (OISST)

In SQUAM, three different Reynolds OISST products are employed: weekly 1° OI.v2 (Reynolds *et al.*, 2002) (WOI) and two daily 0.25° (Reynolds *et al.*, 2007).

The WOI SST 1° (180 × 360 grid) data are available for the time period from 1981 to present, in a binary format at *ftp://ftp.emc.ncep.noaa.gov/cmb/sst/oisst_v2*. The data are centered at the middle of the week (Wednesday).

Two daily 0.25° (720 × 1440 grid) OISST products are available at *ftp://eclipse.ncdc.noaa.gov/pub*. One is a blend between AVHRR and *in situ* SSTs (DOI_AV), and the other additionally uses Advanced Microwave Scanning Radiometer (AMSR-E) SST data (DOI_AA). Both DOI_AV and DOI_AA SST data are reported in several formats including the network Common Data Form (netCDF). All Reynolds products used Pathfinder SST (Kilpatrick *et al.*, 2001) from January 1985 to December 2005, and operational Naval Oceanographic Office (NAVOCEANO) AVHRR SST (May *et al.*, 1998) from January 2006 onward. The DOI_AA SST is available since June 2002, after AMSR-E data became available from the *Aqua* satellite. The main benefit of using AMSR-E data is its near-all-weather SST coverage. The DOI had switched from *version 1* to *version 2* on 6 January 2009 (*http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily.php*) and so did NRT SQUAM analyses.

3.2. Real Time Global (RTG) daily SSTs

Two RTG SSTs (*http://polar.ncep.noaa.gov/sst*) are available, one at a relatively low resolution of 0.5° (RTG_LR, 360 × 720 grid; Thiébaux *et al.*, 2003) and the other at a higher resolution of 1/12° (RTG_HR, 2160 × 4320 grid; Gemmill *et al.*, 2007). The RTG_LR is available from 30 January 2001 to present at *ftp://polar.ncep.noaa.gov/pub/history/sst*. It uses the same input data as DOI_AV (*i.e.*, operational NAVOCEANO AVHRR and *in situ* SSTs) but processes them differently. The RTG_HR became operational on 27 September 2005 and the product is available at *ftp://polar.ncep.noaa.gov/pub/history/sst/ophi* from April 2007 to present. The SST input to the RTG_HR is based on a new physical retrieval system developed at the Joint Center for Satellite Data Assimilation (Gemmill *et al.*, 2007).

3.3. Operational SST and Sea Ice Analysis (OSTIA)

The OSTIA product was developed at the UK Met Office in response to the requirements of the GHRSST (Stark *et al.*, 2007; 2008). The data are generated daily at a 0.05° (3600 × 7200 grid) spatial resolution and made available in netCDF format at *ftp://podaac.jpl.nasa.gov/GHRSST/data/L4/GLOB/UKMO/OSTIA*. The data are available free-of-charge for non-commercial purposes. However, users are required to obtain a license agreement (*http://www.metoffice.gov.uk/corporate/legal/data_lic_form.html*). The dataset is available from 1 April 2006 to present.

3.4. Ocean Data Analysis System for Mersea (ODYSSEA)

The ODYSSEA SST is a daily 0.1° resolution (1600 × 3600 grid; -80° to 80° latitude) product. It was developed in the framework of the MERSEA project (*http://www.mersea.eu.org*) and complies with the GHRSST standards (Autret & Piollé, 2007). The data in netCDF format are available at ftp://ftp.ifremer.fr/ifremer/medspiration/data/l4hrsstfnd/eurdac/glob/odyssea since October 2007 to present.

3.5. Comparisons of global L4 SST fields

Although the primary objective of the SQUAM validation tool is to monitor satellite SSTs, diagnostics of the L4 fields are provided as well. Fig. 2 shows example time series plots of global mean differences and standard deviations in several T_R fields w.r.t. RTG_LR.

INSERT FIG. 2 ABOUT HERE

The two daily Reynolds products (DOI_AV and DOI_AA) appear mutually consistent. Their mean differences and standard deviations w.r.t. RTG_LR show a clear seasonal cycle, with amplitudes from -0.1 to +0.2K and 0.5-0.9K, respectively. The OSTIA and ODYSSEA differences w.r.t. RTG_LR also show seasonality, although the corresponding mean differences and standard deviations are somewhat smaller. Shortly after its inception in February 2006, OSTIA had a cold mean difference of -0.2K, which reduced to -0.1K later in 2006 but then shortly spiked again to -0.2K in early 2007. After this initial period, the OSTIA product has been fairly consistent with DOI. ODYSSEA SST shows some short-term spikes up to ~+0.2K in early 2008 and 2009, and occasional data gaps (*e.g.*, 20-21 August 2008).

To better understand the observed L4 differences, the comparisons were further stratified into those with $T_R \geq 0°C$ (Fig. 2, middle panels) and with $T_R < 0°C$ (Fig. 2, bottom panels). These analyses suggest that the major differences between L4 products takes place in the high latitudes (Fig. 2, bottom panels), likely due to different treatment of the marginal ice zone (*e.g.*, Reynolds *et al.*, 2007).

Although some products show a global mean difference to be close to zero, their large standard deviation suggests some significant regional differences. Fig. 3 shows two examples of Hovmöller diagrams (latitude-time evolution of T_R *minus* DOI_AV), to understand the zonal differences.

INSERT FIG. 3 ABOUT HERE

In mid-latitudes, the DOI_AV, OSTIA, and RTG_LR match each other quite well. However, much larger and seasonal mean differences are found in the higher latitudes.

Analyses in this subsection suggest that more work is needed to reconcile different L4 products, especially in the high latitudes. In the remainder of this paper, only three reference SSTs are used (DOI_AV, RTG_LR, and OSTIA). More intercomparisons between daily L4 fields are available at *http://www.star.nesdis.noaa.gov/sod/sst/squam/L4/*.

3.6. Match-up of MUT SST with L4 fields

Satellite SSTs are matched to the reference SST datasets using the nearest neighbor approach and no interpolation in space and time is attempted. All T_R fields provide near-global and almost gap-free coverage, so that there are only a few MUT SST retrievals found outside the domains covered by these fields. The MUT SSTs without corresponding reference SSTs are excluded from the SQUAM analyses.

4. Global quality control (QC) and handling outliers in SQUAM

Fig. 4 shows examples of nighttime and daytime maps of ΔT_S for MetOp-A and OSTIA SSTs. Generally, ΔT_S is close to zero. However, in some pixels, T_S is either too warm or too cold relative to T_R, suggesting that these points are likely outliers.

INSERT FIG. 4 ABOUT HERE

The distribution of ΔT_S can be significantly distorted by outliers. The outliers may be due to “contaminant” points in T_S, T_R, or both, or they may be caused by “discordant” data points due to, *e.g.*, T_S - T_R space-time mismatch in areas of high SST gradients (*e.g.*, at the boundaries of oceanic currents, upwellings, *etc.*). If the objective is to provide a high-quality, satellite SST product, then only “contaminant” T_S’s should be excluded and “discordant” T_S’s retained, to preserve SST information in the dynamic oceanic areas. If the objective is to routinely monitor the global performance of SST products, which is the subject of SQUAM analyses, then both “contaminant” and “discordant” observations are to be excluded.

Customarily, outliers in data are handled using one of the two principal approaches, “identification” or “accommodation” (*e.g.*, Tietjen, 1986). “Identification” involves labeling and removing outliers from the data, whereas “accommodation” belongs in the area of robust estimation. In SQUAM, these two approaches are used in concert, in order to most effectively handle outliers in the data.

A common “identification” approach is removing data points beyond a confidence interval based on 3 or 4 standard deviations about the mean (*cf.*, Bevington and Robinson, 1992). The exact number of standard deviations used in QC can be based on Chauvenet’s criterion, which links the probability to the sample size (*cf.*, Bevington and Robinson, 1992). This study employs a simpler approach based on using a fixed N=4, irrespective of the sample size (*e.g.*, Ostle and Malone, 1988).

In reality, the conventional mean and standard deviation themselves are contaminated by outliers, rendering their use for “identification” progressively less effective as the fraction of outliers increases. To circumvent this problem, robust statistics (median and robust standard deviation, RSD) are employed in SQUAM to construct the screening thresholds, *i.e.*, “median ± 4×RSD” (*cf.*, Merchant and Harris, 1999). The RSD of a distribution is given as: IQR/*S*, where, IQR is interquartile range (75^{th} percentile – 25^{th} percentile, in an ordered dataset) and *S* is a scaling factor (1.348 for an ideal normal distribution).

4.1. Histograms of ΔT_S

Fig. 5 shows typical nighttime histograms of ΔT_S for NOAA-18 against *in situ* (within 20 km × 1 hour) and OSTIA data, before and after removal of outliers. The equivalent number of match-ups w.r.t. *in situ* data is ~250 times smaller than w.r.t. OSTIA (~7,000 per month and ~450,000 per 8-day period, respectively). Also, the fraction of outliers is a factor ~2 higher than w.r.t. OSTIA data, indicating that *in situ* SSTs themselves are strongly contaminated by bad data. Statistical parameters and a Gaussian fit, *X~N(median, RSD)*, are also annotated in the histograms.

INSERT FIG. 5 ABOUT HERE

Prior to the removal of outliers (Fig. 5, left), min(ΔT_S) and max(ΔT_S) reach ~±20 °C for *in situ* data and ~±10 °C for OSTIA. The extreme ΔT_S values (minimum and maximum) before removing the outliers are likely due to failed cloud detection and land/glint contamination (see Section 4.2 for distribution of outliers). Mean and median estimates of the global average ΔT_S distribution are close to each other, with a magnitude of only a few hundredths of a kelvin. However, the RSDs and conventional standard deviations differ significantly. For instance, the RSD=0.25 K w.r.t. *in situ* data, corresponding to only ~14% of the variance measured by the conventional standard deviation=0.67 K. For OSTIA, RSD~0.29 K, compared with standard deviation~0.47 K. This is because the conventional standard deviation is artificially inflated by outliers. The conventional values of skewness (s~2.28 for *in situ* and s~2.62 for OSTIA) and kurtosis (k~252 and ~39, respectively) indicate strong asymmetry and peakedness of the empirical histograms. (Note that no robust measures of the 3^{rd} and 4^{th} moments are employed in SQUAM.)

After excluding outliers (Fig. 5, right), the robust statistics (median and RSD) remain practically unchanged, as expected. The conventional statistics however do change, the higher moments improving dramatically. In particular, the standard deviation is significantly reduced and becomes closer to the RSD, which changes only a little. The kurtosis (1.01 and 1.07 for *in situ* and OSTIA, respectively) becomes much more realistic and representative of the observed distribution. The min(ΔT_S) and max(ΔT_S) are now within ~±1.2 K as data are not allowed to depart from the median by more than four RSDs (with typical RSD < ~0.3 K).

To summarize analyses of histograms, the distributions of ΔT_S are indeed close to Gaussian but contaminated by a small fraction of outliers. The global differences (mean and median) are close to zero and RSDs range from ~0.3-0.5 K, quite close to the similar metric against *in situ* SST (*cf.*, McClain *et al.*, 1985; Walton *et al.*, 1998a; May *et al.*, 1998). One thus concludes that validation against global reference fields can be successfully used to monitor satellite SST products globally and in NRT.

