Author's personal copy

Deep-Sea Research II 77-80 (2012) 31–43

Contents lists available at SciVerse ScienceDirect

Deep-Sea Research II

journal homepage: www.elsevier.com/locate/dsr2

Group for High Resolution Sea Surface Temperature (GHRSST) analysis fields inter-comparisons—Part 2: Near real time web-based level 4 SST Quality Monitor (L4-SQUAM)

Prasanjit Dash [a,b,*], Alexander Ignatov [a], Matthew Martin [c], Craig Donlon [d], Bruce Brasnett [e], Richard W. Reynolds [f], Viva Banzon [g], Helen Beggs [h], Jean-Francois Cayula [i], Yi Chao [j], Robert Grumbine [k], Eileen Maturi [a], Andy Harris [a,l], Jonathan Mittaz [a,l], John Sapper [m], Toshio M. Chin [n], Jorge Vazquez-Cuervo [n], Edward M. Armstrong [n], Chelle Gentemann [o], James Cummings [p], Jean-François Piollé [q], Emmanuelle Autret [q], Jonah Roberts-Jones [c], Shiro Ishizaki [r], Jacob L. Høyer [s], Dave Poulter [t]

[a] NOAA/NESDIS, Center for Satellite Application and Research (STAR), Camp Springs, MD, USA
[b] Colorado State University, Cooperative Institute for Research in the Atmospheres (CIRA), Fort Collins, CO, USA
[c] Met Office, Exeter, UK
[d] ESA/ESTEC, Earth Observation Mission Science Division, 2201 AZ Noordwijk, The Netherlands
[e] Canadian Meteorological Centre, Dorval, Quebec, Canada
[f] NOAA/Cooperative Institute for Climate and Satellites (CICS), Asheville, NC, USA
[g] NOAA/National Climate Data Center (NCDC), Asheville, NC, USA
[h] Centre for Australian Weather & Climate Research, Bureau of Meteorology, Melbourne, Australia
[i] QinetiQ North America, Technology Solutions Group, Stennis Space Center, MS 39522, USA
[j] Remote Sensing Solutions, Inc., Pasadena, CA, USA
[k] NOAA/NWS/NCEP, Camp Springs, MD, USA
[l] University of Maryland, Cooperative Institute for Climate and Satellites (CICS), College Park, MD, USA
[m] NOAA/NESDIS, Office of Satellite Processing & Operations (OSPO), Camp Springs, MD, USA
[n] JPL/Caltech, Pasadena, CA, USA
[o] Remote Sensing Systems, Santa Rosa, CA, USA
[p] Naval Research Laboratory, Monterey, CA, USA
[q] Ifremer, Spatial Oceanography Laboratory, CERSAT, Brest, France
[r] Japan Meteorological Agency, Tokyo, Japan
[s] Danish Meteorological Institute, Copenhagen, Denmark
[t] Formerly at National Oceanography Centre, Southampton, SO14 3ZH, UK

ARTICLE INFO

Available online 17 April 2012

Keywords:
Sea surface temperature
Intercomparison
Climate data
Sea ice
Data centers

ABSTRACT

There are a growing number of level 4 (L4; gap-free gridded) sea surface temperature (SST) products generated by blending SST data from various sources which are available for use in a wide variety of operational and scientific applications. In most cases, each product has been developed for a specific user community with specific requirements guiding the design of the product. Consequently differences between products are implicit. In addition, anomalous atmospheric conditions, satellite operations and production anomalies may occur which can introduce additional differences. This paper describes a new web-based system called the L4 SST Quality Monitor (L4-SQUAM) developed to monitor the quality of L4 SST products.

L4-SQUAM intercompares thirteen L4 products with 1-day latency in an operational environment serving the needs of both L4 SST product users and producers. Relative differences between products are computed and visualized using maps, histograms, time series plots and Hovmöller diagrams, for all

