Group for High Resolution SST (GHRSST) Analysis Fields Inter-Comparisons
Part 2. Near real time web-based Level 4 SST Quality Monitor (L4-SQUAM)

Prasanjit Dash[1,2], Alexander Ignatov[1], Matthew Martin[3], Craig Donlon[4], Bruce Brasnett[5], Richard W. Reynolds[6], Viva Banzon[6], Helen Beggs[7], Jean-Francois Cayula[8], Yi Chao[9], Robert Grumbine[10], Eileen Maturi[1], Andy Harris[1,11], Jonathan Mittaz[1,11], John Sapper[12], Toshio M. Chin[9], Jorge Vazquez[9], Edward M. Armstrong[9], Chelle Gentemann[13], James Cummings[14], Jean-François Piollé[15], Emmanuelle Autret[15], Jonah Roberts-Jones[3], Shiro Ishizaki[16], Jacob L. Høyer[17], Dave Poulter[18]

[1] NOAA/NESDIS, Center for Satellite Application and Research (STAR), Camp Springs, MD, USA

[2] Colorado State Univ., Cooperative Institute for Research in the Atmospheres (CIRA), Fort Collins, CO, USA

[3] Met Office, Exeter, UK

[4] ESA/ESTEC, Ocean and Ice, Noordwijk, The Netherlands, 2201 AZ

[5] Canadian Meteorological Centre, Dorval, Quebec, Canada

[6] NOAA/NCDC/CICS, Asheville, NC, USA

[7] Centre for Australian Weather & Climate Research, Bureau of Meteorology, Melbourne, Australia

[8] QinetiQ North America, Technology Solutions Group, Stennis Space Center, MS 39522, USA

[9] NASA/JPL/Caltech, 4800 Oak grove Drive, Pasadena, CA, USA

[10] NOAA/NWS/NCEP, Camp Springs, MD, USA

[11] Univ. of Maryland, Cooperative Institute for Climate and Satellites (CICS), College Park, USA

[12] NOAA/NESDIS, Office of Satellite Processing & Operations (OSPO), Camp Springs, MD, USA

[13] Remote Sensing Systems, Santa Rosa, CA, USA

[14] Naval Research Laboratory, Monterey, CA, USA

[15] Ifremer, Spatial Oceanography Laboratory, /CERSAT, Brest, France

[16] Japan Meteorological Agency, Tokyo, Japan

[17] Danish Meteorological Institute, Copenhagen Ø, Denmark

[18] National Oceanographic Centre, Southampton, SO14 3ZH, UK

Manuscript submitted to Journal of Deep Sea Research II
Special issue on Satellite Oceanography and Climate Change
Washington DC, Submitted: 12 August 2011

Correspondence:

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

Contributors’ email:

Prasanjit Dash	prasanjit.dash@noaa.gov	NOAA/NESDIS/STAR CSU/CIRA
Alexander Ignatov	alex.ignatov@noaa.gov	NOAA/NESDIS/STAR
Matthew Martin	matthew.martin@metoffice.gov.uk	UK Met Office
Craig Donlon	craig.donlon@esa.int	European Space Agency
Bruce Brasnett	bruce.brasnett@ec.gc.ca	Canadian Meteorological Centre
Richard W. Reynolds	richard.w.reynolds@noaa.gov	NOAA/NESDIS/NCDC CICS
Viva Banzon	viva.banzon@noaa.gov	NOAA/NESDIS/NCDC
Helen Beggs	h.beggs@bom.gov.au	CAWCR/Australian BoM
Jean-Francois Cayula	jean-francois.cayula@QinetiQ-NA.com	QinetiQ NA
Yi Chao	yi.chao@jpl.nasa.gov	NASA/JPL/Caltech
Robert Grumbine	robert.grumbine@noaa.gov	NOAA/NWS/NCEP
Eileen Maturi	eileen.maturi@noaa.gov	NOAA/NESDIS/STAR
Andy Harris	andy.harris@noaa.gov	NOAA/NESDIS/STAR UMD/CICS
Jonathan Mittaz	jon.mittaz@noaa.gov	NOAA/NESDIS/STAR UMD/CICS
John Sapper	john.sapper@noaa.gov	NOAA/NESDIS/OSPO
Toshio M. Chin	mike.chin@jpl.nasa.gov	NASA/JPL/Caltech
Jorge Vazquez	jorge.vazquez@jpl.nasa.gov	NASA/JPL/Caltech
Edward M. Armstrong	edward.armstrong@jpl.nasa.gov	NASA/JPL/Caltech
Chelle Gentemann	gentemann@remss.com	Remote Sensing Systems
James Cummings	cummings@nrlmry.navy.mil	Naval Research Laboratory
Jean-François Piollé	jean.francois.piolle@ifremer.fr	Ifremer/CERSAT/
Emmanuelle Autret	emmanuelle.autret@ifremer.fr	Ifremer/CERSAT/
Jonah Roberts-Jones	jonah.roberts-jones@metoffice.gov.uk	UK Met Office
Shiro Ishizaki	s_ishizaki@met.kishou.go.jp	Japanese Meteorological Agency
Jacob L. Høyer	jlh@dmi.dk	Danish Meteorological Institute
Dave Poulter	david.poulter@noc.soton.ac.uk	National Oceanographic Centre, UK

Table of contents

Abstract

NESDIS has established a near real-time web based SST Quality Monitor (SQUAM; *http://www.star.nesdis.noaa.gov/sod/sst/squam/*). The initial objective of SQUAM was to monitor NESDIS AVHRR Level 2 (L2; data at the observed pixels) SST products. Subsequently, following the interest from the Level 4 (L4; gap-free gridded data) SST community, and in the spirit of the Group for High Resolution Sea Surface Temperature (GHRSST; *http://www.ghrsst.org/*) Inter comparison Technical Advisory Group (IC-TAG; *https://www.ghrsst.org/ghrsst-science/science-team-groups/ic-tag/*) collaborative efforts, SQUAM functionality has been extended to include cross-comparisons of various L4 SST products. The L4-SQUAM (*http://www.star.nesdis.noaa.gov/sod/sst/squam/L4/*), described here, in Part 2 of this three-part paper, complements the GHRSST Multi Product Ensemble (GMPE) and High Resolution Diagnostic Data-set (HR-DDS) systems, documented in Parts 1 and 3 of this paper, respectively.

The L4-SQUAM is aimed at serving the needs of both L4 users and producers. It performs quasi near real-time monitoring of thirteen L4 products, with 1-day latency, while retaining their full history. Analyses of "L4 *minus* L4" SST differences are performed by plotting global maps, histograms, time series and Hovmöller diagrams, for all available combinations of L4 products. The emphasis is on quantitative comparisons in a global domain. Additionally, all L4 products are consistently compared with quality controlled *in situ* data (drifting buoys, ships, and coastal and tropical moorings), available from the NESDIS *in situ* SST Quality Monitor (*i*QUAM; *http://www.star.nesdis.noaa.gov/sod/sst/iquam/*), and "L4 *minus in situ*" statistics are analyzed the same way as "L4 *minus* L4".

Currently, the following daily L4 SSTs are monitored in L4-SQUAM: two Reynolds

OISST (AVHRR, AVHRR+AMSR-E), two OSTIA (operational and reanalyzed), two RTG (high and low resolution), NAVO K10, NESDIS Multi-SST, JPL G1SST, CMC 0.2°, ODYSSEA, BoM GAMSSA, and GMPE product. Work is underway to add JPL MUR, RSS -MW and -IR+MW, NRL NCODA, JMA MGDSST, and DMI analysis.

The largest differences between various L4 SSTs are typically observed in high latitudes, partly due to different treatment of the sea-ice transition zone. When an ice flag is available, the inter-comparisons are performed in two ways: including and excluding ice grids. Differences are also observed in coastal areas, as well as in many open ocean areas. These large differences call for a community effort to understand and reconcile them. Some L4 products tend to cluster together and form groups (such as the OSTIA and the CMC 0.2º, which also agree well with the GMPE, or alternatively the RTG high resolution and the NESDIS Multi-SST analysis). Few products cover the full AVHRR era (1981-on), while many products have only several years of data. Their extension back in time would reduce uncertainty in the historical SST data, and provide a more reliable first-guess SST for improved cloud detection and SST retrievals in satellite data reprocessing efforts, such as the Pathfinder Ocean or ESA's Climate Change Initiative (*http://www.esa-cci.org/*), aimed at generating improved SST Climate Data Records (CDRs). SQUAM, along with near real-time monitoring of L4-products, provides a framework for climate research and applications by performing evaluation of such SST CDRs.

Keywords: *Sea surface temperature, Intercomparison, Climatic data, Sea ice, Data centers*

1. Introduction

Satellite-based sea surface temperature (SST) products have been operationally derived from low earth orbiting (LEO) and geostationary (GEO) platforms, initially at NESDIS and subsequently at other agencies (*e.g.*, McClain *et al.*, 1985; Walton, 1988; Walton *et al.*, 1998; May *et al.*, 1998; Wu *et al.*, 1999; Kilpatrick *et al.*, 2001; Brisson *et al.*, 2002; LeBorgne *et al.*, 2007; Maturi *et al.*, 2008). Satellite Level 2 (L2) products are derived from Level 1B (L1B) brightness temperatures and may be further processed into Level 3 (L3) products. These L2 and L3 products are used for a variety of meteorological and oceanographic applications, but their potential is limited due to data gaps caused by satellite scan geometry, cloud coverage, *etc*. Therefore, efforts at various data centers have been directed towards generating global, gridded, blended, gap-free SST fields with attached error statistics, known as Level 4 (L4) SSTs. In addition to various L2 SSTs from multiple sources, many L4 products also use *in situ* data, and blend them together using various interpolation techniques (*cf.*, Martin *et al.*, 2011, Part 1). Resulting global L4 fields provide information crucial for a variety of real-time and research applications, including seasonal and short-term weather forecasting, fisheries and coral-reef monitoring requiring temperature and temperature anomalies which affect aquatic health, and for developing SST retrieval algorithms employing radiative transfer simulations which require a first-guess SST. The L4 SSTs, in particular those with a longer history, are invaluable for generating Climate Data Records (CDRs) and their retrospective and near real-time monitoring are crucial for monitoring climate changes.

Because various applications have different requirements for a global L4 SST product, about twenty L4 products have been developed worldwide. This fast progress has posed an additional challenge and requirement to understand their relative merit and performance, in terms of data coverage, resolution and accuracy. An "L4 inventory" with comparison tools could assist the data users to choose a product appropriate for their applications, as well as provide feedback to the data producers and help them improve and reconcile their products.