4.2. Distribution of outliers in space and time

Although outliers are generally considered a nuisance for validation purposes, their distribution in space and time may carry important information about their source and help identify potential areas for improvement in the satellite or reference SSTs. Fig. 6 shows examples of global distributions of low (*i.e.*, ΔT_S < median - 4×RSD) and high (*i.e.*, ΔT_S > median + 4×RSD) outliers in the nighttime ΔT_S for MetOp-A.

INSERT FIG. 6 ABOUT HERE

Reproducible low outliers (*e.g.*, in the Northern Pacific, "roaring forties", off East African coast, and southeast Arabian Sea) are predominantly associated with persistent cloud and aerosols, and suggest the need for improvements in satellite SST. On the other hand, consistent pattern of prominent high outliers (especially in the high latitudes, and in the Northern Hemisphere) may be due to a low bias in all L4 products, although high bias in

AVHRR SST may not also be ruled out (*cf.*, also comparisons between different L4's in Section 3.5 which show highest uncertainties in the high latitudes). A reduced number of high outliers in the Arctic (above ~65°N) in the DOI_AV SST, relative to RTG and OSTIA, may be due to their different processing of sea-ice boundary (*e.g.*, Reynolds *et al.*, 2007). Many reproducible distribution patterns, with high and low outliers closely interleaved, are found in the high-gradient regions (such as Gulf Stream, Brazil Current, Mozambique and Agulhas Current to the south of Africa, and East Australia Current). Those are likely caused by mismatches between T_S and T_R, partly due to the inherent variability within a given T_R grid and partly due to the different spatial resolutions of T_S and T_R, whose combined effects may become significant in highly dynamic oceanic areas.

INSERT FIG. 7 ABOUT HERE

Fig. 7 shows a time series of nighttime outliers. Consistent with Fig. 5, the fraction of outliers against *in situ* data is a factor 2 to 3 larger compared to L4, which suggests a persistently strong contamination in the *in situ* SSTs. For all platforms, the rate of low outliers (likely indicating residual cloud in MUT nighttime SST data) is relatively flat in time and ranges from ~0.5 to 1.0 %. The right panels of Fig. 7 show corresponding time series of high outliers, which exhibit a strong annual cycle with maximum reaching ~2.5% in July-August. This seasonality mainly comes from the high latitudes of the Northern Hemisphere, which are sampled by the polar-orbiting NOAA and MetOp platforms more frequently during the boreal summer (*cf.*, Fig. 6). In December-January, fraction of high outliers is reduced to ~0.5%, consistent with the general level of low outliers. This seasonality suggests either consistent problems with L4 products in the ice melting zone, or problems with AVHRR cloud screening and/or SST algorithms, or both. More analyses are needed to reconcile different satellite and L4 data in this complex area, which is generally lacking *in situ* data (*cf.*, top panels of Figs. 6 and 7).

5. Monitoring satellite SST for stability and cross-platform consistency

In SQUAM, the statistical moments of the ΔT_S distributions are monitored to assess satellite data for stability and cross-platform consistency. Following the discussions in Section 4.2, only outlier-free data are analyzed here.

5.1. Monitoring stability of satellite SST products

Fig. 8 shows a time series of median ΔT_S. Although the major trends are captured well in all time series, the L4 plots show more fine structure compared to *in situ* results, due to a much larger number of match-ups supporting each data point, and higher temporal resolution (8 days instead of 1 month).

INSERT FIG. 8 ABOUT HERE

Two major types of anomalies are observed in the time series.

The first group is due to problems with satellite SST. For instance, the NOAA-17 and -18 ΔT_S's track each other closely, whereas MetOp-A has been biased ~0.1 K high until mid-2009. NOAA-16 shows a highly anomalous behavior, including two large dips in late 2006 and 2007, followed by a series of smaller dips in 2008 and 2009. Recall that NOAA-16 currently flies close to the terminator and its AVHRR continuously experiences rapid changes in its thermal regime, and its black body is subject to frequent solar impingement (*cf.*, Cao *et al.*, 2001). Additional offline analyses (not shown here) confirm that its calibration coefficients in all bands undergo cyclic changes. Work is underway to better understand and resolve this NOAA-16 anomaly. For the rest of this study, NOAA-16 data will be excluded from further analyses and discussion.

The second group of anomalies in Fig. 8 comes from the reference fields themselves. The degree and magnitude of these artifacts is L4 product specific. For instance, there are two "jumps" in the DOI_AV plots in 2004 and 2005, and one "jump" in the first half of 2007. These artifacts are also seen, although to a lesser extent, in the RTG_LR. Another example of a non-reproducible feature is observed in the OSTIA time series, which shows an elevated SST anomaly in 2006 and a spike in the 1st quarter of 2007. Also, different L4 products show a different degree of short-term "noise", which is smaller in OSTIA and RTG_LR and larger in DOI_AV time series.

Note that despite artifacts in individual L4 products, the time series of ΔT_S from different platforms track each other very closely, suggesting that selection of T_R is not critical for monitoring cross-platform consistency of different T_S products.

5.2. Using Double Differences (DD) to monitor satellite SST for cross-platform consistency

A more direct way to monitor T_S for cross-platform consistency is based on using the double differencing (DD) technique (*cf.*, Alber *et al*., 2000). The DD methodology has been employed in remote sensing for many applications including transferring calibration from one satellite sensor to another *via* a third "transfer standard" sensor (*cf.*, Wang and Wu, 2008). For our analyses, the NOAA-17 ΔT_S was selected as the respective "transfer standard", and subtracted from the corresponding ΔT_S's for other platforms as follows: DD = $(T_{S,SAT}–T_R) – (T_{S,N17}–T_R) \approx T_{S,SAT}–T_{S,N17}$. Note that monitoring cross-platform consistency with direct differencing, *i.e.*, "$T_{S,SAT}–T_{S,N17}$" is also possible, but only in the "intersection sub-sample" of the two satellites, and therefore is more geographically non-uniform from one temporal snapshot to another. This issue is largely alleviated when the DD technique is used. The major premise is that the T_R, which is subject to artifacts and irregularities, cancels out and the DD thus provides a measure of average cross-platform consistency, in a global domain.

Fig. 9 shows time series of the DDs, for several different T_R's. The patterns are quite consistent for different L4s, suggesting that the respective DDs are largely insensitive to the selected T_R. Based on the nighttime local overpass times for different satellites (~9:30 p.m. for MetOp-A, ~10 p.m. for NOAA-17, and ~2 a.m. for NOAA-18 and -19), one would expect the best consistency between NOAA-17 and MetOp-A, and between NOAA-18 and -19. Since all global L4 products currently do not resolve the diurnal cycle, the second cluster is expected to be several hundredths of a degree kelvin cooler than the first cluster, based on the expected diurnal cooling at night (*cf*., Stuart-Menteth *et al*., 2005; Gentemann *et al*., 2007; Kennedy *et al*., 2007). Indeed, NOAA-18 and -19 closely agree, but MetOp-A is biased high w.r.t. NOAA-17 by ~+0.1 K, until about mid-2009. Note that these relationships are also seen in Fig. 8, but DDs provide a better way to quantify the cross-platform biases.

INSERT FIG. 9 ABOUT HERE

The DDs look different when *in situ* SST is used as T_R (Fig. 9, top panel). Recall that *in situ* bulk SSTs account for the diurnal variation in SST, but only partially, because, the diurnal cycle in bulk SST is always suppressed compared with skin SST. In the top panel of Fig. 9, it is again expected that NOAA-17 and MetOp-A form one cluster, and NOAA-18 and -19 form another (and colder) one. The data do follow this expected pattern but not fully.

5.3. Using Double Differences (DD) to monitor satellite SST for day-night consistency

The DD technique can also be employed to quantify platform-specific day minus night (or "DN") SST biases. Recall also that in addition to the diurnal cycle in SST, artificial DN biases may occur because the regression coefficients in the daytime (NLSST) and nighttime (MCSST) algorithms are tuned independently against *in situ* SSTs, and the DN check is also useful to verify the relative consistency of these tunings. Fig. 10 shows global average of "DN" satellite SST biases calculated as: $DN = (T_{S,D}–T_R) – (T_{S,N}–T_R) \approx T_{S,D}–T_{S,N}$.

INSERT FIG. 10 ABOUT HERE

Should *in situ* SST fully account for the diurnal cycle in skin SST, then the time series on the top panel of Fig. 10 would have been flat and at 0 K. However, as discussed before, this accounting is only partial. As a result, all DNs show a small positive bias, with a clear seasonal cycle from 0-0.15 K. This cycle is caused by systematic changes in skin-bulk difference, as affected by the solar insolation and wind speed and modulated by the changing global coverage.

Turning to the L4 results in Fig. 10, the shape of the corresponding DNs is largely insensitive to the reference SST. This is expected, since the current L4 products do not resolve the diurnal cycle, and therefore the global DN here captures the average differences between the satellite skin SSTs between the day and night satellite overpasses. For the two ~10 a.m./p.m. platforms, NOAA-17 and MetOp-A, the DNs range from 0-0.2 K, and track each other closely. For the 2 a.m./p.m. platforms, the DNs change from 0.2-0.4 K. The DN for 2 a.m./p.m. is larger than for ~10 a.m./p.m. platforms, because the corresponding local overpass times are close to the diurnal minimum and maximum of SST (*cf.*, Stuart-Menteth *et al.*, 2005).

Work is underway to explore the potential of the DD technique to better quantify and minimize cross-platform and day-night biases.

5.4. Using higher moments for SST monitoring

Fig. 11 shows time series of robust standard deviation (RSD) corresponding to Fig. 8.

INSERT FIG. 11 ABOUT HERE

For all T_R fields, there is excellent cross-platform consistency. The nighttime RSDs w.r.t. *in situ* and OSTIA SSTs are ~0.3 K, followed by RTG_LR and DOI_AV (< 0.4 K).

Note a close proximity of the OSTIA and *in situ* SST validation results.

Non-uniformities in the time series are deemed to be due to changes in the quality of the reference fields themselves. For instance, the RSD w.r.t. to *in situ* SST has decreased from 0.4 to 0.3 K since 2003, likely due to improved quality of *in situ* SST. The drop in DOI_AV RSD from >0.5 K to < 0.4 K on January 1, 2006 coincides with the switch in DOI production from Pathfinder to NAVOCEANO SST as the primary input (Reynolds *et al.*, 2007). Similar non-uniformities (although of somewhat smaller magnitude) are also observed in the RTG_LR time series in 2004 and 2005 for NOAA-17 and in late 2005 for NOAA-18. Some of these changes might have been caused by incorporation in the RTG processing of the NOAA-17 and -18 data, respectively.

5.5. Additional self-consistency diagnostics

SQUAM additionally performs self-consistency checks of SST products, by plotting global ΔT_S as a function of relevant observational and geophysical variables, such as the view zenith angle (VZA). A case study is shown in Fig. 12 where nighttime NOAA-17 and -18 ΔT_S are plotted against VZA, for two different periods: one before January 2006 and one in the beginning of January 2006. The dependence prior to January 2006 (Fig. 12, left panel) shows an artificial across-swath bias of > 0.3 K. This bias was caused by a faulty assignment of VZA in MUT and was uncovered with analyses from an early prototype of SQUAM. Notice the reduction in dependence and improved symmetry after correction (Fig. 12, right panel).