* Corresponding author at: NOAA/NESDIS, Center for Satellite Application and Research (STAR), Camp Springs, MD, USA.
Tel.: +1 301 763 8053x168; fax: +1 301 763 8572.
E-mail addresses: prasanjit.dash@noaa.gov, prasanjit.dash@colostate.edu (P. Dash), alex.ignatov@noaa.gov (A. Ignatov), matthew.martin@metoffice.gov.uk (M. Martin), craig.donlon@esa.int (C. Donlon), bruce.brasnett@ec.gc.ca (B. Brasnett), richard.w.reynolds@noaa.gov (R.W. Reynolds), viva.banzon@noaa.gov (V. Banzon), h.beggs@bom.gov.au (H. Beggs), jean-francois.cayula@QinetiQ-NA.com (J.-F. Cayula), ychao@remotesensingsolutions.com (Y. Chao), robert.grumbine@noaa.gov (R. Grumbine), eileen.maturi@noaa.gov (E. Maturi), andy.harris@noaa.gov (A. Harris), jon.mittaz@noaa.gov (J. Mittaz), john.sapper@noaa.gov (J. Sapper), mike.chin@jpl.nasa.gov (T.M. Chin), jorge.vazquez@jpl.nasa.gov (J. Vazquez-Cuervo), edward.armstrong@jpl.nasa.gov (E.M. Armstrong), gentemann@remss.com (C. Gentemann), cummings@nrlmry.navy.mil (J. Cummings), jean.francois.piolle@ifremer.fr (J.-F. Piollé), emmanuelle.autret@ifremer.fr (E. Autret), jonah.roberts-jones@metoffice.gov.uk (J. Roberts-Jones), s_ishizaki@met.kishou.go.jp (S. Ishizaki), jlh@dmi.dk (J.L. Høyer).

0967-0645/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.dsr2.2012.04.002

Author's personal copy

combinations of products. In addition, products are compared to quality controlled *in situ* SST data (available from the *in situ* SST Quality Monitor, *iQUAM*, companion system) in a consistent manner. A full history of products statistics is retained in L4-SQUAM for time series analysis. L4-SQUAM complements the two other Group for High Resolution SST (GHRSST) tools, the GHRSST Multi Product Ensemble (GMPE) and the High Resolution Diagnostic Data Set (HRDDS) systems, documented in part 1 of this paper and elsewhere, respectively.

Our results reveal significant differences between SST products in coastal and open ocean areas. Differences of $>2\,^{\circ}$C are often observed at high latitudes partly due to different treatment of the sea-ice transition zone. Thus when an ice flag is available, the intercomparisons are performed in two ways: including and excluding ice-flagged grid points. Such differences are significant and call for a community effort to understand their root cause and ensure consistency between SST products. Future work focuses on including the remaining daily L4 SST products, accommodating for newer L4 SSTs which resolve the diurnal variability and evaluating retrospectively regenerated L4 SSTs to support satellite data reprocessing efforts aimed at generating improved SST Climate Data Records.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Satellite-based sea surface temperature (SST) products have been operationally derived from low earth orbiting (LEO) and geostationary (GEO) platforms, initially at National Environmental Satellite, Data, and Information Service (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; Le Borgne et al., 2007; Maturi et al., 2008). Satellite level 2 (L2; data at the observed pixels) products are derived from level 1B (L1B; raw data with appended calibration and Earth location information) brightness temperatures and may be further processed into level 3 (L3; gridded data with gaps) 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 (Martin et al., 2012). There is a variety of real-time and research applications requiring global L4 fields. These applications include seasonal and short-term weather forecasting, fisheries and coral-reef monitoring and the development of SST retrieval algorithms employing radiative transfer simulations. The L4 SSTs, in particular those with a longer history, are invaluable for generating Climate Data Records (CDRs, defined by the United States National Research Council as "A time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change."). Their retrospective and near real-time analyses are crucial for monitoring climate changes.

In order to satisfy these requirements for global SST information, there are now approximately twenty global L4 products produced worldwide. This poses a challenge to understand their relative merit and performance, in terms of data coverage, resolution and accuracy. To assist with this challenge, we have created L4-SQUAM, an "L4 inventory" with comparison tools that can help users to choose a product appropriate for their application, as well as provide feedback to data producers that could help them improve their products.