One would think that, ideally, an L4 product should optimally blend multiple satellite and *in situ* SSTs into one "true" SST. However, it has become apparent that significant differences exist between various products, especially in high latitudes and in coastal areas, as well as often in the open ocean. Such differences are also significant in areas of warm Western boundary currents and in semi-enclosed basins such as the Mediterranean and the Gulf of California. In the time series of global statistics, some products may cluster in groups, *e.g.*, the foundation SST (the SST free of diurnal warming) family, while significant differences may be observed between different groups. These primary reasons for such differences may be attributed to: (a) developing specific L4 SSTs for specific applications, depending on prevailing requirements and resources in corresponding data centers, (b) use of different input data (satellite infrared, microwave and *in situ* SSTs) of varying space-time resolutions, quality, cloud-masks, and quality control (QC) procedures, (c) use of different blending and optimal interpolation methods and multiple correlation lengths, (d) different representations of SST (skin, depth, foundation *etc*.) and feature resolutions and (e) non-uniform treatment of land-sea and ice masks.

These challenges have been acknowledged by the Group for High Resolution SST (GHRSST; *http://www.ghrsst.org/*), which formed the Inter-Comparison Technical Advisory Group (IC-TAG; *https://www.ghrsst.org/ghrsst-science/science-team-groups/ic-tag/*) to facilitate cross-evaluation of L4 SSTs. Today, the IC-TAG comprises three major near real-time web-based systems: the GHRSST Multi Product Ensemble (GMPE; Part 1, Martin *et al.*, 2011), the Level-4 SST Quality Monitor (L4-SQUAM; Part 2, this study) and the High-

Resolution Diagnostic Data Set (HR-DDS; Part 3, Poulter *et al.*, 2011). The major objective of Part 2 is to document the L4-SQUAM system and illustrate its functionalities, highlighting its potential to quickly evaluate the consistency between various L4 fields.

As of this writing, thirteen L4 fields are monitored in L4-SQUAM, and work is underway to include the remaining fields (see Section 2.1). The L4-SQUAM is an extension of the L2-SQUAM described in Dash *et al.* (2010). It automatically calculates "L4 *minus* L4" differences for all available product combinations, within ~24 hours of their availability, and plots global maps, histograms, time series and Hovmöller plots of SST differences. Also, to understand and reconcile ice mask differences, analyses in L4-SQUAM are performed two ways, both "including" and "excluding" ice masks, when corresponding ice flags are available in the product. The resulting diagnostics are posted at *http://www.star.nesdis.noaa.gov/sod/sst/squam/L4/*. The primary motivation for L4-SQUAM was near real-time (NRT) monitoring, but retrospective diagnostics are also calculated and posted on the web, and the full available time series are analyzed every time a newer product is included in the processing stream.

Besides L4 cross-comparisons, all products are also validated against uniformly quality controlled *in situ* data available from the NESDIS *in situ* SST Quality Monitor (*i*QUAM; *http://www.star.nesdis.noaa.gov/sod/sst/iquam/*). This validation may not be fully independent as many L4 SSTs use *in situ* data in their blending methods. However, having consistent validation statistics against the same data provides an easy way to compare all products. Ideally, the products should be validated against an independent data source, *e.g.*, Argo floats (*e.g.*, Part 1, Martin *et al.*, 2011) or ship-borne infrared radiometers (Donlon *et al.*, 1998; Minnett *et al.*, 2001; Donlon *et al.*, 2011). The advantage of adding independent Argo data to an "*in situ* inventory", such as the *i*QUAM, has been recognized by its developers (*cf.*, Xu and Ignatov, 2010) and will be explored in the future. However, there is no publicly available community-consensus radiometer dataset for use in such validation.

The paper is organized as follows: Section 2 describes the L4-SQUAM concept, system, and the L4 SST fields monitored in it. Inter-comparison results and other observations are discussed in Section 3. Potential extensions of L4-SQUAM are explored in Section 4. Section 5 summarizes and concludes the paper and provides an outlook for the future.

2. The L4-SQUAM concept and system

The basic premise of L4-SQUAM is that differences, ΔT_S="$L4_i$–$L4_j$" or "$L4_i$–*in situ*", are centered about zero and distributed near-normally (see discussion for L2-SQUAM in Dash *et al.*, 2010). The first several moments of the distribution (mean, standard deviation, skewness and kurtosis) are used as a measure of the proximity of the two products and monitored in L4-SQUAM.

2.1. Daily L4 SST fields monitored in L4-SQUAM

Currently, the following daily L4 SST fields are monitored in L4-SQUAM: two NOAA daily OISST (AVHRR, AVHRR+AMSR-E) as described in Reynolds et al. (2007), referred herein as DOI_AV and DOI_AA, respectively, two OSTIA (operational and retrospectively reanalyzed), two RTG (high and low resolution, referred herein as RTG_HR and RTG_LR, respectively), NAVO K10, NESDIS Multi-SST analysis, JPL G1SST, CMC 0.2°, ODYSSEA, BoM GAMSSA and GMPE products. Also, JPL MUR and RSS MW are being processed and work is underway to include the remaining L4 products: RSS IR+MW, NRL NCODA, JMA MGDSST and DMI analysis (see Table 1 for product details). Many of the products included in L4-SQUAM are also included in GMPE and described in Part 1 by Martin *et al.* (2011). However, there are some differences between the GMPE and L4-SQUAM inputs, and we list in Table 1 the products monitored in L4-SQUAM.

INSERT TABLE 1 ABOUT HERE

The SST products listed in Table 1 have either been developed within the GHRSST framework (except the RTG low resolution product) or comply with its standards and specifications. As per the GHRSST specifications, SSTs are categorized into one of the following types: interface, skin, sub-skin, depth and foundation (Donlon *et al*., 2007). Accordingly, for each of the L4 SSTs listed in Table 1, the type is also shown. (Note that the Reynolds and RTG SSTs are adjusted to *in situ* SST and therefore are often referred to as "bulk" SSTs; however, this term is not recommended by the GHRSST.) The OSTIA, CMC, GAMSSA, G1SST, MUR, RSS, MGDSST, ODYSSEA and DMI products are referred to as "foundation SSTs" as they minimize the effect of diurnal thermocline by using either only nighttime satellite data, or additionally daytime data with wind speed above 6 ms^{-1}, or otherwise excluding L2 SSTs with high diurnal variation. The input data to all L4 products are also listed in Table 1, along with information about ice masks and ice bit values to exclude them from statistical analyses. Note that some products are reported with ice masks applied but do not provide bit information to identify those grid cells (*e.g.*, GMPE, CMC), whereas other products do not include ice mask (*e.g.*, NAVO K10, DMI). Also, some products did not have an ice mask included in the initial stage of production, but subsequently added it (*e.g.*, NESDIS Multi-SST analysis in May, 2010). See Table 1 for more information.

2.2. Merging procedure in L4-SQUAM for analyses of SST differences

To analyze SST differences, L4 pairs have to be matched up in space. This may be achieved by: (a) averaging or interpolating all the L4 SST data into a common grid (GMPE approach), (b) interpolating the first term ($L4_1$ in $\Delta T_S = L4_1 - L4_2$) to the resolution of the second term ($L4_2$), using various linear or cubic formulations or inverse distance-weighted methods, or, (c) selecting the nearest neighbor (NN). A detailed offline study was performed for an extreme combination of ultra-high resolution G1SST (0.01º) and low resolution RTG (0.5º) employing both bilinear interpolation and NN approach. Results are shown in Fig. 1. They unambiguously suggest that the effect of interpolation scheme on the global comparison statistics is negligible. (Note that this global result may not be valid when working in highly dynamic regions.) The simpler NN approach was thus adopted in L4-SQUAM.

INSERT FIG. 1 ABOUT HERE

Note that in L4-SQUAM, analyses are performed two ways. As an example, for OSTIA and CMC combination, differences are calculated both as "OSTIA−CMC" and "CMC−OSTIA". The second term is always selected as anchor (*i.e.*, CMC in the first case and OSTIA in the second). As a result of differences in the spatial interpolation, the comparison statistics may slightly differ, but this difference is always small as expected from Fig. 1.

3. Comparisons of global L4 SST fields in L4-SQUAM

This section describes the four types of diagnostics currently implemented in L4-SQUAM.

3.1. Maps and Histograms of ΔT_S

Fig. 2a shows an example map of ΔT_S between two foundation SSTs, GAMSSA and OSTIA.

INSERT FIG. 2 ABOUT HERE

Over most of the global ocean, ΔT_S is close to zero. However, the differences are prominent in the southern oceans, where GAMSSA is >1ºC warmer *w.r.t.* OSTIA over some regions, and in the Arctic, where the magnitude of differences may exceed 2ºC. Biases of both signs are also observed in many coastal locations. Note that different combinations of L4s show different patterns and magnitudes of biases. For instance, for the same date, 13 July 2011, DOI_AV shows highly variable biases *w.r.t.* OSTIA reaching more than ±1ºC (not shown) in many areas of the global ocean, in particular where GAMSSA and OSTIA appear to be consistent.

Fig. 2b shows a histogram of the differences corresponding to Fig. 2a. The ΔT_S statistics are annotated, including number of "match-ups" (due to anchoring to the second term in the NN interpolation, it approximately represents the number of valid grid points in OSTIA SST), minimum, maximum, mean, standard deviation (Std Dev), median, robust standard deviation (RSD), skewness and kurtosis. A dotted gray line shows an ideal Gaussian fit, $X \sim N(Median, RSD)$. Additionally, numbers of "match-ups" beyond "$Median \pm 4 \times RSD$" are shown on the top right. Note that time series of these "outliers" are plotted in L4-SQUAM but not excluded from comparison statistics. Overall, the distribution of ΔT_S is close to Gaussian, with mean and median close to zero, and Std Dev ~0.69ºC and RSD ~0.36ºC.

The difference between the conventional and robust statistics is noticeably high, indicating the large effect of outliers. A significant negative skewness is consistent with a large fraction of negative GAMSSA–OSTIA outliers mostly found in the Arctic (Fig. 2a), suggesting differences in treatment of ice in the two products. Both L4 products contain ice bit flags but different ice products. The bottom panels in Fig. 2 re-plot the top panels, but with ice-covered grid cells excluded when ice is reported in any one mask or both. The statistics change significantly. First, the number of match-ups is reduced by ~20%, from ~16.8 million in “all-grid” to ~13.4 million in “ice-free” ensemble. In the removed 3.4 million ice grid points, the temperature was likely set to default “melting ice” ~–2ºC in at least one of the products. There are grid points in which the ice cells have same values for both products, resulting in an artificial spike at zero in Fig. 2a. On the other hand, there are also grid cells in where one product reports ice and the other does not, resulting in a cold tail in the histogram and a somewhat distorted bell curve (an artificial small mode). As a result, the mean (ΔT_S) changes from −0.07°C in “all-grid” to +0.05°C in “ice-free” sample, and the Std Dev is reduced from ~0.69ºC to ~0.59ºC. However, the apparent worsening of skewness (compare Fig. 2b with Fig. 2a) is related to its decrease in Fig. 2a by the artificial small mode. (Interestingly, excluding icy pixels can also increase the Std Dev for those combinations of L4s where the ice masks are highly consistent, *e.g.*, for “DOI_AV *minus* DOI_AA)” (not shown), due to excluding many grid points with zero ΔT_S.)