Analyses similar to those shown in Fig. 12 are routinely performed in SQUAM to identify and remove any artificial dependencies. Such synoptic diagnostics can be reliably obtained in NRT only using a global field as the reference. Note that selection of a particular reference field is not critical for these analyses. With *in situ* data, similar synoptic diagnostics can also be o but in significantly delayed and time-integrated mode (*cf.*, Merchant *et al.*, 2008).

INSERT FIG. 12 ABOUT HERE

6. Summary and future work

The web-based SST quality monitor (SQUAM) is employed to continuously control the quality of NESDIS operational AVHRR SST products (T_S). Similarly to the customary validation against *in situ* SST, SQUAM performs analyses of SST differences, $\Delta T_S = T_S - T_R$, but calculated with respect to various L4 products, including Reynolds, RTG, OSTIA, and ODYSSEA. Processing is done automatically and results are posted at http://www.star.nesdis.noaa.gov/sod/sst/squam in near-real time (NRT).

The major trends and anomalies seen against *in situ* SSTs are also well captured against L4 fields. Due to its extensive "validation statistics", SQUAM performs global quality control of satellite SSTs by checking ΔT_S for proximity to a Gaussian shape and by handling outliers, in NRT. Global maps, histograms and dependencies plots of ΔT_S are generated for synoptic assessment of satellite SST products, and moments of the ΔT_S distributions are trended in time. Satellite SSTs are further monitored for cross-platform and day-night consistency using double differences (DD).

Testing NESDIS heritage AVHRR SSTs from NOAA-16, -17, -18, -19, and MetOp-A from 2001 to present shows that, overall, the products are stable and cross-platform consistent. The initial warm bias in nighttime MetOp-A SSTs of +0.1 K, which was likely due to specifying suboptimal regression coefficients in its MCSST equation, has been greatly reduced in mid-2009. The NOAA-16 product shows a distinct "out-of family" behavior, apparently due to unstable AVHRR calibration in recent years, likely caused by its near-terminator orbit. Improvements of NOAA-16 AVHRR calibration (*cf.*, Trishchenko, 2002; Mittaz *et al.*, 2009) may be explored in a future work. The remaining differences are largely attributed to different temporal sampling from different platforms, and to the diurnal variability in the satellite SST, which is currently not resolved in the global L4 fields.

Using multiple T_R's facilitates distinguishing artifacts in satellite SSTs from those in T_R fields. In particular, all AVHRR products show widespread positive biases in the Arctic, suggesting that low biases are possible in all current L4 fields. Comparisons between different L4 fields are also performed in SQUAM. They show important differences, particularly in high latitudes, which presumably originate from different treatment of the sea-ice marginal zone in different L4 analysis schemes. Some L4 products show various non-uniformities in time and a larger degree of day-to-day "noise".

Identifying one "most suitable" L4 field would simplify SQUAM analyses. However, different T_R fields emphasize different aspects of SST (bulk, foundation, sub-skin), are available for different time periods, and have different spatial resolution, quality, and data stability. Validation statistics against some L4 fields (*e.g.*, OSTIA) approach the biases and standard deviations measured against *in situ* data, while for others (RTG, Reynolds) the validation statistics are slightly degraded (larger). SQUAM analyses can contribute to objective evaluation of different satellite and L4 SST products and facilitate their improvement and possibly their convergence. In particular, it supplements a high resolution diagnostic data set (HR-DDS; *http://www.hrdds.net*) system which allows, at specified locations (not global), interactive analysis of several satellite, *in situ*, and model data, and a global and regional monitoring facility at the National Centre for Ocean Forecasting (NCOF) (http://ghrsst-pp.metoffice.com/pages/latest_analysis/sst_monitor/). The SQUAM, HR-DDS, and NCOF tools can be used in concert for comprehensive intercomparison of global products. We also plan to explore the GHRSST ensemble of the standard L4 products in SQUAM (*http://ghrsst-pp.metoffice.com/pages/latest_analysis/sst_monitor/daily/ens*).

The near-term SQUAM objective will be working towards reconciliation of all AVHRR SST products from different platforms, during day and night, and establishing a consistent benchmark SST product. Two particular tasks which will be pursued towards that goal are modeling diurnal variability in SQUAM (*e.g.*, implementing model of Gentemann *et al*., 2007 or climatological data of Kennedy *et al*., 2007), and exploring improved AVHRR calibration.

Recently, NESDIS's newly developed Advanced Clear Sky Processor for Oceans (ACSPO) and NAVOCEANO AVHRR GAC SST products were also included in the SQUAM processing. Analyses of the ACSPO SST products and establishment of reliable links with the heritage MUT SST products are underway. SQUAM will also be adapted to monitor other existing (such as MSG SEVIRI) and future (such as MetOp-B and –C AVHRRs, NPOESS VIIRS and GOES-R ABI) sensors. The SQUAM will also be instrumental in quality control of climate data records (*cf*., Vázquez-Cuervo *et al.*, 2004) and in establishing links between the past, present, and next generation SST products.

Acknowledgments

This work was supported by the Product System Development and Implementation program managed by the NESDIS Office of Systems Development, by the Internal Government Studies managed by the NPOESS Integrated Program Office, and by the GOES-R Algorithm Working Group. We thank members of the SST Team (F. Xu, X. Liang, B. Petrenko, N. Shabanov, J. Stroup, and D. Frey) for helpful discussions and constructive feedback. P. Dash acknowledges the CIRA visiting scientist fellowship. We would also like to thank the anonymous reviewers for their comments. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or US Government position, policy, or decision.

References

Alber C., R. Ware, C. Rocken, and J. Braun, 2000: Obtaining single path phase delays from GPS double differences. *Geophys. Res. Lett.*, **27(7)**, 2661-2664.

Autret, E., and J.-F. Piollé, 2007: ODYSSEA Global SST Analysis - User manual, MERSEA-WP02-IFR-STR-001-1A, CERSAT–IFREMER. [Available online at: http://cersat.ifremer.fr/data/discovery/by_parameter/sea_surface_temperature/odyssea_global_sst_analysis]

Bauer, R. A., and M. K. Robinson, 1985: *Description of the Bauer-Robinson Numerical Atlas.* Compass Systems, Inc., 4640 Jewell St., #204, San Diego, CA 92109, 13 pp.

Bevington, P. R., and D. K. Robinson, 1992: Estimates of mean and errors. *Data Reduction and Error Analysis for the Physical Sciences*. 2d ed. McGraw-Hill, 328 pp.

Brisson, A., P. Le Borgne, and A. Marsouin, 2002: Results of one year preoperational production of sea surface temperatures from GOES-8. *J. Atmos. Oceanic Technol.*, **19**, 1638-1652.

Casey, K. S., and P. Cornillon, 1999: A comparison of satellite and in situ-based sea surface temperature climatologies. *J. Climate*, **12**, 1848–1863.

Cao, C., M. Weinreb, and J. Sullivan, 2001: Solar contamination effects on the infrared channels of the advanced very high resolution radiometer, *J. Geophys. Res.*, 106, 33,463–33,469, doi:10.1029/2001JD001051.

Dash, P., A. Ignatov, Y. Kihai, J. Sapper, and X. Liang, 2009: Web-based global quality control and monitoring of NESDIS AVHRR SST products for long term stability and cross-platform consistency in near real-time. Extended abstract, *89th AMS Annual Meeting & 16th Conf. on Sat. Met. & Oceanography,* 11-15 Jan. 2009, Phoenix, AZ.

Donlon, C., I. Robinson, K.S. Casey, J. Vázquez-Cuervo, E. Armstrong, O. Arino, and C.

Gentemann, *et al.*, 2007: The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bull. Amer. Meteor. Soc.*, **88**, 1197–1213.

Dong, S., S. T. Gille, J. Sprintall, and C. Gentemann, 2006: Validation of the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) sea surface temperature in the Southern Ocean. *J. Geophys. Res.,* **11**, doi:10.1029/2005JC002934.

Donlon, C. J., P. J. Minnett, P. J., C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray, 2002: Toward improved validation of satellite surface skin temperature measurements for climate research. *J. Climate*, **15**, 353–369.

Donlon, C. J., S. J. Keogh, D. J. Baldwin, I. S. Robinson, I. Ridley, T. Sheasby, I. J. Barton, E. F. Bradley, T. J. Nightingale, and W. Emery, 1998: Solid-state radiometer measurements of sea surface skin temperature. *J. Atmos. Oceanic Technol.*, **15***,* 775-787.

Emery, W. J., D. J. Baldwin, P. Schlussel, and R. W. Reynolds, 2001: Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *J. Geophys. Res.*, **106**, 2387-2405.

Garraffo, Z. D., A. J. Mariano, A. Griffa, C. Veneziani, and E. P. Chassignet, 2001: Lagrangian data in a high-resolution numerical simulation of the North Atlantic I. Comparison with in situ drifter data. *J. Mar. Syst.*, **29**, 157-176.

Gemmill, W., B. Katz, and X. Li, 2007: Daily real-time, global sea surface temperature- high-resolution analysis: RTG_SST_HR, NOAA/NCEP. NOAA / NWS / NCEP / MMAB Office Note Nr. 260, 39 pp. [Available online at: http://polar.ncep.noaa.gov/sst/]

Gentemann C. L., C. J. Donlon, A. Stuart-Menteth, and F. J. Wentz, 2007: Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.*, **30**, doi:10.1029/2002GL016291.

Goodrum G., K. Kidwell, and W. Winston, Eds., 2003: NOAA-KLM user's guide.

NOAA/NESDIS, National Climatic Data Center, 151 Patton Ave., Asheville, NC 28801-5001. [Available online at http://www2.ncdc.noaa.gov/docs/klm/index.htm]

Haines, S. L., G. J. Jedlovec, and S. M. Lazarus, 2007: A MODIS Sea Surface Temperature Composite for Regional Applications, *IEEE Trans. Geosci. Remote Sens.*, **45**, 2919-2927.

Ignatov, A., J. Sapper, I. Laszlo, N. Nalli, A. Harris, W. Pichel, A. E. Strong, E. Bayler, X. Li, and E. Maturi, 2004: Extended abstract, Global operational sea surface temperature and aerosol products from AVHRR: current status, diagnostics, and potential enhancements. *13th AMS Conf. on Sat. Met. & Oceano.,* 20-24 Sep, Norfolk, VA, Amer. Meteor. Soc.

Kearns, E. J., J. A. Hanafin, R. H. Evans, P. J. Minnett, and O. B. Brown, 2000: An Independent Assessment of Pathfinder AVHRR Sea Surface Temperature Accuracy Using the Marine Atmosphere Emitted Radiance Interferometer (MAERI). *Bull. Amer. Meteor. Soc.*, **81**, 1525–1536.

Kennedy J. J., P. Brohan, and S. F. B. Tett, 2007: A global climatology of the diurnal variations in sea-surface temperature and implications from MSU temperature trends. *Geophys. Res. Lett.*, **34**, doi:10.1029/2006GL028920.

Kilpatrick K. A., G. P. Podesta, 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**, 9179-9198.

Lazarus, S. M., C. G. Calvert, M. E. Splitt, P. Santos, D. W. Sharp, P. F. Blottman, and S. M. Spratt, 2007: Real-Time, High-Resolution, Space–Time Analysis of Sea Surface Temperatures from Multiple Platforms. *Mon. Wea. Rev.*, **135**, 3158-3173.