In producing an L4 SST, the goal is to optimally blend SST data from different sources so that analysis error is minimized. Despite this objective, inconsistencies between these products exist. Differences of several degrees appear regionally between various products, particularly at high latitudes, in the vicinity of Western boundary currents and in semi-enclosed basins, *e.g.*, the Mediterranean Sea and the Gulf of California. The time series of global statistics also reveal that some products cluster in groups. For example, the analyses of foundation SST, the temperature of the water column at a depth where the temperature is free of diurnal variability (Donlon et al., 2007), tend to be similar while significant differences may be observed between other products. 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., 2012), the Level-4 SST Quality Monitor (L4-SQUAM; Part 2, this study) and the High-Resolution Diagnostic Data Set (HR-DDS; Donlon et al., 2009). The major objective of Part 2 is to document the L4-SQUAM system and illustrate how the functionalities of this system can be used to quickly evaluate the consistency between these various L4 fields.

To date, 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 product combinations, within $\sim$24 h of their availability, and plots global maps, histograms, time series and Hovmöller plots of SST differences. Also, to understand the differences between ice masks, analyses in L4-SQUAM are performed two ways, both "including" and "excluding" ice masks, when corresponding ice flags are available. 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 (*iQUAM*; http://www.star.nesdis.noaa.gov/sod/sst/iquam/). Most L4 SSTs include *in situ* data in their blending methods and are therefore not independent of

Author's personal copy

these data. For example, a less accurate analysis that gives a large weight to *in situ* data will agree better with those data than a more accurate analysis that gives little weight to *in situ* data. Thus care must be taken when interpreting the fit-to-data statistics. However, generating consistent validation statistics against the same data provides an easy way to compare all products. Ideally, all L4 products should be produced in a common data format and conform to GHRSST standards, which includes listing data sources used to produce the analysis. This would allow products to be validated against an independent data source, *e.g.*, Argo floats (*e.g.*, Part 1, Martin et al., 2012) or ship-borne infrared radiometers (Donlon et al., 1998, 2011; Minnett et al., 2001). Also, the *in situ* drifters, ships and buoys data which are not included in the blending procedure provide a ready source of independent observations. The advantage of adding independent Argo data to an "*in situ* inventory", such as the *iQUAM*, has been recognized by its developers (*cf.*, Xu and Ignatov, 2011) and will be explored in the future. Unfortunately, there is currently no publicly available community-consensus ship-borne 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. Intercomparison 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 assumption for L4-SQUAM analyses is that paired differences, ΔT_S="$L4_i - L4_j$" or "$L4_i - in\ situ$", are approximately 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 OISST (AVHRR, AVHRR+AMSR-E) as described in Reynolds et al. (2007), referred herein as AVHRR_OI and AVHRR_AMSR_OI, 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 analyses (see Table 1 for details). Many of the products included in L4-SQUAM are also included in GMPE and described in Part 1 (Martin et al., 2012). However, there are some differences between the GMPE and L4-SQUAM inputs. The products monitored in L4-SQUAM are listed in Table 1.

The SST products listed in Table 1 comply with GHRSST standards and specifications except the RTG low resolution product. 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). Each of the L4 SSTs listed in Table 1 are designated with a type listed above. *[Note that the Reynolds and RTG SSTs are adjusted to in situ SST and are often referred to as "bulk" SSTs; however, this term is not recommended by the GHRSST. Nevertheless, "bulk" is comparable to "depth" SSTs which according to GHRSST is defined as measurements beneath the sub-skin, such as from drifting buoys and vertical profiling floats, at depths ranging from 10^{-2}–10^3 m].* The OSTIA, CMC, GAMSSA, G1SST, MUR, RSS, MGDSST, ODYSSEA and DMI products are referred to as "foundation SSTs". These analysis schemes minimize the use of retrievals affected by diurnal variability by employing one or more of the following strategies: (a) using only nighttime satellite data; (b) using additional daytime data with wind speed above 6 $m-s^{-1}$ and (c) excluding L2 SSTs flagged by the producer as having high diurnal variability. The input data to all L4 products are also listed in Table 1, along with information about ice masks which allow the user to interactively exclude ice covered grid cells from statistical analyses. Some products have integrated ice information into their SST fields but do not provide a separate mask to identify ice-covered grid cells (*e.g.*, GMPE), whereas other products have been produced without ice information (*e.g.*, NAVO K10). Some products with integrated ice information did not provide a separate mask to extract ice-covered cells in the beginning but added it at a later stage (*e.g.*, CMC in September, 2011). 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, SSTs have to be matched up in space to generate L4 pairs. This may be achieved by: (a) averaging or interpolating all the L4 SSTs 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 (0.01°) and low resolution RTG (0.5°) employing both bilinear interpolation and NN approach. Results are shown in Fig. 1. They unambiguously suggest that the effect of the 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.