The shape of the “ice-free” histogram is more regular and symmetric, and shows improved consistency between the robust and conventional statistics, indicating reduced effect of “outliers”, consistent with their reduced fraction. Note that “ice-free” analyses emphasize product comparison in the physical SST domain, whereas the “all-grids” analyses should assist L4 producers to diagnose and reconcile different ice masks. Hence both analyses are kept in L4-SQUAM and are available to its users by a click of a button.

3.2. Time series of “L4 minus L4” consistency and in situ validation

The statistical parameters annotated on the ΔT_S histograms are plotted as a function of time for various combinations of L4s to monitor products for relative stability and consistency.

Fig. 3a-b show examples of global “ice-free” mean differences and standard deviations in L4 fields *w.r.t.* DOI_AV, Fig. 3c-d show the same statistics *w.r.t.* drifters and Fig. 3e-f show the same *w.r.t.* GMPE. (*Note that statistics w.r.t. any L4 are available in L4-SQUAM webpage and the ones referred here are for illustration only*.)

INSERT FIG. 3 ABOUT HERE

The time series in Fig. 3 are very busy due to a large number of L4 products. A special provision was made in the L4-SQUAM webpage to allow users to perform interactive analyses by plotting and focusing on time series for one or several products of a user's choice. Although some observations discussed in this section may not be easily seen in Fig. 3, they are easily verified using the interactive capability in L4-SQUAM.

The two daily NOAA OISST products are largely consistent, with DOI_AA being ~+0.05ºC warmer than DOI_AV. The majority of the products are within ±0.15ºC of each other, with a few noticeable exceptions. For instance, G1SST is observed to be largely colder (within +0.05 to −0.2ºC) *w.r.t.* DOI_AV. Similar relatively "cold observations" are also seen in the NESDIS Multi-SST analysis and RTG products since about the beginning of 2010, although to varying magnitudes and with occasional spikes. Compared to DOI_AV, RTG_LR was a little warmer until 6 January 2005, after which time the two products became consistent until the end of 2007, and then RTG_LR became slightly colder than DOI_AV. The CMC was from 0.0 to 0.2ºC warmer than DOI_AV until about the end of 2004, after which time the two products have shown a negligible mean bias. Also, a pre-2006 trend flattening out subsequently is observed, which coincides with the change in input from Pathfinder to NAVOCEANO SST on 1 January 2006 (Reynolds *et al.*, 2007).

An interesting observation is the clustering of some products into groups. For example, the RTG_HR and NESDIS Multi-SST analysis products closely follow each other, forming a tight cluster. (Note that the NESDIS Multi-SST analysis uses a "thinned" RTG_HR for bias correction.) Similar observations are also seen for the foundation SSTs, with GAMSSA being sometimes slightly warmer than the rest of the foundation family, *e.g.*, from 13 April to 13 May, 2010 (Fig. 3a). Shortly after its start in early 2006, OSTIA had a cold mean bias of ~–0.2 ºC *w.r.t.* DOI_AV, which reduced to –0.1 ºC later in 2006 but then briefly spiked again in February 2007, May 2008 and May 2009. [*OSTIA reanalysis has not been processed in L4-SQUAM excluding ice yet and consequently is not shown in Fig. 3a and Fig. 3b; work is underway to add it and have in the revision*].

The standard deviations *w.r.t.* DOI_AV show a clear seasonal cycle, for all L4 products, but with different amplitudes. For instance, the two RTG, G1SST, and ODYSSEA products show Std Dev between ~0.5 and 0.95ºC. For OSTIA, K10 and GAMSSA, *w.r.t.* DOI_AV, Std Devs range between 0.45 and 0.65ºC, and the NESDIS Multi-SST analysis shows slightly higher values. The two NOAA OISST products are very consistent. A clear discontinuity in Std Dev is also observed for "CMC *minus* DOI_AV" and "RTG_LR *minus* DOI_AV" around the beginning of 2007 (the reason is unclear and will be explored in future which may be related to mutually inconsistent versions of products used in NRT L4-SQUAM).

L4-SQUAM *in situ* validation is stratified into drifters, ships, and tropical and coastal moorings, following the four major *in situ* data types available in *i*Quam.

Fig. 3c-d show global mean bias and standard deviation in L4 products *w.r.t.* drifters. Many of the observations in Fig. 3a-b are also reproduced in Fig. 3c-d, but with a reduced magnitude. For example, "RTG_HR *minus* DOI_AV" Std Dev ranges between 0.5 to 0.95ºC with strong seasonality, whereas for "RTG_HR *minus* Drifters" it ranges between 0.35 to 0.55ºC. It is also observed that "L4 *minus* GMPE" and "L4 *minus* Drifters" show remarkable consistency although of slightly different magnitudes. For example, Std Dev of "RTG_HR *minus* GMPE" ranges between 0.35 to 0.5ºC and shows patterns similar to "RTG_HR *minus* Drifters" ("RTG_HR *minus* GMPE" is available only for all-grids as both L4s do not provide ice-bits). These results suggest that GMPE may be used as an alternative (proxy) reference for validation, when *in situ* data are either unavailable or are sparse. (It should be noted that drifter SSTs are input into most of the L4 analyses in this study, see Table 1). This result is consistent with Part 1 (Martin *et al.*, 2011) which has shown that GMPE has lower errors than other SST analyses when compared with Argo floats. However, reprocessing GMPE back in time is needed, to extend the time coverage.

Comparisons against ship data and moorings also show interesting observations. The corresponding plots are not shown here in the interest of space, but the major observations are discussed below (interested readers are referred to L4-SQUAM webpage). Compared to ship data, all the L4s show negative differences, *i.e.*, ship records are warmer (due to engine intake) and also show much stronger seasonality (*cf.*, Xu and Ignatov, 2010). The standard deviations *w.r.t.* ship data are also much higher ranging from 0.75 to 1.3ºC. One interesting observation in the seasonality of "L4 *minus* Ships" mean differences is that many products show seasonal (sinusoidal) patterns of comparable amplitudes but different *signs*. For example, the trends for CMC and RTG seem to be anti-correlated to the trends shown by DOI_AV and DOI_AA.

Validation statistics against coastal moorings also vary significantly between different products. For example, Std Dev approximately ranges from 0.35 to 0.8ºC for DOI_AV and CMC, 0.4 to 1.0ºC for the NESDIS Multi-SST, 0.38 to 1.5ºC for G1SST, and 0.6 to 1.4ºC for RTG, K10, GAMSSA, ODYSSEA and GMPE. Another interesting observation is a jump in the Std Dev for OSTIA on 28 July 2009. Prior to this date, the Std Dev ranged between 0.2 and 0.6ºC, but after that it ranges from 0.4 to 1.6ºC. The reasons for these differences are not fully apparent at this stage and should form the matter of future investigations.

Although some pairs of products show a close to zero global mean difference, the large standard deviation suggests significant regional differences which are further analyzed next using Hovmöller diagrams.

3.3. Hovmöller diagrams

Hovmöller diagrams provide a way to visualize and understand zonal time series evolution of ΔT_S and detect seasonal cycles and climatic trends. Fig. 4 (top-left and bottom-left panels) show examples of Hovmöller diagrams of ice-free mean biases and standard deviations for "RTG_HR *minus* DOI_AV".

INSERT FIG. 4 ABOUT HERE

On average, the RTG_HR and the DOI_AV SSTs agree well everywhere except in the high latitudes around ~60°S and ~70°N, where large persistent biases and seasonal cycles are observed. The Std Dev are small in the sub-tropical oceans, increasing in the Inter-Tropical Convergence Zone (ITCZ) and the mid-latitudes, and reaching 0.75-1ºC from 40°N to 75°N. The cause of these differences is not fully clear. Recall that DOI_OI uses the NAVOCEANO L2 SST as input, whereas RTG high resolution SST employs a unique physical SST retrieval as a part of their L4 production. Similar patterns are observed in RTG_HR and NESDIS Multi-SST analysis products, compared to any other L4 products.

To better understand the causes of differences between products, mean biases and Std Dev of "RTG_HR *minus* Drifters" and "DOI_OI *minus* Drifters" are also plotted in Fig. 4. Both L4 products show a near zero mean bias in the full domain covered by drifters. The large "RTG_HR *minus* DOI_AV" mean biases and Std Dev are not captured in the *in situ* validation, suggesting that the large L4 differences are observed in the areas not covered by *in* *situ* data. On the other hand, when *in situ* data are present, both L4s agree with them well, suggesting that they are assimilated in L4s with a relatively high weight and therefore using the very same *in situ* data as were ingested to validate an L4 is not fully representative of its true global performance.

Another interesting observation includes warmer biases in GAMSSA over the Southern Ocean and colder biases over the Arctic Ocean compared to most other products (see L4-SQUAM webpage for figures). In fact, over the Arctic Ocean, most of the products show distinctive seasonal biases *w.r.t.* each other (not only GAMSSA). Besides the differences in sea-ice treatment discussed in Section 3.1, these differences may also be attributed to different bias correction schemes and zonal inconsistencies between input L2P products. For example, the GAMSSA system removes "global" biases in the input L2P SSTs using the associated per-pixel bias estimates obtained from global buoy match-ups. In contrast, the Met Office uses "regional" AATSR and buoy SSTs to de-bias the L2P inputs (Stark *et al.*, 2007). The Reynolds OISST and CMC systems de-bias all satellite inputs "regionally" using both buoy and ship SSTs (Reynolds *et al.*, 2007; Brasnett, 2008). Besides differences in bias removals, the L2P inputs also show significant mutual zonal differences. For example, Reynolds *et al.* (2010; see Fig. 5 therein) showed that AVHRR, AATSR and AMSR-E SSTs diverge at high latitudes as well as over the equator when referenced to DOI_AA (AVHRR+AMSR-E) SST. Noticeably, NOAA-18 and -17 SSTs are warmer over the Southern Ocean. Similar patterns are also seen from comparisons between AVHRR GAC and DOI_AV SSTs (*cf.*, *http://www.star.nesdis.noaa.gov/sod/sst/squam/NAVO/navo_sst_diff_hovmoller.htm*). It may therefore be inferred that much of the warm bias between GAMSSA and other L4 products over the Southern Ocean and mutual inconsistencies between most L4 SSTs over the Arctic Ocean can be mitigated by using L2 SSTs which use regional (zonal) calibrations and provide per-pixel bias estimates based on regional *in situ* observations (rather than global). This would also reduce the need for analysis systems to perform their own bias-correction of satellite data and allow for greater consistency between L4 products.