Le Borgne, P., G. Legendre, and A. Marsouin, 2007 : Operational SST Retrieval from MetOp/AVHRR, *Proc. 2007 EUMETSAT Conf., Amsterdam*. [Available online at http://www.eumetsat.int/Home/Main/Publications/Conference_and_Workshop_Proceedings/]

May, D. A., M. M. Parmeter, D. S. Olszewski, and B. D. McKenzie, 1998: Operational processing of satellite sea surface temperature retrievals at the Naval Oceanographic Office. *Bull. Amer. Meteor. Soc.*, **79**, 397–407.

McClain, E. P., 1989: Global sea surface temperatures and cloud screening for aerosol optical depth estimates. *Int. J. of Remote Sens.*, **10**, 763-769.

McClain, E. P., W. G. Pichel, and C. C. Walton, 1985: Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophys. Res.*, **90**, 11,587-11,601.

Merchant, C. J., P. Le Borgne, A. Marsouin, and H. Roquet, 2008: Optimal estimation of sea surface temperature from split-window observations. *Remote Sens. Environ.,* **112**, 2469-2484.

Merchant, C. J., and A. R. Harris, 1999: Towards the elimination of bias in satellite retrievals of sea surface temperatures, 2. Comparison with in situ measurements. *J. Geophys. Res.*, **104**, 23,579-23,590.

Minnett, P. J., R. O. Knuteson, F. A. Best, B. J. Osborne, J. A. Hanafin, and O. B. Brown, 2001: The Marine-Atmospheric Emitted Radiance Interferometer: A high-accuracy, seagoing infrared spectroradiometer. *J. Atmos. Oceanic Technol.*, **18***,* 994-1013.

Mittaz, J. P. D., A. R. Harris, and J. T. Sullivan, 2009: A physical method for the calibration of the AVHRR/3 thermal IR channels 1: the prelaunch calibration data. *J. Atmos. Oceanic Technol.*, **26***,* 996-1019.

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

O'Carroll, A. G., R. W. Saunders, and J. G. Watts, 2006a: The measurement of the sea surface temperature by satellites from 1991 to 2005. *J. Atmos. Oceanic Technol.*, **23***,* 1573-1582.

O'Carroll, A. G., J. G. Watts, L. A. Horrocks, R. W. Saunders, and N. A. Rayner, 2006b: Validation of the AATSR meteo product sea surface temperature. *J. Atmos. Oceanic Technol.*, **23**, 711-726.

Ostle, B., and L. Malone, 1988: Statistics in Research. Iowa State University Press, 664pp.

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. Climate*, **20**, 5473-5496.

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

Stark, J. D., C. Donlon, A. O'Carroll, and G. Corlett, 2008: Determination of AATSR biases using the OSTIA SST analysis system and a matchup database. *J. Atmos. Oceanic Technol.*, **25**, 1208-1217.

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch, 2007: OSTIA: An operational, high resolution, real time, global sea surface temperature analysis system. Extended abstract *Oceans '07 IEEE Aberdeen, Marine challenges: coastline to deep sea*, 18-22 June, 2007, Aberdeen, Scotland.

Stuart-Menteth, A. C., I. S. Robinson, and C. J. Donlon, 2005: Sensitivity of the diurnal warm layer to meteorological fluctuations. Part 2: A new parameterization for diurnal warming. *Int. J. of Remote Sens.*, **10**, 209-234.

Suarez, M. J., W. J. Emery, and G. A. Wick, 1997: The multi-channel infrared sea truth radiometric calibrator (MISTRC). *J. Atmos. Oceanic Technol.*, **14**, 243-252.

Thiébaux, J., E. Rogers, W. Wang, and B. Katz, 2003: A New High-Resolution Blended Real-Time Global Sea Surface Temperature Analysis. *Bull. Amer. Meteor. Soc.*, **84**, 645-656.

Tietjen, G. L., 1986: The analysis and detection of outliers, *Goodness-of-fit techniques*. 2d ed. R.B. D'Agostino and M.A. Stephens, Editors, Marcel Dekker Inc, 497–522 pp.

Trishchenko, A. P., 2002: Removing unwanted fluctuations in the AVHRR thermal calibration data using robust techniques. *J. Atmos. Oceanic Technol.*, **19***,* 1939-1954.

Vázquez-Cuervo, J., E. Armstrong, and A. Harris, 2004: The effect of aerosols and clouds on the retrieval of satellite derived infrared sea surface temperatures. *J. Climate*, **17**, 3921-3933.

Walton, C. C., W. G. Pichel, and J. F. Sapper, 1998: The development and operational applications of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *J. Geophys. Res.*, **103**, 27,999-28,012.

Walton, C. C., 1988: Nonlinear multichannel algorithms for estimating sea surface temperature with AVHRR satellite data. *J. of Applied Meteorology*, **27**, 115-124.

Wang, L., and X. Wu, 2008: GSICS tools used to compare IASI and AIRS. *GSICS Quarterly*, **2**, *4*, 4. [Available online at http://www.star.nesdis.noaa.gov/smcd/spb/calibration/icvs/GSICS/documents/newsletter/GSICS_Quarterly_Vol2No4_2008.pdf]

Tables

Table 1: Operational regression equations and coefficients for sea surface temperature retrievals from AVHRRs onboard NOAA and MetOp-A platforms, in the NESDIS heritage Main Unit Task (MUT).

Night equation:	$SST = a_0 + a_1 T_4 + a_2 T_3 + a_3 T_5 + a_4 (T_3 - T_5)(\sec\theta - 1) + a_5 (\sec\theta - 1)$					
	a_0	a_1	a_2	a_3	a_4	a_5
NOAA-19	-275.732	0.37345	1.12512	-0.48501	0.12493	1.36803
MetOp-A	-273.205	0.29797	1.21294	-0.505499	1.52873	0.10867
NOAA-18	-274.686	0.46757	1.08556	-0.543265	0.13763	1.12622
NOAA-17	-275.456	0.57317	1.12933	-0.690623	0.07219	1.66172
NOAA-16	-274.875	0.25749	1.25364	-0.502818	0.11060	1.12932
Day equation:	$SST = a_0 + a_1 T_4 + a_2 T_{sfc} (T_4 - T_5) + a_3 (T_4 - T_5)(\sec\theta - 1)$					
	a_0	a_1		a_2	a_3	
NOAA-19	-259.864	0.95606		0.06340	0.83725	
MetOp-A	-256.746	0.94599		0.08391	1.01458	
NOAA-18	-253.308	0.93400		0.07245	0.74804	
NOAA-17	-253.951	0.93605		0.08387	0.92085	
NOAA-16	-247.389	0.91128		0.08088	0.71744	

Here, a_n (n = 1…5) are the coefficients; T_3, T_4, and T_5 are AVHRR brightness temperatures (K) in channels 3B, 4, and 5, respectively, θ is the view zenith angle, and T_{sfc} is a first-guess SST (in MUT, taken from the nearest 100km analysis gridpoint of satellite SST from last 24 hrs).

Table 2: List of analyses SST products used in SQUAM as reference SSTs.

Product	Space/Time resolution	Abbreviation and Type	Reference	Availability period, data format, and ftp source,	Input source
OISST	1.00° Weekly	WOI Bulk	Reynolds *et al.*,2002	1981 to present, Raw binary *ftp.emc.ncep.noaa.gov* */cmb/sst/oisst_v2*	IR: AVHRR (NOAA-17,18 & earlier platforms) MW: --- In situ: Ships, Drifting & Moored buoys
	0.25° Daily	DOI_AV Bulk	Reynolds *et al.*,2007	1985 to present, netCDF *eclipse.ncdc.noaa.gov* */pub/OI-daily-v2/NetCDF*	IR: AVHRR (NOAA-17,18 & earlier platforms) (Pathfinder until Dec-2005, then NAVOCEANO) MW: --- In situ: Ships, Drifting & Moored buoys
		DOI_AA Bulk		Jun-2002 to present, netCDF *eclipse.ncdc.noaa.gov* */pub/OI-daily-v2/NetCDF*	IR: AVHRR (NOAA-17,18 & earlier platforms) (Pathfinder until Dec-2005, then NAVOEANO) MW: AMSR-E (Aqua) In situ: Ships, Drifting & Moored buoys
RTG SST	0.50° Daily	RTG_LR Bulk	Thiébaux *et al.*,2003	Dec-2000 to present, gridded binary (grib) *polar.ncep.noaa.gov* */pub/history/sst*	IR: AVHRR (NOAA-17,18 & earlier platforms) MW: --- In situ: Ships, Drifting & Moored buoys
	1/12° Daily	RTG_HR Bulk	Gemmill, Katz, & Li,2007	Feb-2007 to present, grib *polar.ncep.noaa.gov* */pub/history/sst/ophi* (rotated for last one year)	IR: AVHRR (NOAA-17,18 & earlier platforms) (a newer Physical SST retrieval scheme) MW: --- In situ: Ships, Drifting & Moored buoys
Oper. SST & Sea Ice Analysis	0.05° Daily	OSTIA Foundation	Stark *et al.*,2007; 2008	Apr-2006 to present, netCDF *podaac.jpl.nasa.gov* */GHRSST/data/L4/GLOB/UKMO/OSTIA*	IR: AATSR (Envisat), AVHRR (NOAA-17,18), SEVIRI (Meteosat-8) MW: Aqua AMSR-E, TRMM TMI, DMSDP SSMI ice In situ: Ships, Drifting & Moored buoys
Ocean Data Analysis, Mersea/ GMES	0.10° Daily	ODYSSEA Subskin	Autret & Piollé,2007	Oct-2007 to present, netCDF *ftp.ifremer.fr/ifremer/medspiration* */data/l4hrsstfnd/eurdac/glob/odyssea*	IR: AATSR (Envisat), AVHRR (NOAA-17,18), VISSR (GOES-12), SEVIRI (Meteosat-8) MW: AMSR-E (Aqua), TMI (TRMM) In situ: ---
monthly mean climate SST	1.00° Monthly average (1969-1985)	CLISST Bulk	Bauer & Robinson, 1985	Monthly mean for 1969 to 1981, ASCII http://dss.ucar.edu /datasets/ds278.0	In situ: primarily Mechanical Bathythermograph (MBT)/hydrocast