In L4-SQUAM, analyses are performed in two ways. As an example, for a combination of OSTIA and CMC, differences are calculated both as "OSTIA−CMC" and "CMC−OSTIA". The second term is the product to which the NN matching is done (*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. Note that statistics with respect to any L4 are available on the L4-SQUAM webpage and the ones used here are for illustration only. Also, not all graphs discussed in this paper are reproduced here (*e.g.*, comparison with ship and buoy observations), but interested readers are invited to view these graphs online.

3.1. Maps and histograms of ΔT_S

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

Over most of the global ocean, ΔT_S is close to zero. However, the differences are prominent in the southern oceans, where GAMSSA is >1 °C warmer with respect to OSTIA over some regions, and in the Arctic, where the magnitude of differences

Author's personal copy

Table 1
List of L4 SST products monitored or considered in L4-SQUAM. Note that AMSR-E data was an input to most of the L4 SSTs listed here but its production has been suspended on 04 October 2011.

Product	Space/time res. & type	Abbreviation & mode	Reference	Availability period, data format, and ftp source,	Input data				Ice mask
					Infrared	Microwave	*Insitu*	Other	
Products fully implemented in L4-SQUAM									
Optimal interpolation SST	0.25° Daily depth (bulk)	AVHRR_OI 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	√
		AVHRR_AMSR_OI NRT; delayed reanalysis		01-Jun-2002 to 04-Oct-2011, netCDF ftp://eclipse.ncdc.noaa.gov/pub/OI−daily-v2/NetCDF	AVHRR	AMSR-E* (*suspended on 04-Oct-2011)	√	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	-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 physical retrievals	-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-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. center 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, WindSat	√	CMC ice	√ from Sep., 2011
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, WindSat	√	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., 2012	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	mw_ir_oi 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	DMI_OI NRT	Høyer and She, 2007	Jan-2011 to present, netCDF ftp://ftpserver.dmi.dk/GBL005/ (passwd)	AVHRR (GAC, HRPT), SEVIRI, AATSR, MODIS	AMSR-E, TMI	X	O&SI SAF ice	√
Naval res. lab. NCODA analysis	~0.08° 6 hourly depth	NCODA NRT	Cummings, 2005	Jun-2009 to present, IEEE binary, direct access, http://www.usgodae.org/cgi-bin/datalist.pl?summary=Go&dset=fnmoc_ghrsst	AVHRR (GAC, LAC), SEVIRI, GOES, MTSAT-2	AMSR-E, Windsat (coming)	√	6-h SSM/I & SSMIS ice	X