4. Possible extension of L4-SQUAM analyses

This section explores potential extensions to the L4-SQUAM functionalities.

4.1. Diurnal-cycle resolved L4 products

All L4 products currently monitored in SQUAM are daily, and do not resolve the diurnal cycle. Some L4 developers have started exploring diurnal cycle resolved L4 products (*e.g.*, BoM and NCODA produce experimental products with 3 hour and 6 hour resolution, respectively). Modeling of diurnal variation (DV) may have various degrees of complexity and accuracy, depending on methods of accounting for solar insolation and its propagation in the top few meters of the ocean water (*cf.*, Stuart-Menteth *et al.*, 2005; Gentemann *et al.*, 2007; Donlon *et al.*, 2007; Kennedy *et al.*, 2007). One could expect that the recent trend towards finer time resolution L4 products will continue, and L4-SQUAM will need to be adjusted accordingly to report and monitor such L4 products.

Analyses by Dash *et al.* (2010) suggest that one could validate such DV models, by combining satellite L2 products with diurnally resolved L4s. Towards that objective, a double-differencing (DD) technique was implemented in L2-SQUAM. In particular, Day–Night (DN) DDs are calculated as follows $DN = (T_{SD}–T_R) – (T_{SN}–T_R) \approx T_{SD}–T_{SN}$, where T_{SD} and T_{SN} are daytime and nighttime satellite L2 SSTs, and T_R is the L4 “reference” SST which is used here as a “transfer standard”. Note that DN differences can also be calculated by direct differencing of the respective L2 products, but this can only be done in a sub-sample of the global data domain, where both day and night retrievals are available at the same location. However, the DD technique allows substantial extension of the comparison domain and makes the comparison more stable and consistent from day to day. More discussion is found in Dash *et al.* (2010).

Fig. 5 shows an example DN time series for four AVHRR sensors, generated by the NESDIS heritage SST system, using DOI_AV as the transfer standard.

INSERT FIG. 5 ABOUT HERE

The DN values are always positive, since the AVHRR L2 SST product is subject to diurnal changes and DOI_AV SST is one daily value that does not account for the diurnal cycle. As expected, the afternoon platforms, NOAA-18 and -19, which pass at ~1:30 am/pm, show higher DN values than the morning platforms, NOAA-17 and Metop-A, which overpass at ~10 am/pm. (Note that a systematic residual offset between NOAA-18 and -19 of ~0.10 ºC is likely due to the empirical setting of regression coefficients in NESDIS L2 production and not from the DV physics. Work is underway to understand and remove this bias.) Using a diurnally resolved L4 as a transfer standard in the DD technique should compensate for the diurnal differences observed in the L2 product, and make the DN time series flat and close to zero. Thus calculation of DN differences using DD technique, with various diurnally-resolved L4 products employing different DV models, provides an assessment of global performances of the diurnally-resolved L4 products.

Likewise, any external DV model can also be validated using this technique by applying it to remove the diurnal variation from L2 SSTs, or by adding it on the top of the "daily" L4 field and then recalculating the DN DDs. These analyses are the subject of future work and will contribute to the GHRSST DV Working Group activities (*https://www.ghrsst.org/ghrsst-science/science-team-groups/dv-wg/*).

4.2. Dependencies

The SST differences may also be plotted as a function of retrieval conditions, *e.g.*, latitude, proximity to the coast and bathymetry. Such "dependencies" plots are helpful to stratify the differences and focus on domains with the largest differences. Examples of wind speed dependencies are shown in Fig. 6 for "MUR *minus* GMPE" and "CMC *minus* GMPE".

INSERT FIG. 6 ABOUT HERE

Both MUR and CMC are foundation SST products. According to L4-SQUAM, the GMPE provides a good average representation of the foundation family. It is thus expected that these products should be consistent in the full range of wind speeds. Indeed, there is a high degree of consistency between MUR, CMC and GMPE. However, the corresponding ΔT_S vary across the wind speed range, with product-specific amplitudes. For example, at low winds MUR is colder than GMPE by 0.1ºC, whereas at high winds it is warmer by 0.1ºC. Under low wind conditions, this may be attributed to a cool-skin effect, MUR being a satellite-only product (no *in situ*; see Table 1), which reduces with increasing wind speed. The corresponding standard deviations are largest at low winds (~0.5ºC) and decrease towards larger winds reaching ~0.35-0.40ºC. The CMC product shows similar trends but with lesser magnitudes. Including such dependencies in SQUAM and verifying over longer time series, will help to better understand the cause of these residual biases and fix them in L4 data.

4.3. Correlograms and N-way error analyses

Another potential extension of L4-SQUAM is adding autocorrelation analyses (*cf.*, Box and Jenkins, 1976). The autocorrelation of the time series is defined as a lagged correlation between the same variable measured at two different times (days), x_t and x_{t+lag}, and is used to detect non-randomness in the time series. The autocorrelation coefficient "*r*" for lag "*k*" is calculated as $r = \Sigma(x_t - \bar{x})(x_{t+k} - \bar{x})/\Sigma(x_t - \bar{x})^2$. The "*r*" *vs*. "*k*" for time series biases and standard deviations in "L4 *minus* drifters" are shown in Fig. 7a and Fig. 7b, respectively.

INSERT FIG. 7 ABOUT HERE

In general, if day-to-day variations in "L4 *minus* drifters" mean biases and standard deviations are random then the error in the L4 field has no "memory" and "*r*" would be close to zero. Deviation of "*r*" from zero can be used as a measure of this memory. Both Fig. 7a-b show that autocorrelations are positive and very strong for the first several days and then decay exponentially. However, the magnitudes of "*r*" can be significantly different for different L4 SSTs, and also between mean bias and standard deviations for a given product. For example, in Fig. 7a, OSTIA shows the lowest and RTG_HR shows the highest "randomness", whereas in Fig. 7b, DOI_AV and OSTIA_RAN show lowest and GAMSSA and ODYSSEA show highest "randomness". Fig. 7a suggests that the bias in some fields, *e.g.*, OSTIA, are rather smooth and consistent *w.r.t.* drifters whereas for some fields (*e.g.*, RTG_HR) they are noisier. It should however be noted that interpretation of such "preliminary" conceptual plots must be performed in conjunction with validation time series and spatial autocorrelation maps because the results are subjective (future work). For example, an L4 with high "*r*" but consistent low bias and Std Dev might be a positive thing whereas the reverse may indicate otherwise.

Another potential addition would be to estimate individual contributions of a given product to the observed differences. This may be achieved by employing an N-way error analysis, *cf.*, three-way error analysis by O'Carroll *et al*. (2008), which was applied to three global but limited in time data sources (ATSR, AMSR-E and drifters) and individual errors for these three products were derived (three average numbers representing root mean squared errors in these products). Similarly, it can be employed to derive product-specific spatial error fields rather than a global average number (*cf.*, Xu and Ignatov, 2010, who explored derivation of error fields using Pathfinder SST, DOI_AV and *in situ* data). In SQUAM, where many L2 and L4 products and *in situ* data are available, the three way analyses may also be extended into N-way error analyses, provided the assumption of mutually independent error structures and negligible systematic biases in the spatially and temporarily collocated products hold. Also, time-averaged L4 SST differences, *e.g.*, monthly mean difference maps, may be useful for identifying persistent and seasonal features, as has been suggested by some L4 producers.

To close the discussions in Section 4, although the current L4-SQUAM metrics address its objectives, it could potentially be further expanded. This has been briefly explored here and will form the subject of future investigations.

5. Summary and future work

The web-based L4 SST quality monitor (L4-SQUAM) was developed at NOAA NESDIS to monitor global L4 SST fields, in near real-time and retrospective modes. The L4-SQUAM is complementary to the two other existing systems of the IC-TAG: the GHRSST Multi Product Ensemble (GMPE; Martin *et al.*, 2011; Part 1 companion paper) and the High Resolution Diagnostic Data Set (HR-DDS; Poulter, 2011; Part 3 companion paper).

As of this writing, the thirteen daily L4 SSTs are monitored in L4-SQUAM, with two additions underway and another four planned.

L4-SQUAM metrics are based on analyses of “L4 *minus* L4” and “L4 *minus in situ*” ΔT_S. The maps and the Hovmöller plots provide synoptic snapshots of differences and similarities between various products, the histograms check for their proximity to a Gaussian shape, and the time series assess relative stability of consistency statistics. To better understand and reconcile the ice masks in individual products, analyses are performed in two ways: including and excluding ice, when the corresponding bit information to extract ice-mask is available. All processing is done automatically, within 24 hours of data availability, and the diagnostics are posted at [*http://www.star.nesdis.noaa.gov/sod/sst/squam/L4*](http://www.star.nesdis.noaa.gov/sod/sst/squam/L4).

The foundation SSTs seem to show more consistency with each other whereas some of the depth-SSTs show persistent zonal differences. The differences are often more pronounced in high latitudes, associated with ice and sparse data coverage in both satellite and *in situ data*, and in coastal areas. However, large differences also exist in the open oceans. Further efforts should be directed towards understanding and reconciling different L4 SSTs.

Analyses in SQUAM emphasize the need for diurnally-resolved L4 SSTs, and their global validation using L2 SSTs. Dependence of SST differences on geophysical parameters, autocorrelation and N-way error analyses are potential useful additions to L4-SQUAM.