Figure captions

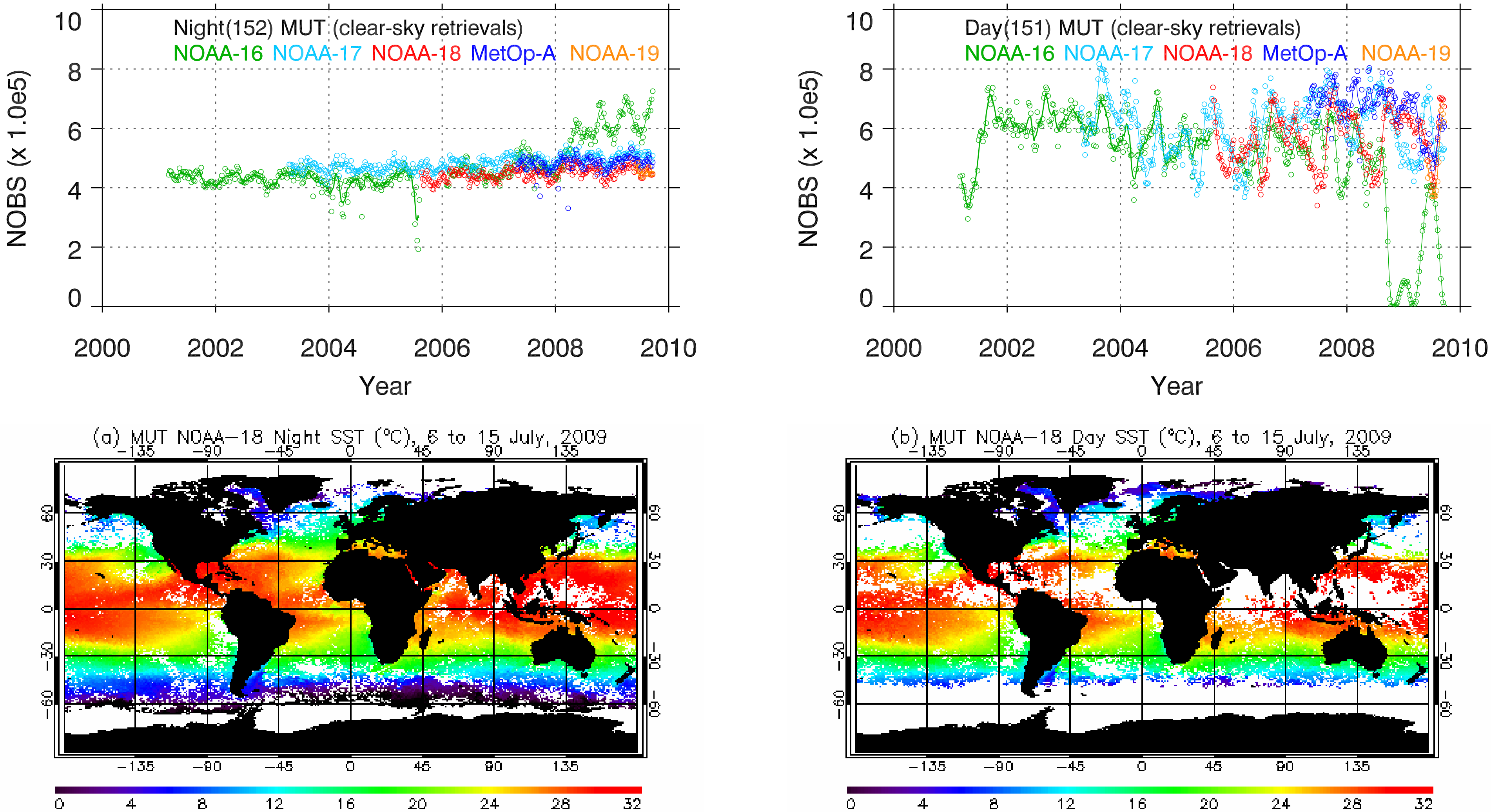


Fig. 1: Top panels show time series of numbers of clear-sky observations (NOBS) in NESDIS heritage Main Unit Task (MUT) SSTOBS weekly files for five platforms. MUT retrievals are restricted to ±53° view zenith angle. Bottom panels show AVHRR SST (°C) from NOAA-18 for 6-15 July 2009 sampled at 0.25º spatial resolution; (a) night, (b) day.

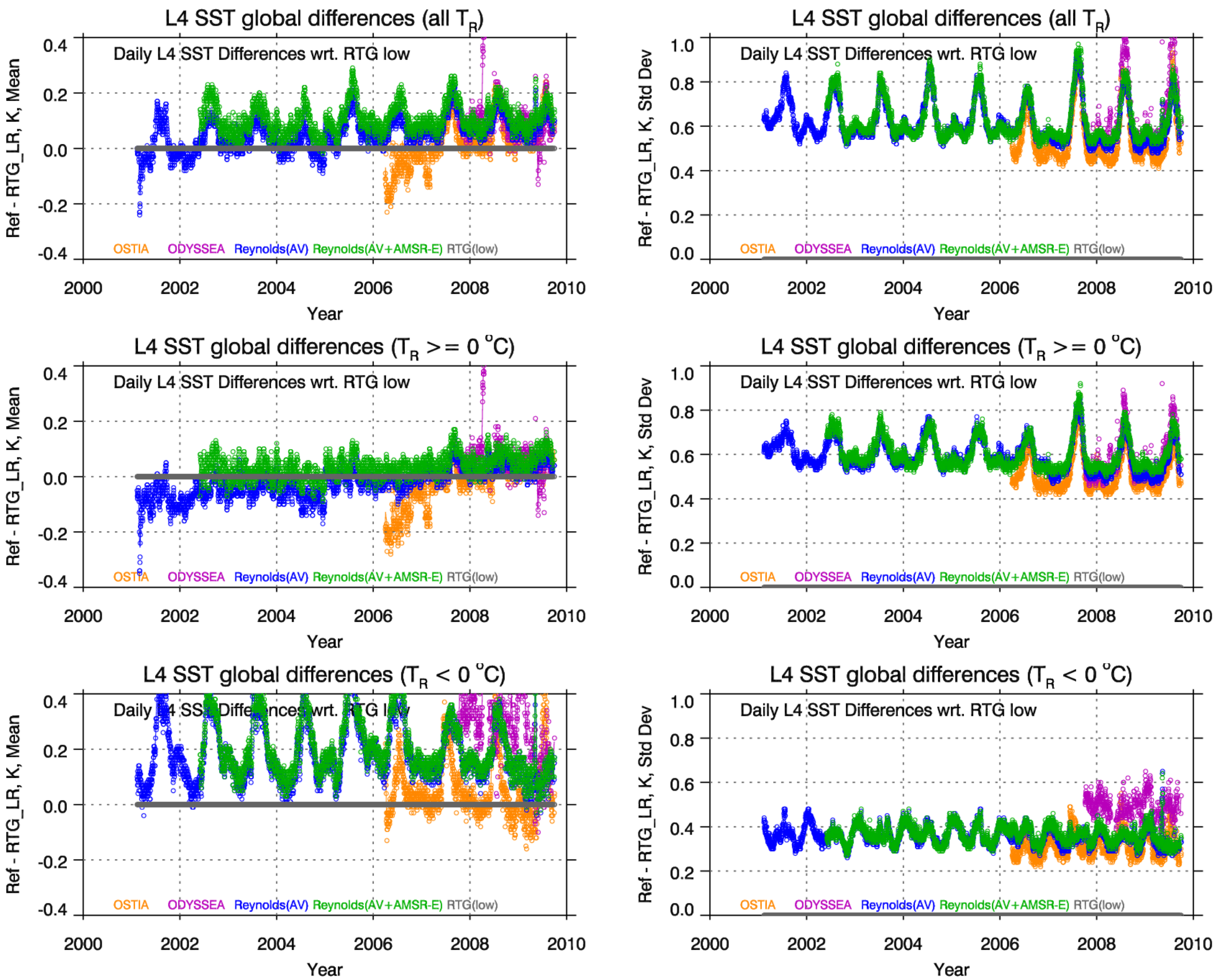


Fig. 2: Global mean differences between several reference SST fields w.r.t. low resolution RTG SST. Left panels: mean; Right panels: standard deviation; Top panels: full range of SSTs, Middle panels: T_R and RTG_LR ≥ 0 °C, and Bottom panels: T_R and RTG_LR < 0 °C.

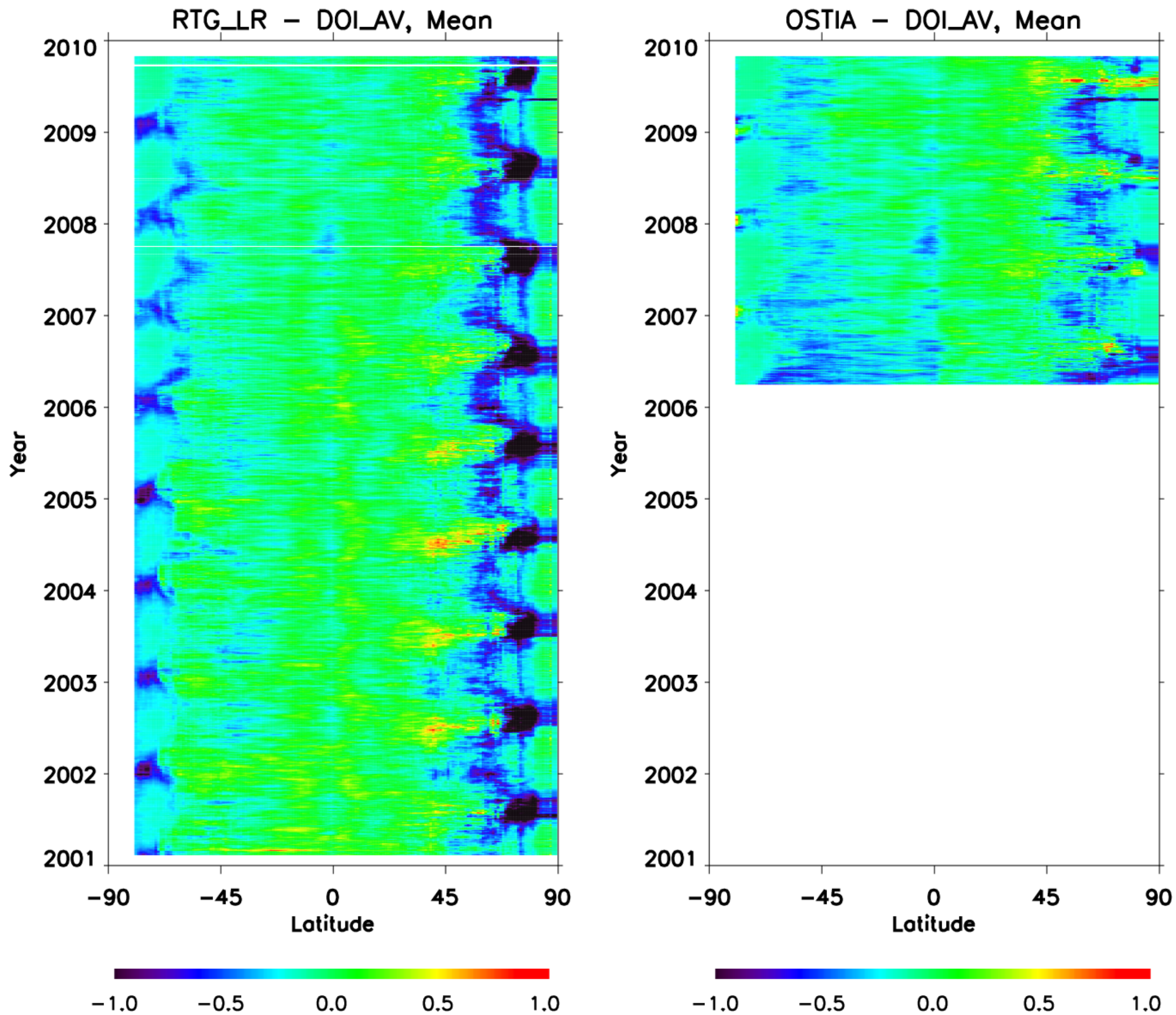
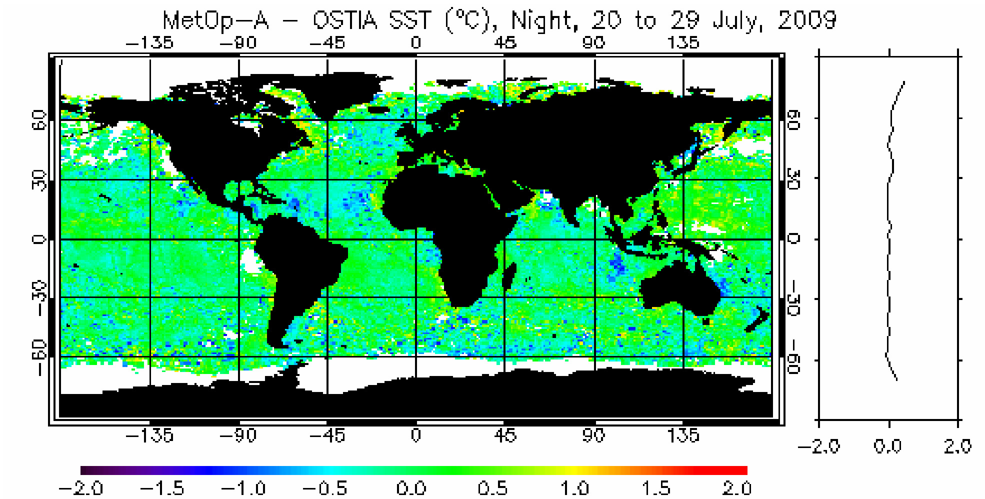


Fig. 3: Hovmöller diagrams of average zonal differences in T_R SST fields (T_R *minus* DOI_AV) for T_R = RTG_LR and OSTIA.