Author's personal copy

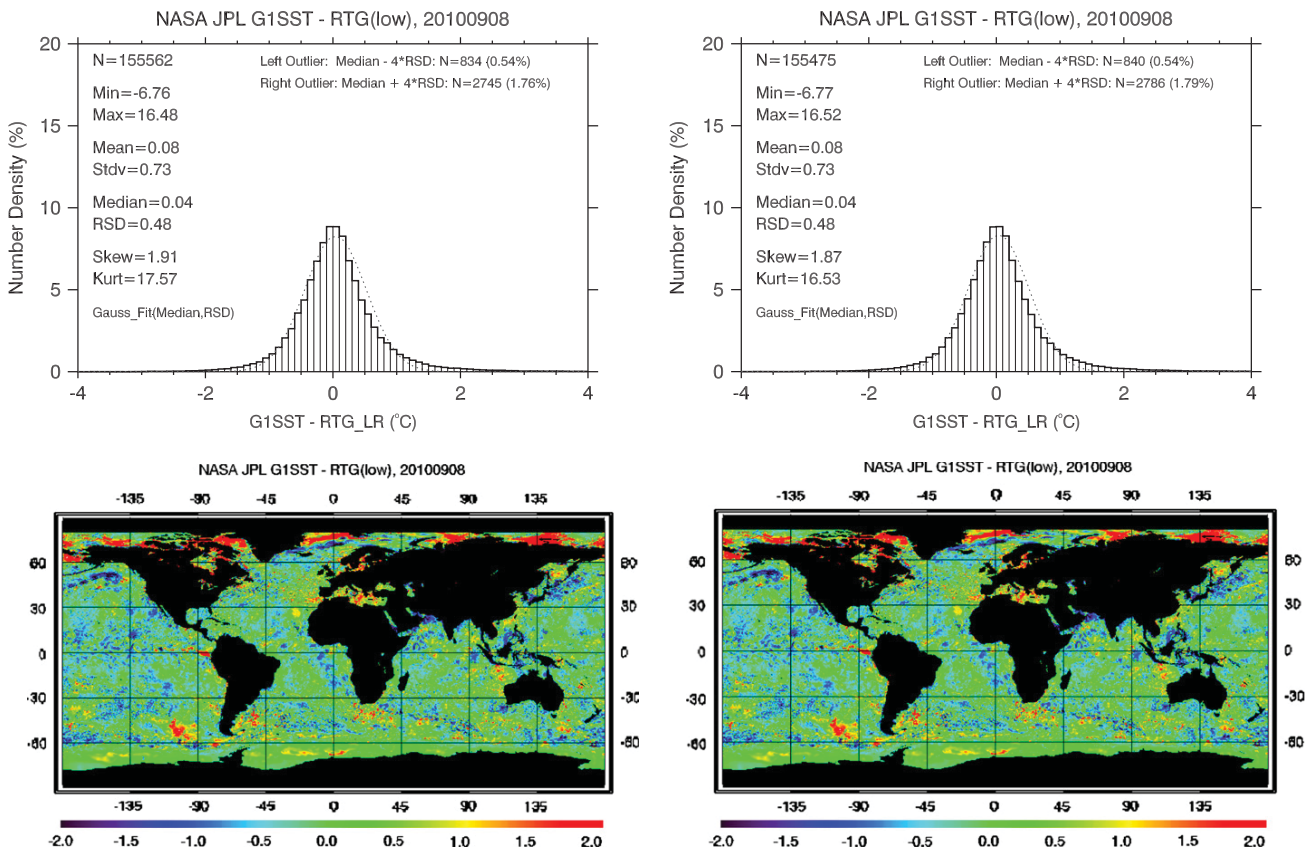

Fig. 1. Effect of interpolation on merging L4 SST fields (0.01° ultra high resolution G1SST *minus* 0.5° lat-lon RTG). Statistical moments are annotated on the histograms (see Section 3.1 for description). Left panels: nearest neighbor selection anchored to RTG; Right panels: bilinear interpolation of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

may exceed 2 °C. Differences of both signs are also observed in many coastal locations. Also, different combinations of L4s show different patterns and magnitudes of differences. For instance, for 13 July 2011, AVHRR_OI shows highly variable differences with respect to OSTIA reaching more than ± 1 °C (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 the number of SST pairs, minimum, maximum, mean, standard deviation (Std Dev), median, robust standard deviation (RSD), skewness and kurtosis. *[RSD here is defined as: IQR/S, where, IQR is interquartile range (75th percentile—25th percentile, in an ordered dataset) and S is a scaling factor which is 1.348 for an ideal normal distribution, cf., Merchant and Harris, 1999]*. The number of SST pairs approximately represents the number of valid OSTIA SSTs because NN matching is done to OSTIA grid. A dotted gray line shows an ideal Gaussian fit, $X \sim N(Median, RSD)$. Additionally, numbers of SST pairs 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 $\sim$0.69 °C and RSD $\sim$0.36 °C.

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 found largely in the Arctic (Fig. 2A), suggesting differences in treatment of ice in the two products. Both L4 products contain ice masks which are derived from 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 either mask or both. The statistics change significantly. First, the number of SST pairs is reduced by $\sim$20%, from $\sim$16.8 million in "all-grid" to $\sim$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" ~ -2 °C in at least one of the products. There are grid points in which the ice cells have the same values for both products, resulting in an artificial spike at zero in Fig. 2B. On the other hand, there are also grid cells 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 for the "ice-free" sample, and the Std Dev is reduced from $\sim$0.69 °C to $\sim$0.59 °C. However, the apparent worsening of skewness (compare Fig. 2B with Fig. 2D) is related to its decrease in Fig. 2B, caused by the artificial small mode in the -1.3 °C to -1.5 °C bins (Fig. 2B). Excluding icy pixels can also increase the Std Dev for those combinations of L4s where the assumed value of SST in partially ice-covered regions is the same, *e.g.*, for "AVHRR_OI *minus* AVHRR_AMSR_OI" (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. The "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.