Having all the various L4 SSTs in one place, uniformly analyzed and compared to the same *in situ* data allows L4 SQUAM to provide L4 data users and producers with valuable information on which L4 products are available, their relative merit for particular product applications and their potential areas of improvement. While the objective of L4-SQUAM is to provide the users and producers with representative diagnostics, highlighting differences or similarities between various products and their associated strengths and weaknesses, it is beyond the scope of this work to conclusively determine the “best” SST. Furthermore, it is not the purpose of L4-SQUAM to determine which data set is the “best” or select “one” product suitable for all applications. It is up to the users to choose which product is better suited to their applications based on diagnostics from GMPE, L4-SQUAM and HR-DDS. To assist in this goal, maps, histograms and time series plots are made available at *http://www.star.nesdis.noaa.gov/sod/sst/squam/L4/* for all combinations of "L4 *minus* L4" for all available dates. One may also ask whether it is justified to compare foundation and depth SSTs, which are expected to be inherently inconsistent, and whether this may be a reason for differences in some regions, *e.g.*, in sub-tropical latitudes with light winds and high insolation. This is also true for other combinations, *i.e.*, "foundation *vs.* foundation" and "depth *vs.* depth" SSTs. In-depth analyses to diagnose the causes of these differences, however, are beyond the scope of L4-SQUAM which only highlights the differences or similarities as they are. Nevertheless, the presence of multiple combinations of L4 SSTs in L4-SQUAM can provide confidence in the diagnostics, *e.g.*, if one product deviates from the majority of the products for any given region, it is more likely (although not conclusively) that the issue or problem is in the deviant product.

Finally, L4-SQUAM was initially developed as a near real-time system aimed at short-term diagnostics. Nevertheless, it is "climate-ready" with demonstrated capabilities to analyze and compare datasets for longer periods and will be instrumental for monitoring and reconciling long-term SST records.

Acknowledgments

This work was supported by the Joint Center for Satellite Data Assimilation (JCSDA) Science and Development Implementation (JSDI) program. We thank Feng Xu (formerly at NESDIS) for setting up the *i*Quam system and many NESDIS SST colleagues for helpful discussions. E. Armstrong, M. Chin and J. Vazquez acknowledge the support from NASA's Making Earth Science Data Records for Use in Research Environments (MEaSUREs) program. 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

Autret, E., Piollé, J.-F., 2011: Product User Manual for ODYSSEA Level 3 and 4 global and regional products, MYO-PUM-SST-TAC-ODYSSEA, Ifremer/CERSAT. [Available online at: http://projets.ifremer.fr/cersat/Data/Discovery/By-parameter/Sea-surface-temperature/ODYSSEA-Global-SST-Analysis].

Beggs H., Zhong, A., Warren, G., Alves, O., Brassington, G., Pugh, T., 2011: RAMSSA – An Operational, High-Resolution, Multi-Sensor Sea Surface Temperature Analysis over the Australian Region. *Australian Meteorol. & Oceanographic J.*, **61**, 1-22.

Box, G. E. P., Jenkins, G. M., 1976: *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day, *pp*. 28-32.

Brasnett, B., 1997: A global analysis of sea surface temperature for numerical weather prediction. *J. Atmos. Oceanic Technol.*, **14**, 925-937.

Brasnett, B., 2008: The impact of satellite retrievals in a global sea-surface-temperature analysis. *Q. J. R. Meteorol. Soc.*, **134**, 1745-1760. DOI: 10:1002/qj.319.

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

Chao, Y., Li, Z., Farrara, J. D., Hung, P., 2009: Blending Sea Surface Temperatures from Multiple Satellites and In Situ Observations for Coastal Oceans. *J. Atmos. Oceanic Technol.*, **26**, 1415–1426. doi: 10.1175/2009JTECHO592.1

Cummings, J. A., 2005: Operational multivariate statistics and model data comparisons. *Q. J. R. Meteorol. Soc.*, **131**, 3583-3604.

Dash, P., Ignatov, A., Kihai, Y., Sapper, J., 2010: The SST Quality Monitor (SQUAM). *J. Atmos. Oceanic Technol.*, **27**, 1899–1917. doi: 10.1175/2010JTECHO756.1

Donlon, C.J., Martin, M., Stark, J.D., Roberts-Jones, J., Fiedler, E., 2011: The Operational

Sea Surface Temperature and Sea Ice Analysis (OSTIA), *Rem. Sens. Env.* (*in press*).

Donlon, C., Robinson, I., Casey, K. S., Vázquez-Cuervo, J., *et al.*, 2007: The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bull. Amer. Meteor. Soc.*, **88**, 1197–1213.

Donlon, C. J., Keogh, S. J., Baldwin, D. J., *et al.*, 1998: Solid-state radiometer measurements of sea surface skin temperature. *J. Atmos. Oceanic Technol.*, **15**, 775-787.

Gemmill, W., Katz, B., Li, X., 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., Donlon, C. J., Stuart-Menteth, A., Wentz, F. J., 2007: Diurnal signals in satellite sea surface temperature measurements. *Geophys. Res. Lett.*, **30**, doi:10.1029/2002GL016291.

Kennedy J. J., Brohan, P., Tett, S. F. B., 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., Podesta, G. P., Evans, R., 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.

Kurihara, Y., Sakurai, T., Kuragano, T., 2006: Global daily sea surface temperature analysis using data from satellite microwave radiometer, satellite infrared radiometer and in-situ obvservations, *Weather Bulletin*, 73. 1-18 (in Japanese).

Le Borgne, P., Legendre, G., Marsouin, A., 2007: Operational SST Retrieval from MetOp/AVHRR, *Proc. 2007 EUMETSAT Conf., Amsterdam*.

Martin, M., Dash, P., Ignatov, A., Autret, E., Banzon, V., Beggs, H., Brasnett, B., Cayula, J.-

F., Chao, Y., Cummings, J., Donlon, C., Gentemann, C., Grumbine, R., Ishizaki, S., Maturi, E., McKenzie, B., Reynolds, R., Roberts-Jones, J., 2011: Group for High Resolution SST (GHRSST) Analysis Fields Inter-Comparisons: Part 1. A GHRSST Multi-Product Ensemble (GMPE). *Deep Sea Res. II* (*submitted; companion paper*).

Maturi, E., Harris, A., Merchant, C., *et al.*, 2008: NOAA's sea surface temperature products from operational Geostationary satellites. *Bull. Amer. Meteor. Soc.*, **89**, 1877–1888.

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

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

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

O'Carroll, A. G., Eyre, J. R., Saunders, R. W., 2008: Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *J. Atmos. Oceanic Technol.*, **25**, 1197-1207.

Poulter, D., 2011. Group for High Resolution SST (GHRSST) Analysis Fields Inter-Comparisons: Part 3. High Resolution Diagnostic Data-set (HR-DDS) system. (*in preparation; companion paper*).

Reynolds, R.W., Gentemann, C.L., Corlett, G.K., 2010: Evaluation of AATSR and TMI Satellite SST data. *J. Climate*, **23**, 152-165.

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

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

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

Stuart-Menteth, A. C., Robinson, I. S., Donlon, C. J., 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.

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

Walton, C. C., Pichel, W. G., Sapper, J. F., 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.

Wu, X., Menzel, P. W., Wade, G. S., 1999: Estimation of sea surface temperatures using GOES 8/9 radiance measurements. *Bull. Amer. Meteor. Soc.*, **80**(*6*), 1127–1138.

Xu, F., Ignatov, A., 2011: Evaluation of in situ SSTs for use in the calibration and validation of satellite retrievals, *J. Geophys. Res.*, **115**, C09022, doi:10.1029/21010JC006129.

Zhong, A., Beggs, H., 2008: Operational Implementation of Global Australian Multi-Sensor Sea SurfaceTemperature Analysis. *Analysis and Prediction Operations Bulletin* No. 77, *Bureau of Meteorology*, Australia, 2 October 2008. [Available online at: *http://cawcr.gov.au/projects/SST/GAMSSA_BoM_Operational_Bulletin_77.pdf*]

Tables

Table 1: List of L4 SST products monitored or considered in L4-SQUAM.

Product	Space/Time Res. & Type	Abbreviation & mode	Reference	Availability period, data format, and ftp source,	Input data				Ice bit
					Infrared	Microwave	*Insitu*	Other	
Products fully implemented in L4-SQUAM									
Optimal Interpolation SST	0.25° Daily Depth (bulk)	DOI_AV NRT; delayed reanalysis	Reynolds *et al.*, 2007	1981 to present, netCDF *ftp://eclipse.ncdc.noaa.gov /pub/OI–daily-v2/NetCDF*	AVHRR (PF until 2005, then NAVO)	-NA-	√	NCEP ice	√
		DOI_AA NRT; delayed reanalysis		Jun-2002 to present, netCDF *ftp://eclipse.ncdc.noaa.gov /pub/OI–daily-v2/NetCDF*	AVHRR	AMSR-E	√	NCEP ice	√
Operational SST & Sea Ice Analysis	0.05° Daily Foundation	OSTIA NRT	Stark *et al.*, 2007;2008; Donlon *et al.*, 2011	Apr-2006 to present, netCDF *ftp://podaac-ftp.jpl.nasa.gov /allData/ghrsst/data/L4/GLOB/UKMO/OSTIA*	AVHRR, AATSR, SEVIRI	TMI, AMSR-E	√	O&SI SAF ice	√
		OSTIA_RAN Reanalysis		1985-2007, netCDF *ftp://data.ncof.co.uk/ostia_reanalysis/* (passwd)	AVHRR PF, (A)ATSR	-none-	√	O&SI SAF ice	√
Real Time Global SST	0.50° Daily Depth (bulk)	RTG_LR NRT	Thiébaux *et al.*, 2003	Dec-2000 to present, gridded binary (grib) *ftp://polar.ncep.noaa.gov /pub/history/sst*	AVHRR physical retrievals	-none-	√	NCEP ice	X
	1/12° Daily Depth (bulk)	RTG_HR NRT	Gemmill, Katz, & Li, 2007	Feb-2007 to present, grib *ftp://polar.ncep.noaa.gov /pub/history/sst/ophi* (rotated for a year)	AVHRR	-none-	√	NCEP ice	X
NAVOCEANO K10 Analysis	0.10° Daily Depth	K10 NRT	*http://podaac.jpl.nasa.gov/ dataset/NAVO-L4HR1m-GLOB-K10_SST*	Apr-2008 to present, netCDF *ftp://podaac-ftp.jpl.nasa.gov /allData/ghrsst/data/L4/GLOB/NAVO/K10_SST*	AVHRR, GOES	AMSR-E	X	JPL climate	X
NESDIS Multi-SST Analysis *(formerly called POES-GOES)*	0.10° Daily Depth	GOESPOES NRT	Maturi *et al.*, 2008; *http:// www.nesdis.noaa.gov/mecb /blended_validation/*	Feb-2009 to present, HDF *ftp://dds.nesdis.noaa.gov/pull/* (passwd)	AVHRR, GOES, MTSAT, SEVIRI,	*Planned: AATSR, AMSR-E, AMSR-2*	X	NCEP ice (since May 2010)	√
JPL ultra high resolution G1SST	0.01° Daily,±80°lat Foundation	G1SST NRT	Chao *et al.*, 2009	Jun-2010 to present, netCDF *ftp://podaac-ftp.jpl.nasa.gov/allData/ghrsst /data/L4/GLOB/JPL_OUROCEAN/G1SST/*	AVHRR, AATSR, MODIS, GOES, SEVIRI, MTSAT	TMI, AMSR-E	√	Some ice	√