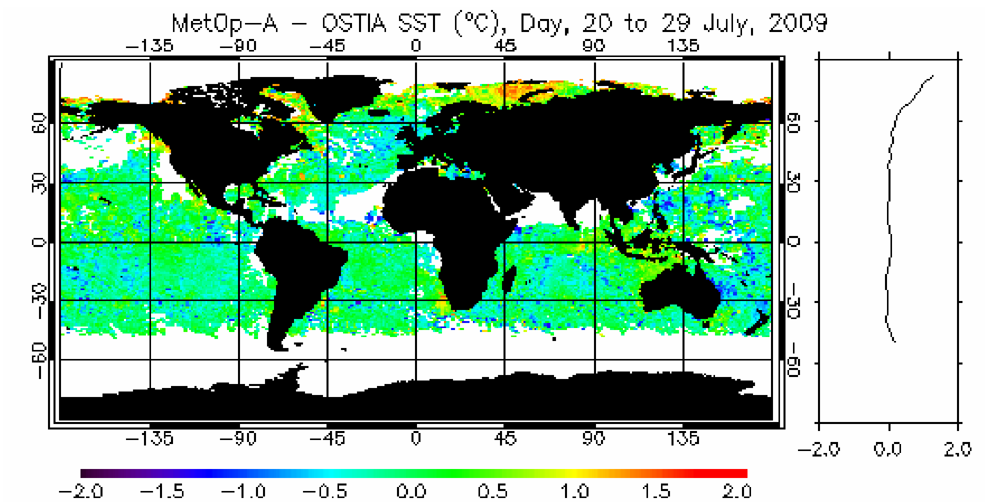
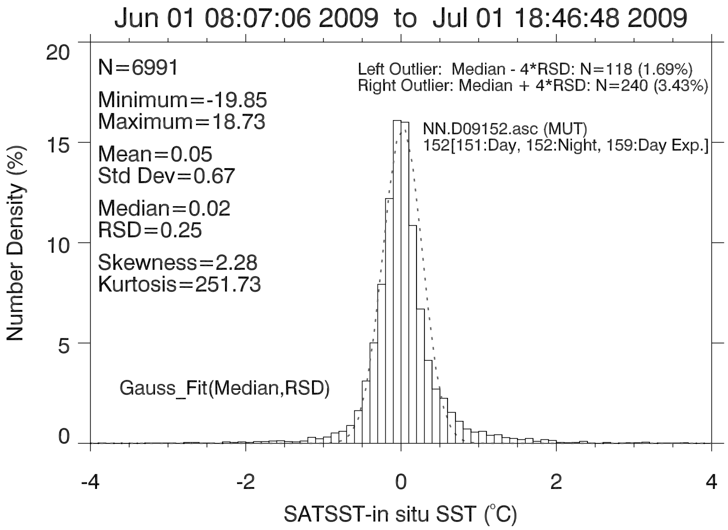
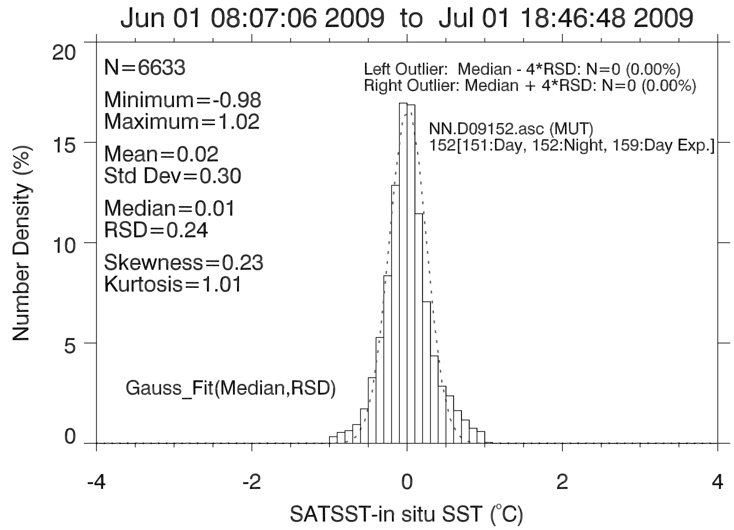


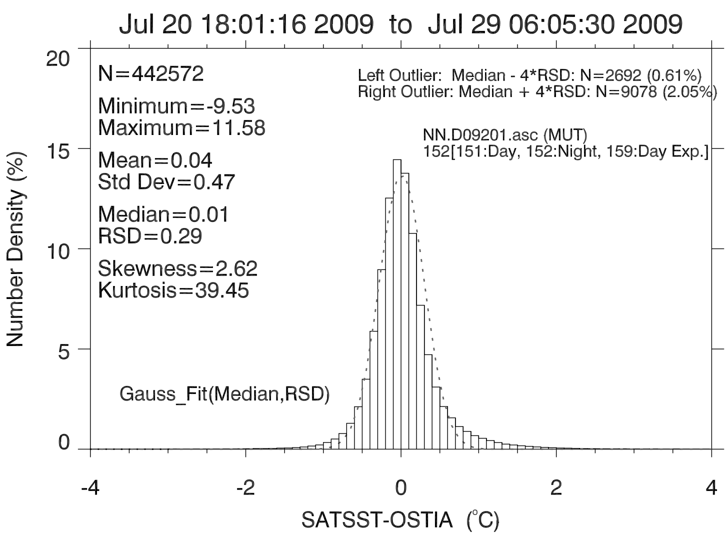
Fig. 4: Global maps of $\Delta T_S = T_R - T_{OSTIA}$ for MetOp-A, from 20-29 July 2009; top: night, bottom: day.



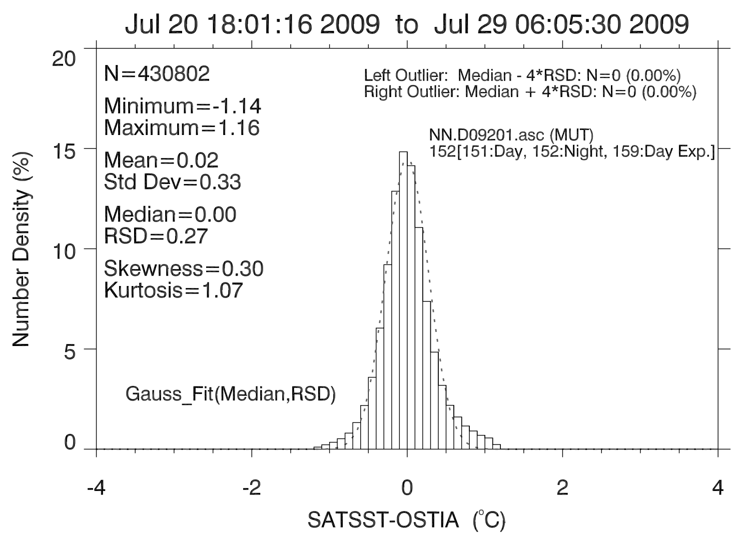
a) NOAA18 - *in situ* SST, outliers present



b) NOAA18 - *in situ* SST, outliers removed



c) NOAA18 - OSTIA SST, outliers present



d) NOAA18 - OSTIA SST, outliers removed

Fig. 5: Nighttime global histograms of ΔT_S for NOAA-18 w.r.t. different reference SSTs [Top panels: $T_R = T_{in\ situ}$; Bottom panels: $T_R = T_{OSTIA}$; Left panels: before removal of outliers; Right panels: after removal of outliers]. ΔT_S statistics are annotated on the left side of the histograms. Dotted gray line shows an ideal Gaussian fit, *N(median, RSD)*. The numbers of outliers (retrievals beyond "Median ± 4×RSD") are shown on the top right. The start and end time of the satellite SST data are shown on the top. The name of the SSTOBS file processed (*e.g.*, NN.D09021.asc) contains platform information (NN stands for NOAA-18) and the start day-of-year of the SSTOBS file. Below filename, the data type is also shown: day (MUT code "151"), night ("152"), or day with relaxed cloud test ("159").