Author's personal copy

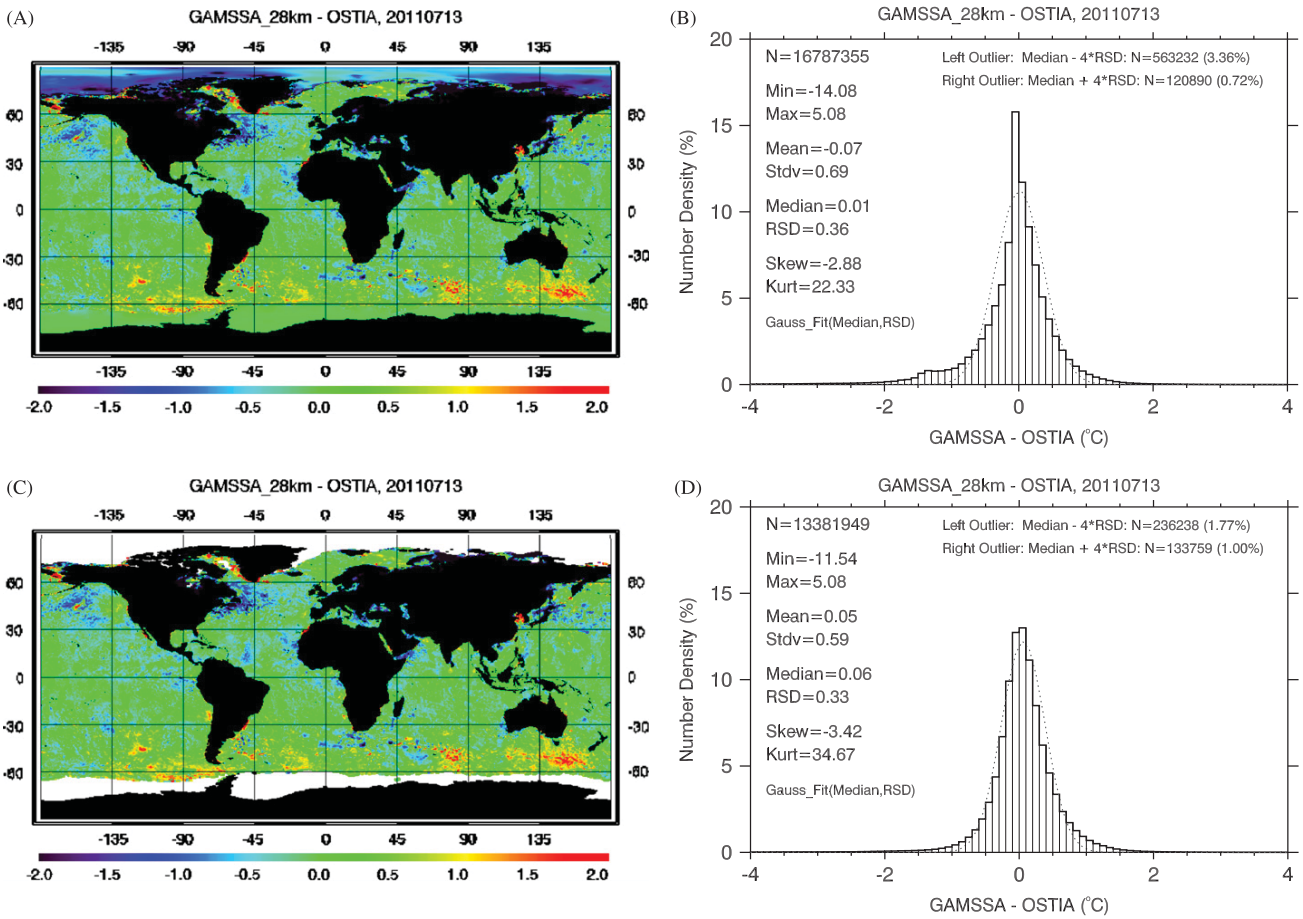

Fig. 2. In the left panels, spatial differences between GAMSSA and OSTIA are mapped. These are generally close to zero but prominent in some areas, *e.g.*, roaring forties and in many coastal locations. The arctic ice areas also show significant differences. In the right panels, ΔT_S statistics are annotated on the left side of the histograms, dotted gray line shows an ideal Gaussian fit, and the numbers of L4 match-ups beyond "Median $\pm 4 \times$ Robust Std Dev" are shown on the top right. Note that due to NN interpolation, anchored to the second term (*i.e.*, OSTIA), the number of SST pairs "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) 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.