Canadian Met. Centre Analysis	0.2° Daily Foundation	CMC 0.2° NRT	Brasnett, 1997; 2008	Jan-2002 to present, netCDF (contact CMC for data access)	AVHRR, AATSR	AMSR-E	√	CMC ice	X
Australian BoM GAMMSA	0.25° Daily Foundation	GAMSSA NRT	Beggs *et al.*, 2011; Zhong & Beggs, 2008	Oct-2008 to present, netCDF *ftp://podaac-ftp.jpl.nasa.gov/allData/ghrsst/data/L4/GLOB/ABOM/GAMSSA_28km*	AVHRR, AATSR	AMSR-E	√	NCEP ice	√
Ocean Data Analysis, MyOcean/GMES	0.10° Daily Foundation	ODYSSEA NRT	Autret & Piollé, 2011	Reinstated Sep-2010 to present, netCDF *ftp://eftp.ifremer.fr/cersat-rt/project/myocean/sst-tac/l4/glob/odyssea/* (passwd)	AVHRR, AATSR, GOES, SEVIRI	TMI, AMSR-E	X	O&SI SAF ice	√
GHRSST Multi Prod. Ensemble	0.25° Daily Ensemble	GMPE NRT	Martin *et al.*, 2011	Sep-2009 to present, netCDF *ftp://data.ncof.co.uk/* (passwd *via* MyOcean)	-NA-	-NA-	-NA-	O&SI SAF ice	X
				Products currently being tested					
JPL Multi-scale Ultra-high Res. SST	0.01° Daily Foundation	MUR Being tested	*http://mur.jpl.nasa.gov/multi_resolution_analysis.php*	Jan-2009 to present, netCDF *ftp://podaac-ftp.jpl.nasa.gov/allData/ghrsst/data/L4/GLOB/JPL/MUR/*	MODIS (Terra, Aqua), AHVRR (GAC)	AMSR-E	X	O&SI SAF ice	√
RSS MW OI	0.25° Daily Minimum	RSS MISST NRT	*http://www.remss.com/*	Jun-2002 to present, netCDF *ftp://ftp.discover-earth.org/sst/misst/l4/tmi_amsre/nc*	-NA-	TMI, AMSR-E	X	-	√
				Products potentially being considered to be included					
RSS IR+MW	0.25° Daily Foundation	RSS_IR, NRT IR+MW, NRT	*http://www.remss.com/*	netCDF *ftp://ftp.discover-earth.org/sst/*	MODIS	AMSR-E, TMI	X	-	√
JMA Merged SST	0.25° Daily Foundation	MGDSST NRT; delayed reanalysis	Kurihara *et al.*, 2006	1985 to present, Plain binary *http://goos.kishou.go.jp/rrtdb/usr/pub/JMA/mgdsst/* (passwd)	AVHRR (GAC, HRPT)	AMSR-E	√	JMA sea-ice	√
DMI OI SST analysis	0.05° Daily Foundation	DMISST NRT	*http://ocean.dmi.dk/*	Jan-2011 to present, netCDF *ftp://ftpserver.dmi.dk/GBL005/* (passwd)	AVHRR (GAC, HRPT), SEVIRI, AATSR	AMSR-E	X	X	X
Naval Res. Lab. NCODA analysis	~0.08° 6 hourly Depth	NCODA experimental	Cummings, 2005	May-2009 to present, netCDF *http://tds.hycom.org/thredds/GLBa0.08/expt_90.8.html*	-	-	√	-	X

Figure captions and figures

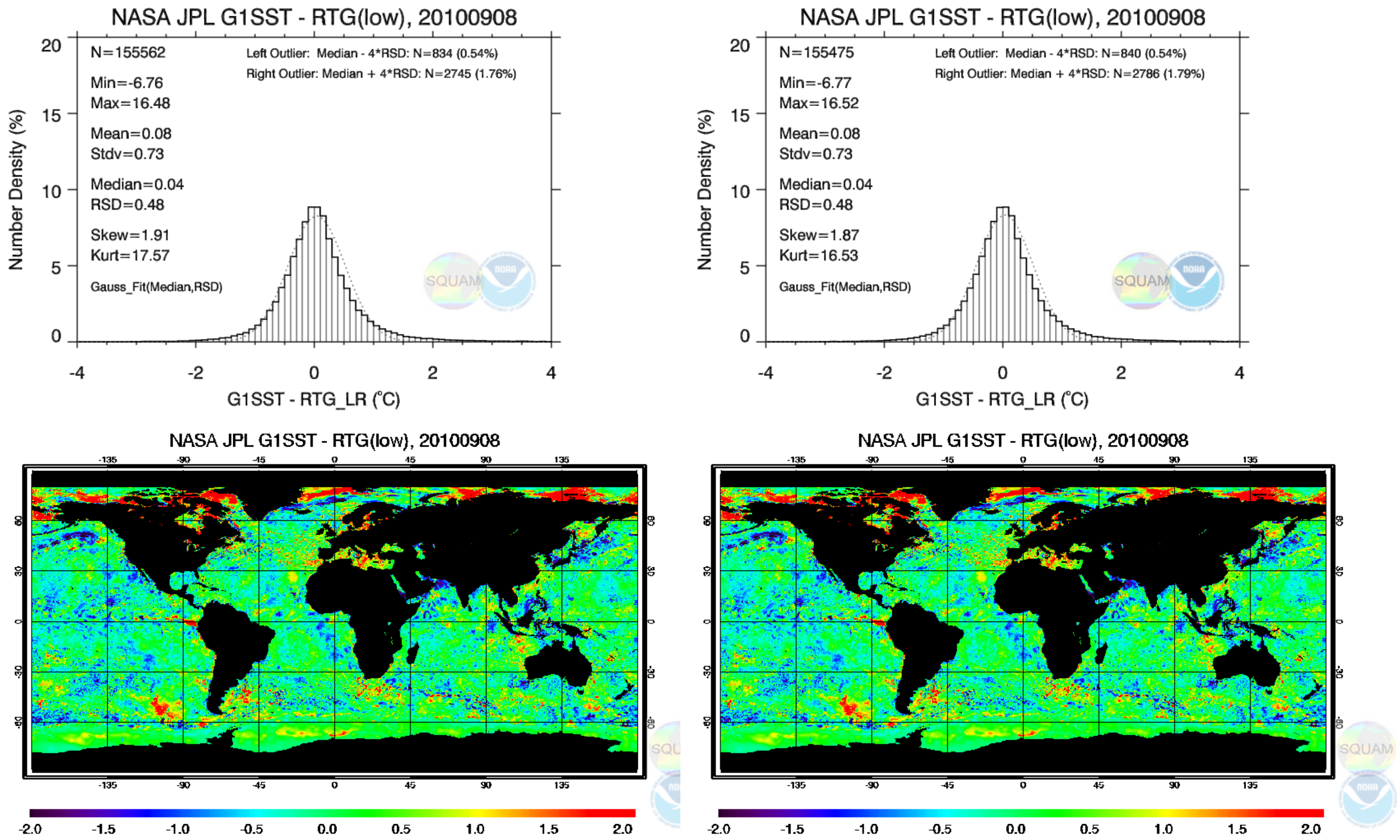
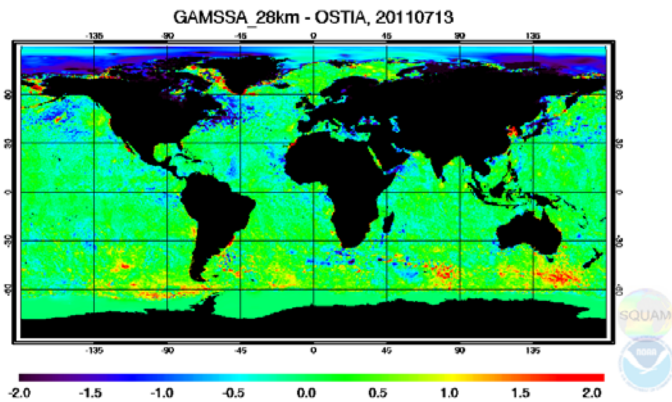
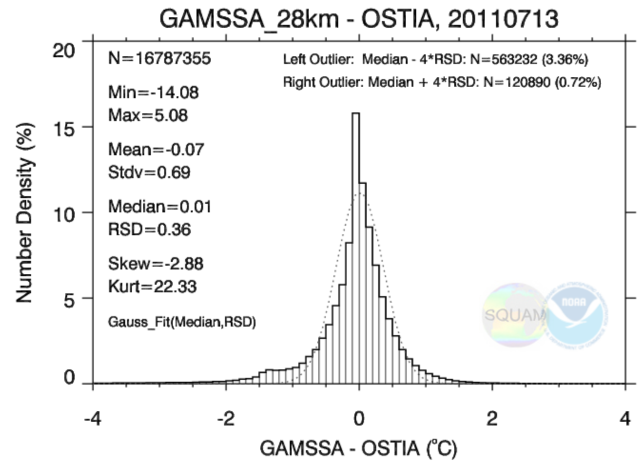


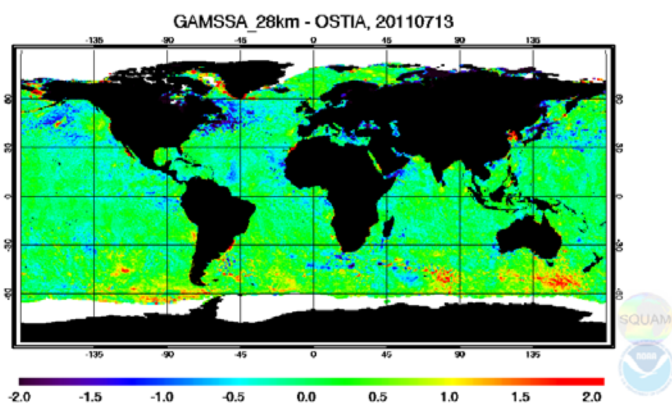
Fig. 1: Effect of interpolation on merging L4 SST fields (0.01° ultra high resolution G1SST *minus* 0.5° lat-lon RTG). Statistical moments are annotated on the histograms (see Section 3.1 for description). Left panels: nearest neighbor selection anchored to RTG; Right panels: bilinear interpolation of G1SST to RTG grid.