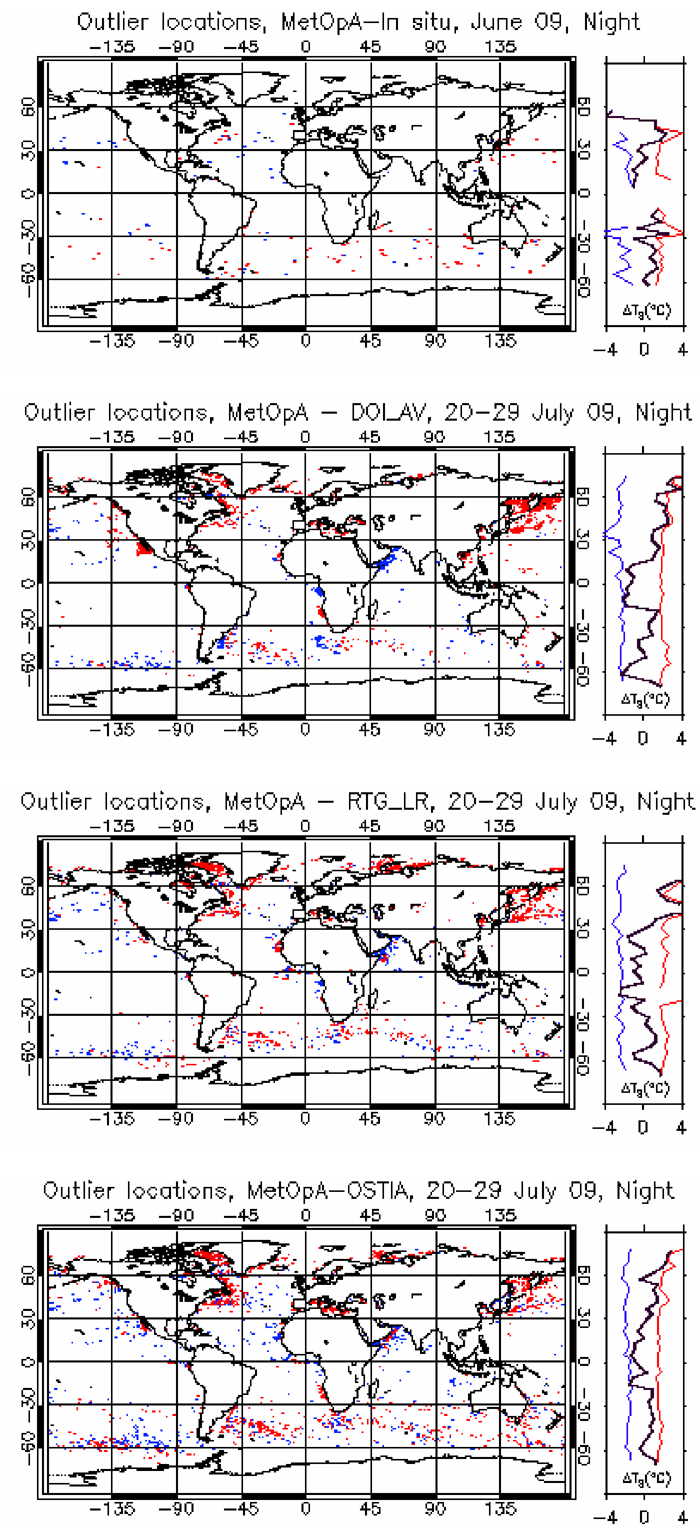


Fig. 6: Locations of nighttime low (blue) and high (red) outliers for ΔT_S (MetOp-A – reference), for June 2009 (*in situ*) and 20–29 July 2009 (global fields). Image pixels are shown larger than their actual size for enhanced visualization. Adjacent plots show zonal distribution for cold only (blue), warm only (red) and total (black) ΔT_S corresponding to outliers.

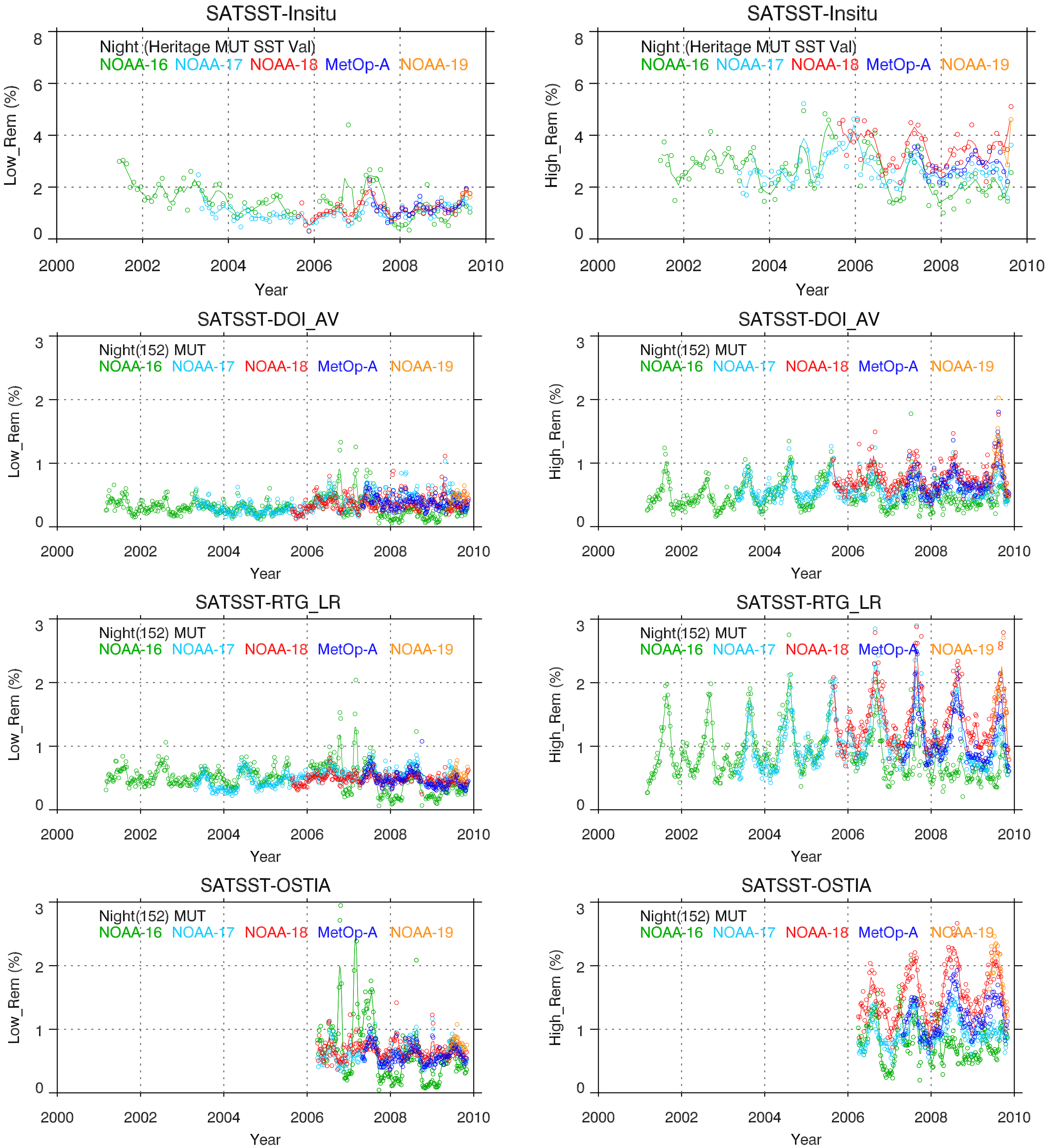


Fig. 7: Time series of nighttime outliers (left panels: low, right panels: high) for several reference SSTs.

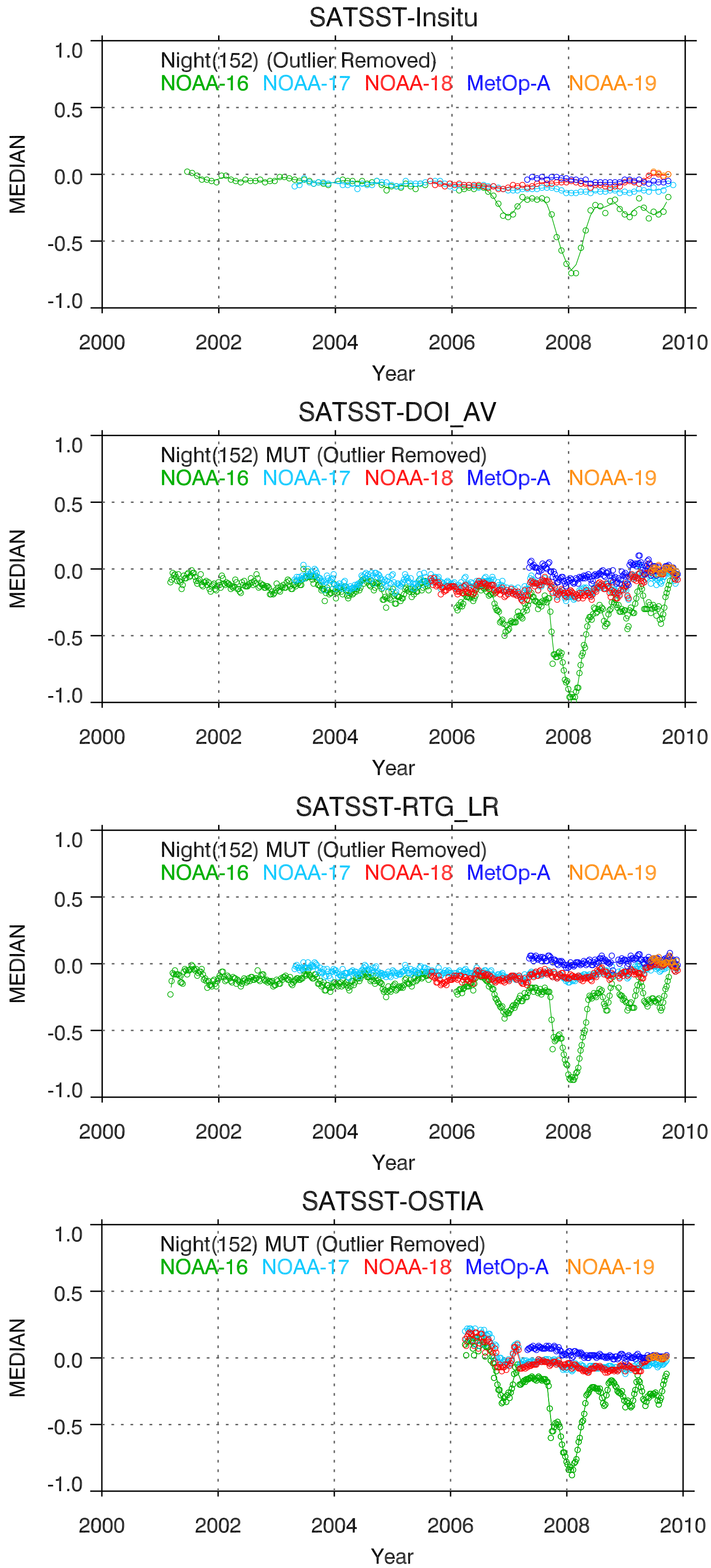


Fig. 8: Time series of nighttime median ΔT_S after outlier removal for four reference SSTs.