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.

Figs. 3A and B show examples of global "ice-free" mean differences and standard deviations in L4 fields with respect to AVHRR_OI, Figs. 3C and D show the same statistics with respect to drifters and Figs. 3E and F show the same with respect to GMPE.

The time series in Fig. 3 are very busy due to a large number of L4 products. However, users of the L4-SQUAM webpage can perform interactive analyses by plotting and focusing on time series for one or several products. It is also possible to interactively apply a time filter to the statistics with the period of the filter specified by the user.

The majority of the products are within ± 0.15 °C of each other. For example, the two daily NOAA OISST products are largely consistent, with AVHRR_AMSR_OI being $\sim +0.05$ °C warmer than AVHRR_OI. However, there are a few noticeable exceptions. For instance, G1SST is generally colder (between $+0.05$ to -0.2 °C) with respect to AVHRR_OI. Similarly, a cold bias relative to AVHRR_OI is 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 AVHRR_OI, 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 AVHRR_OI. The CMC was from 0.0 to 0.2 °C warmer than AVHRR_OI until about the end of 2004, after which time the two products have shown a negligible mean bias. Also, a pre-2006 trend which flattens out subsequently is observed. This coincides with the change in input of NOAA OISST products from Pathfinder to Naval Oceanographic Office (NAVOCEANO) SST on 1 January 2006 (Reynolds et al., 2007).

There appears to be clustering of some products into groups. For example, the RTG_HR and NESDIS Multi-SST analysis products closely follow each other. 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 °C with respect to AVHRR_OI, which reduced to -0.1 °C later in 2006 but then briefly spiked again in February 2007, May 2008 and May 2009. *[OSTIA reanalysis excluding ice has not been processed in L4-SQUAM yet and consequently is not shown in Figs. 3A and B; work is underway to add it]*.

The standard deviations with respect to AVHRR_OI show a clear seasonal cycle, for all L4 products, but with different