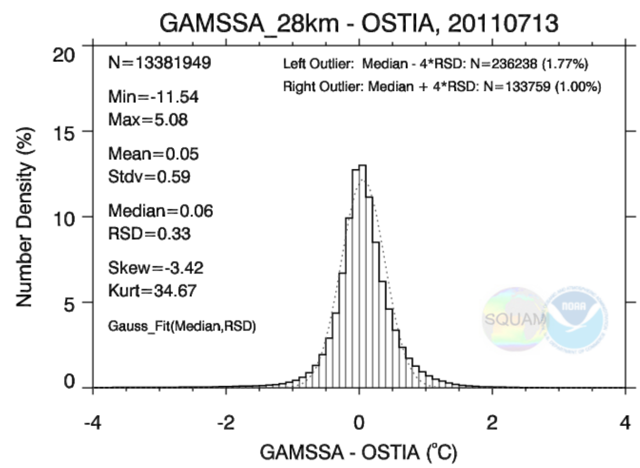
a) GAMSSA *minus* OSTIA, ice included



b) Frequency distribution corresponding to Fig. 2a

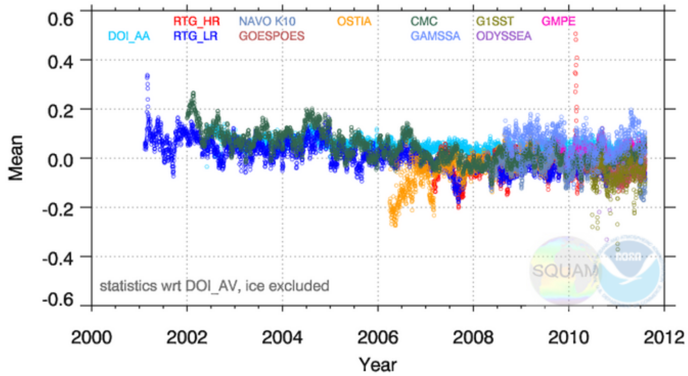


c) GAMSSA *minus* OSTIA, ice excluded

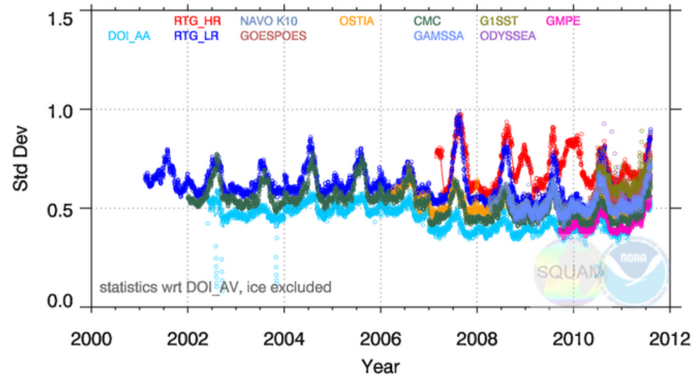


d) Frequency distribution corresponding to Fig. 2c

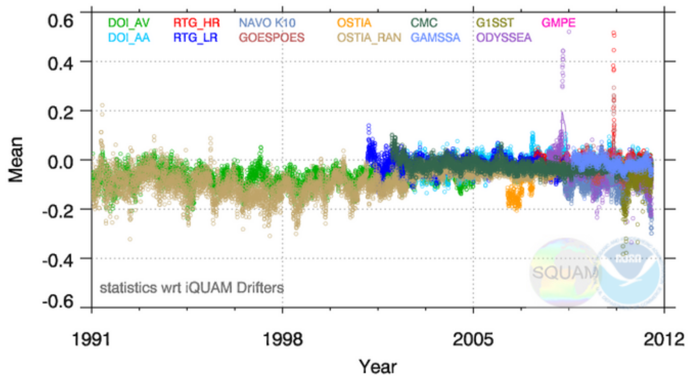
Fig. 2: In the left panels, spatial differences between two L4 SSTs (GAMSSA *minus* OSTIA) are observed, which are close to zero in many areas but are also prominent in some areas, *e.g.*, roaring forties and in many coastal locations. The arctic ice areas also show significant differences between the two products. In the right panels, ΔT_S statistics are annotated on the left side of the histograms, dotted gray line shows an ideal Gaussian fit, and the numbers of L4 match-ups beyond "Median ± 4×Robust Std Dev" are shown on the top right. Note that due to NN interpolation, anchored to the second term (*i.e.*, OSTIA), the match-up "N" is equal or close to the number of valid grid cells in OSTIA. Top-panels: ice included in the analyses; Bottom-panels: ice excluded.



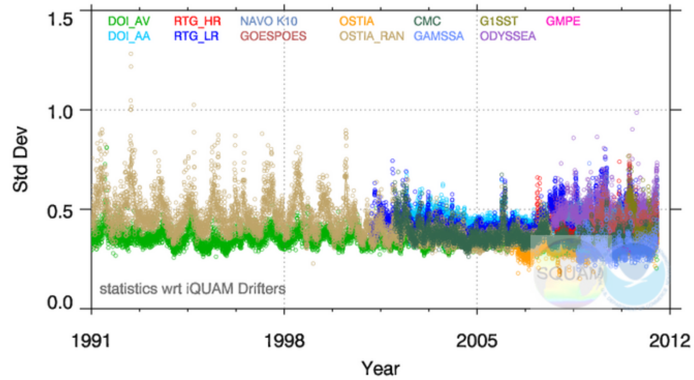
a) Mean, "L4 – Reynolds(AVHRR)", ice excluded



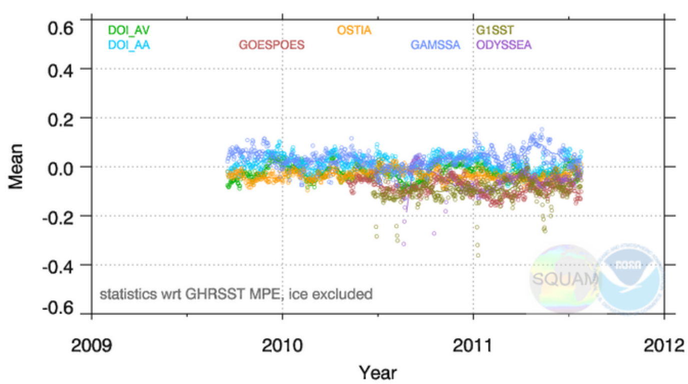
b) Std Dev, "L4 – Reynolds(AVHRR)", ice excluded



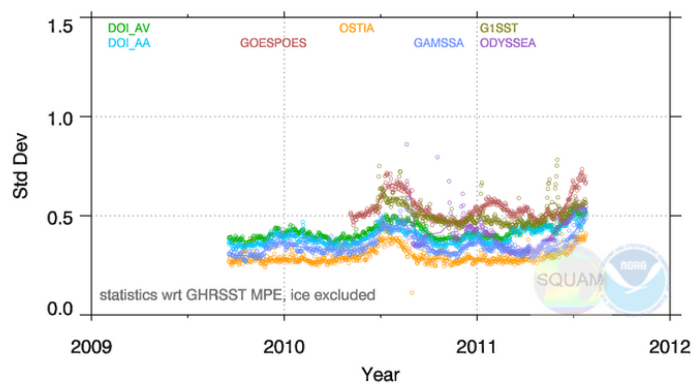
c) Mean, "L4 – Drifters"



d) Std Dev, "L4 – Drifters"



e) Mean, "L4 – GMPE", ice excluded



f) Std Dev, "L4 – GMPE", ice excluded

Fig. 3: Mean and standard deviation of ΔT_S. Left-panels: median; Right-panels: standard deviation. Top-panels: statistics *w.r.t.* Reynolds (AVHRR) excluding ice grids; Middle-panels: same as top-panels but *w.r.t.* drifters; Bottom-panels: same as top-panel but *w.r.t.* GMPE.

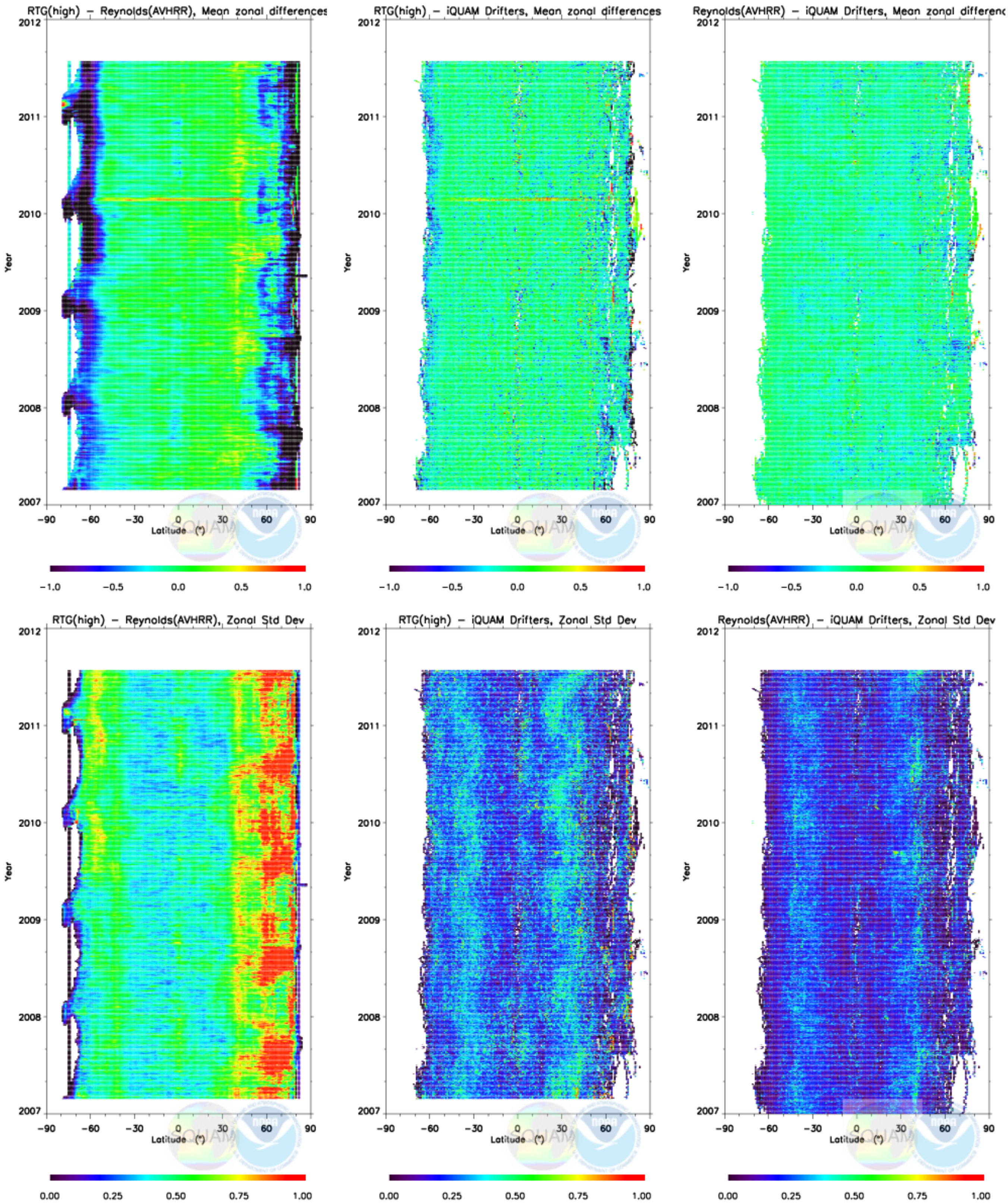


Fig. 4: Hovmöller diagrams of average zonal differences: *First column*: RTG (high)– Reynolds, ice excluded; *Second column*: RTG (high) – Drifters; *Third column*: Reynolds (AVHRR) – Drifters; Top-panels: mean differences; Bottom-panels: standard deviations.