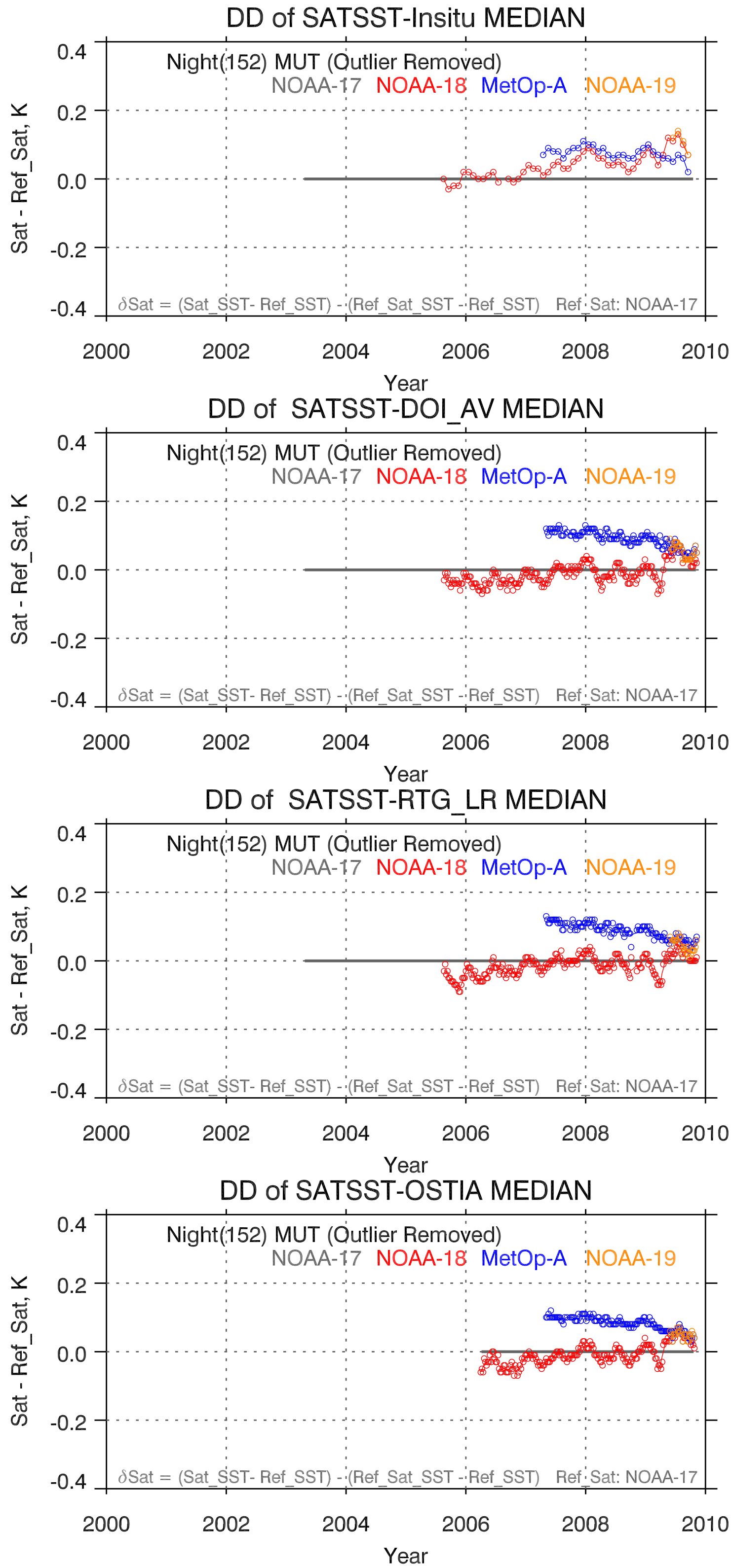


Fig. 9: Cross-platform SST biases derived using double differencing technique for different reference SSTs. $\Delta T_{S,N17}$ was used as a "transfer standard".

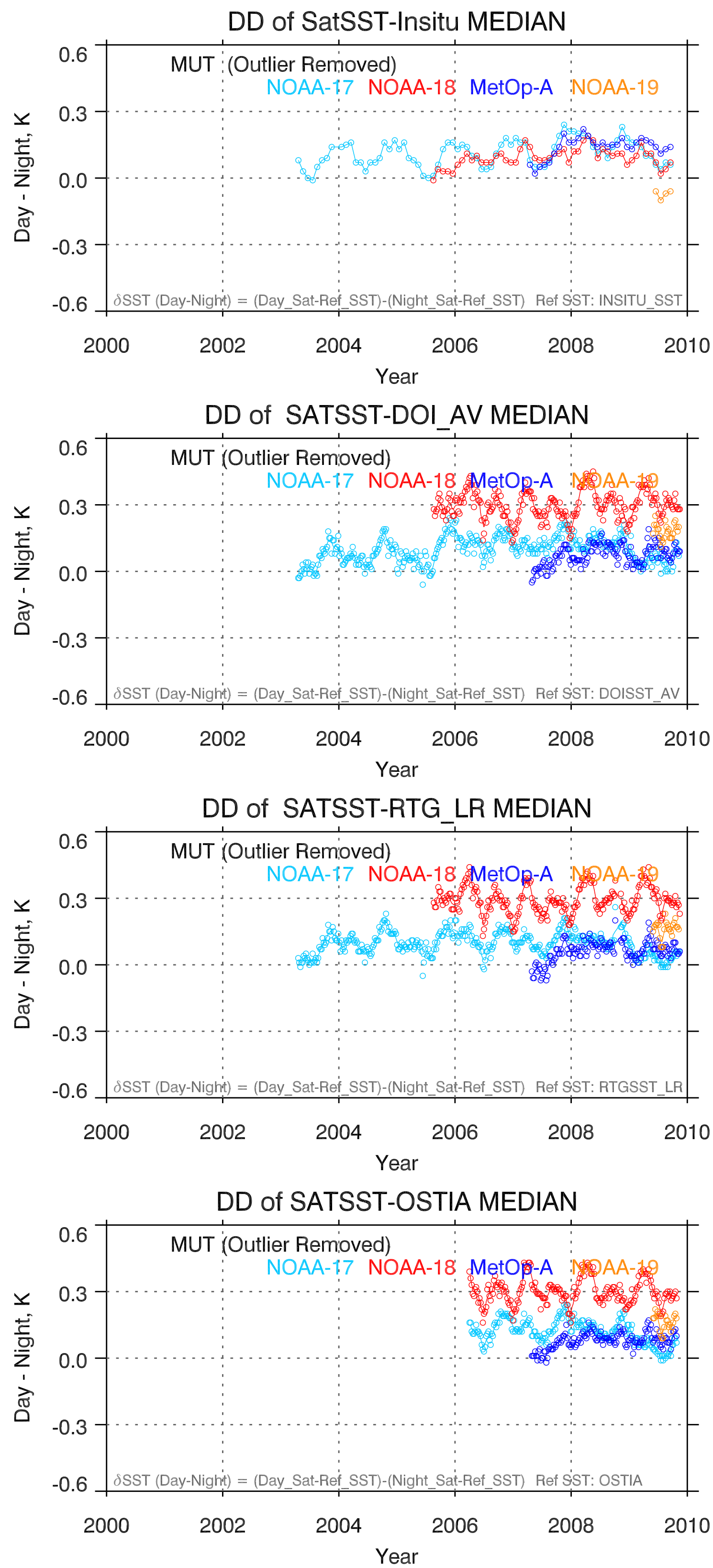


Fig. 10: "Day minus night" SST biases estimated using double differences (DD) for different reference SSTs.

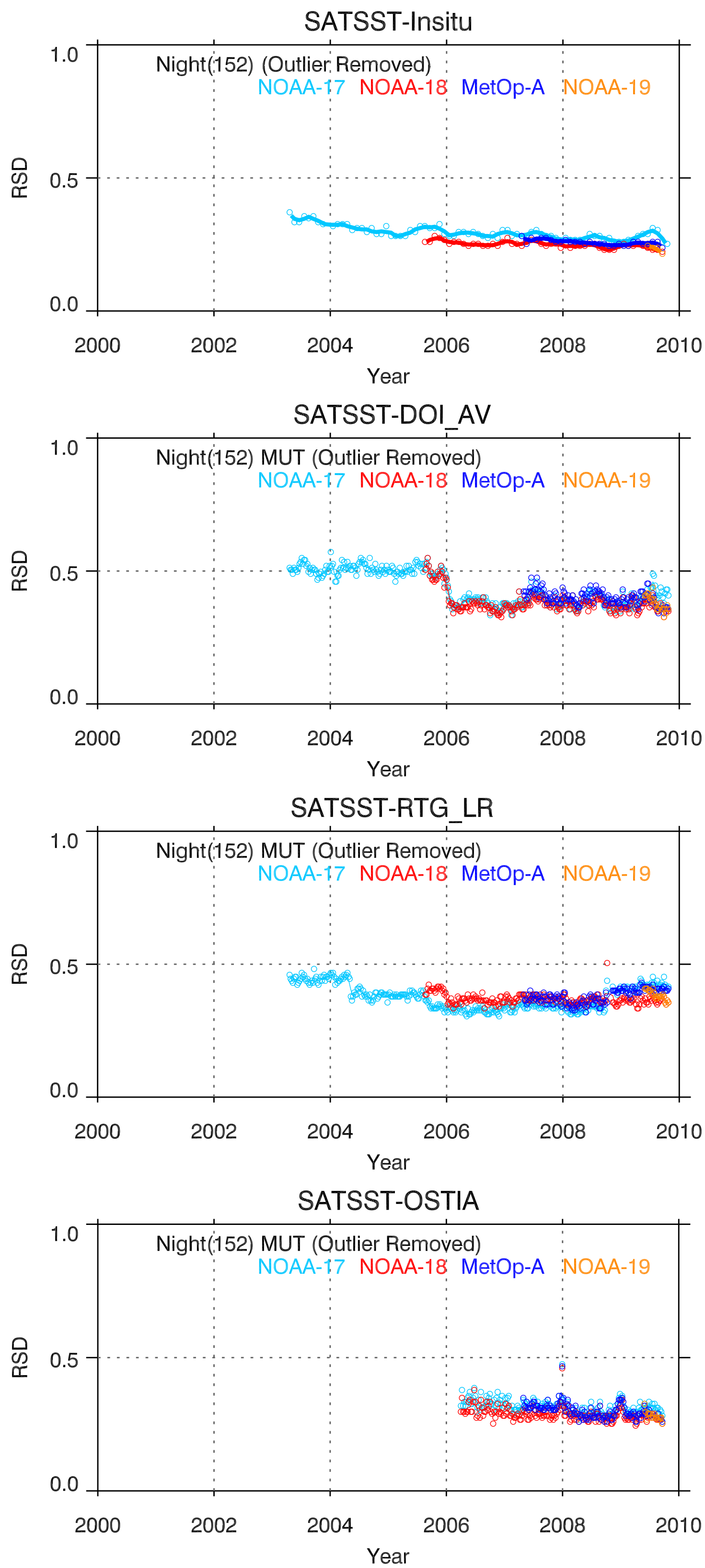


Fig. 11: Time series of the robust standard deviation (RSD) w.r.t. different reference SSTs.

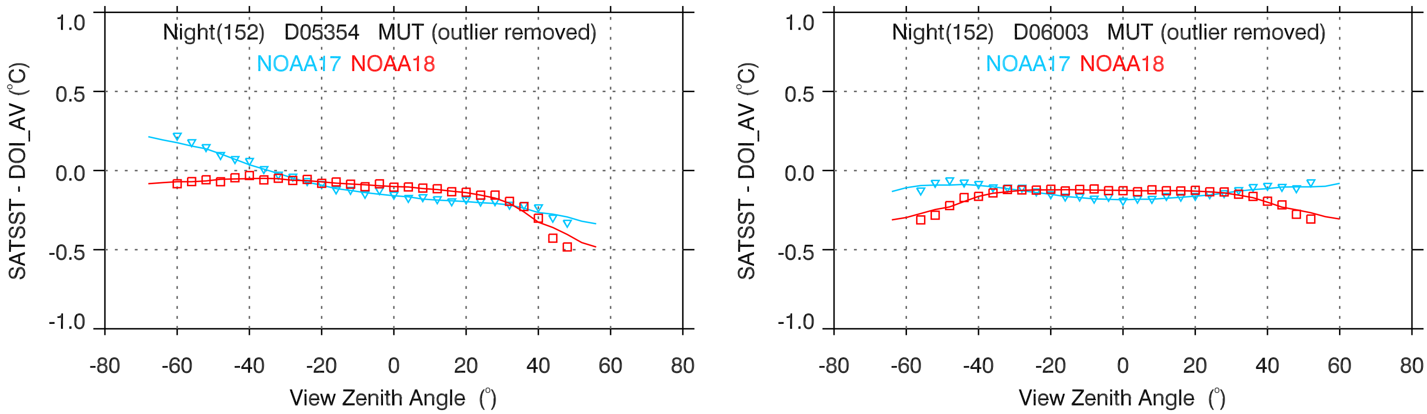


Fig. 12: Angular dependence of ΔT_S for NOAA-17 and -18 before (left) and after (right) resolving the view zenith angle issue in the MUT system.