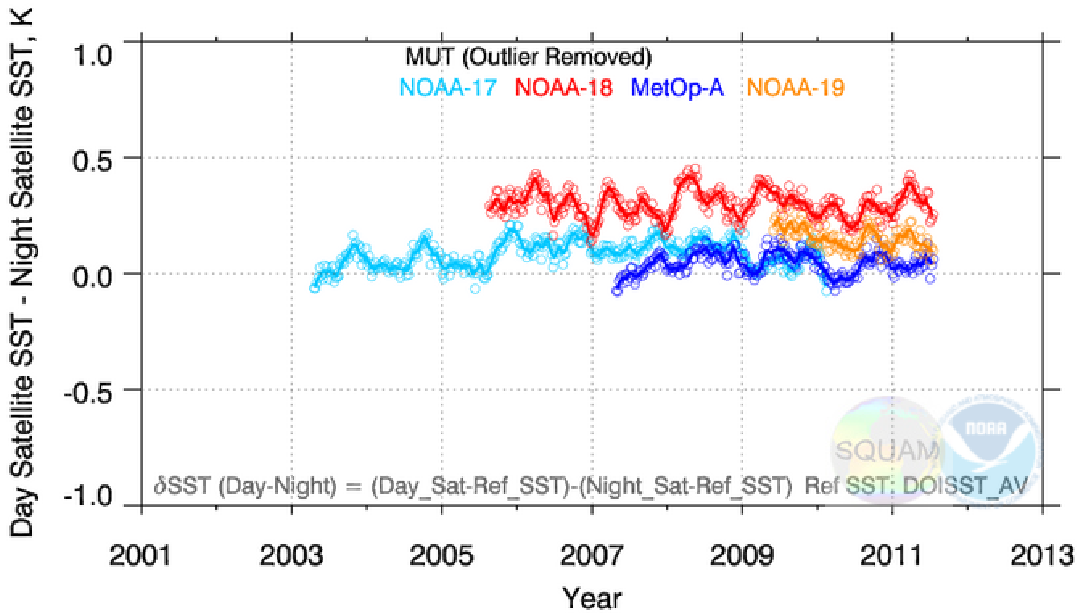


Fig. 5: Average "Day *minus* Night" SST differences estimated employing double differencing (DD) technique, with daily Reynolds SST as the transfer standard.

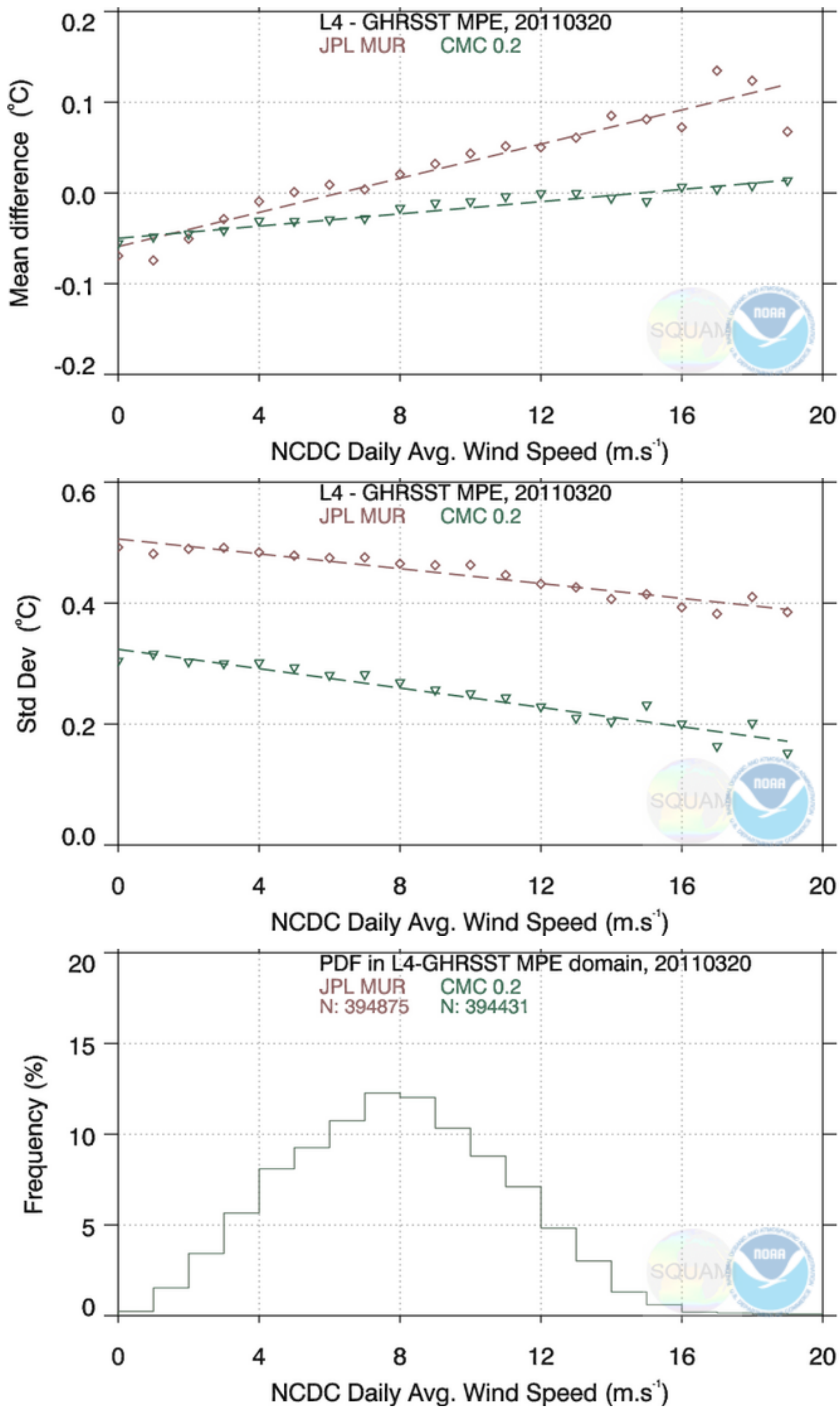
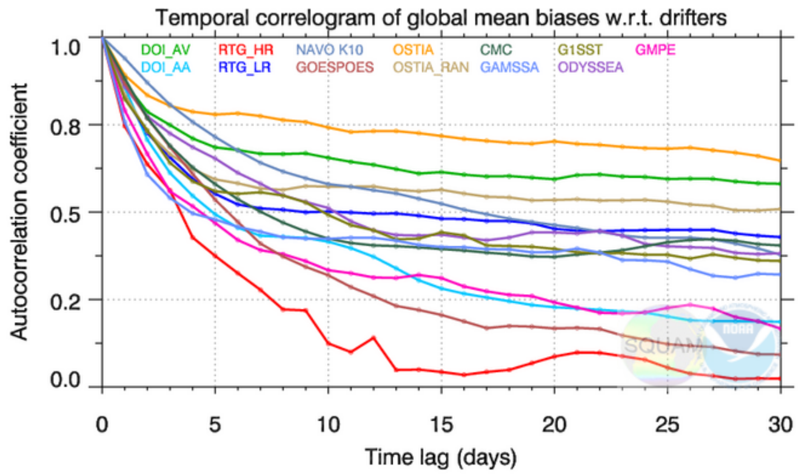
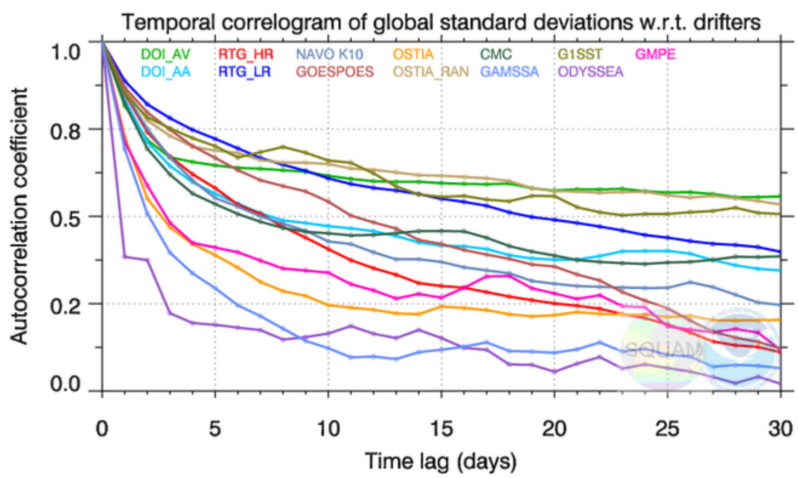


Fig. 6: Dependence of "JPL MUR – GMPE" and "CMC 0.2 – GMPE" ΔT_S on wind-speed. Top-panel: dependence of mean ΔT_S; Middle-panel: dependences of ΔT_S standard deviations; Bottom-panel: Distribution of wind-speed to check where distributions are statistically relevant.



a) Autocorrelation of mean biases: autocorrelation coefficient *vs.* lag in days



b) Same as in a) but for standard deviations

Fig. 7: Correlograms for daily time series data of "L4 *minus* Drifters". Top-panel: autocorrelation coefficients of mean biases; Middle-panel: same as in top-panel but for standard deviation; Bottom-panel: number of match-ups. For top and middle panels, X-axis shows time lag in days (*k* = 0, 1, 2, ... 30).