Author's personal copy

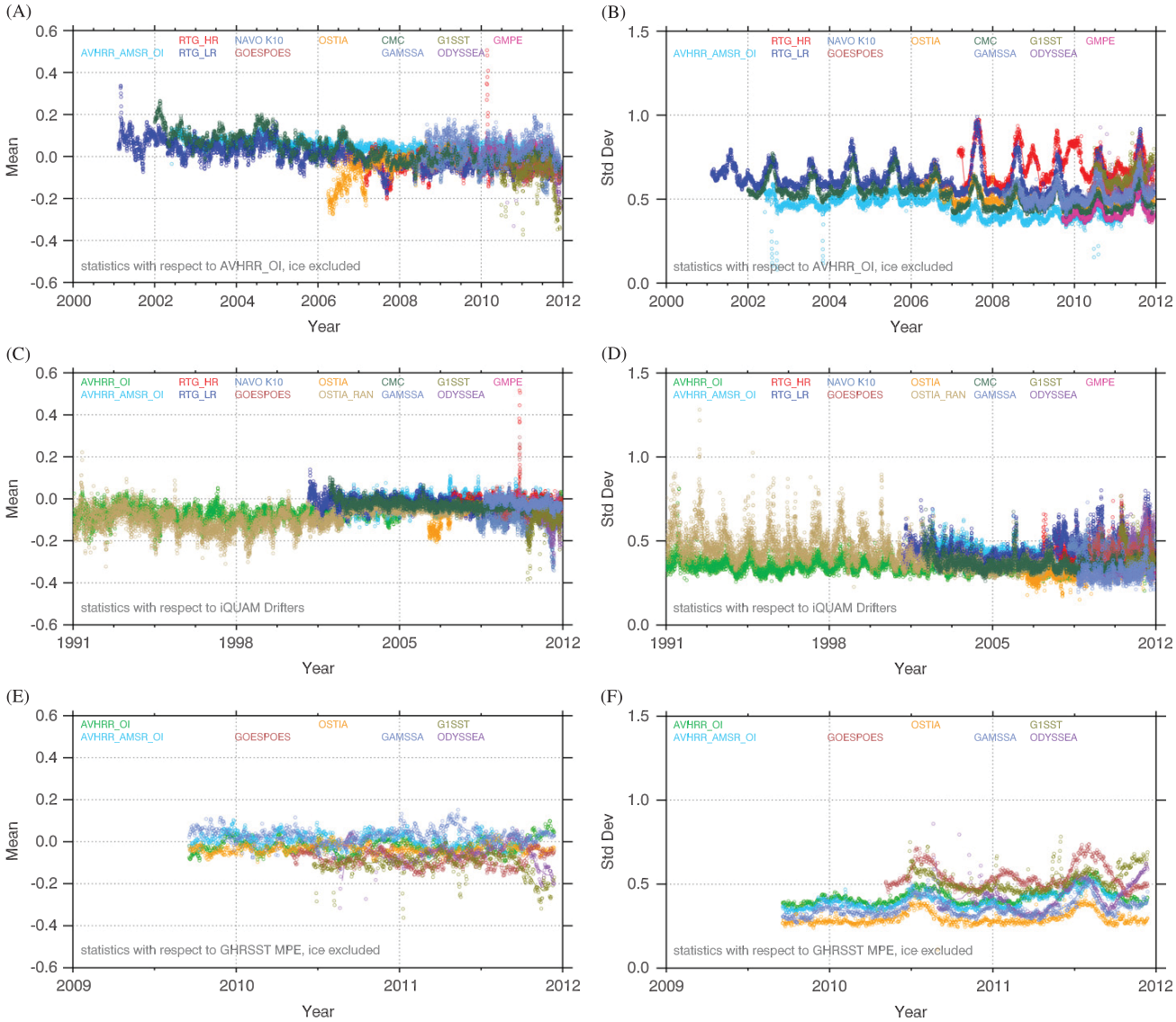

Fig. 3. Mean and standard deviation of ΔT_S. Left-panels: mean; Right-panels: standard deviation. Top-panels: statistics with respect to Reynolds (AVHRR) excluding ice grids; Middle-panels: same as top-panels but with respect to drifters; Bottom-panels: same as top-panel but with respect to GMPE. (A) 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.

amplitudes. For instance, the two RTG, G1SST, and ODYSSEA products show standard deviations between ~0.5 and 0.95 °C. For OSTIA, K10 and GAMSSA, with respect to AVHRR_OI, the standard deviations range between 0.45 and 0.65 °C, and the NESDIS Multi-SST analysis shows slightly higher values. The two NOAA OISST products are very consistent. A clear discontinuity in the Std Dev is also observed for "CMC *minus* AVHRR_OI" and "RTG_LR *minus* AVHRR_OI" on January 1, 2007. On that day, the volume of satellite data used by AVHRR_OI effectively doubled when retrievals from NOAA-18 were added to the retrievals from NOAA-17 that had been used previously.

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 *iQUAM*.

Figs. 3C and D show global mean bias and standard deviation in L4 products with respect to drifters. Many of the observations in Figs. 3A and B are also reproduced in Figs. 3C and D, but with a reduced magnitude. For example, "RTG_HR *minus* AVHRR_OI" Std Dev ranges between 0.5 and 0.95 °C with strong seasonality, whereas for "RTG_HR *minus* Drifters" it ranges between 0.35 and 0.55 °C. 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 and 0.5 °C and shows patterns similar to "RTG_HR *minus* Drifters" ("RTG_HR *minus* GMPE" is available only for all-grids as neither L4 provides an ice mask). These results highlight the utility of GMPE as a reference field. (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., 2012) 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 some noteworthy characteristics, not shown here in the interest of space. All the L4 SSTs show negative differences when compared to ship data, *i.e.*, ship records are warmer due to engine intake, and also show much stronger seasonality (*cf.*, Xu and Ignatov, 2011).

Author's personal copy

The standard deviations with respect to ship data are also much higher ranging from 0.75 to 1.3 °C. One interesting observation in the "L4 *minus* Ships" mean differences is seasonal (sinusoidal) patterns of comparable amplitudes but different *signs*. For example, the plots for CMC and RTG seem to be anti-correlated to the plots shown by AVHRR_OI and AVHRR_AMSR_OI.

Validation statistics against coastal moorings also vary significantly between different products. For example, the Std Dev

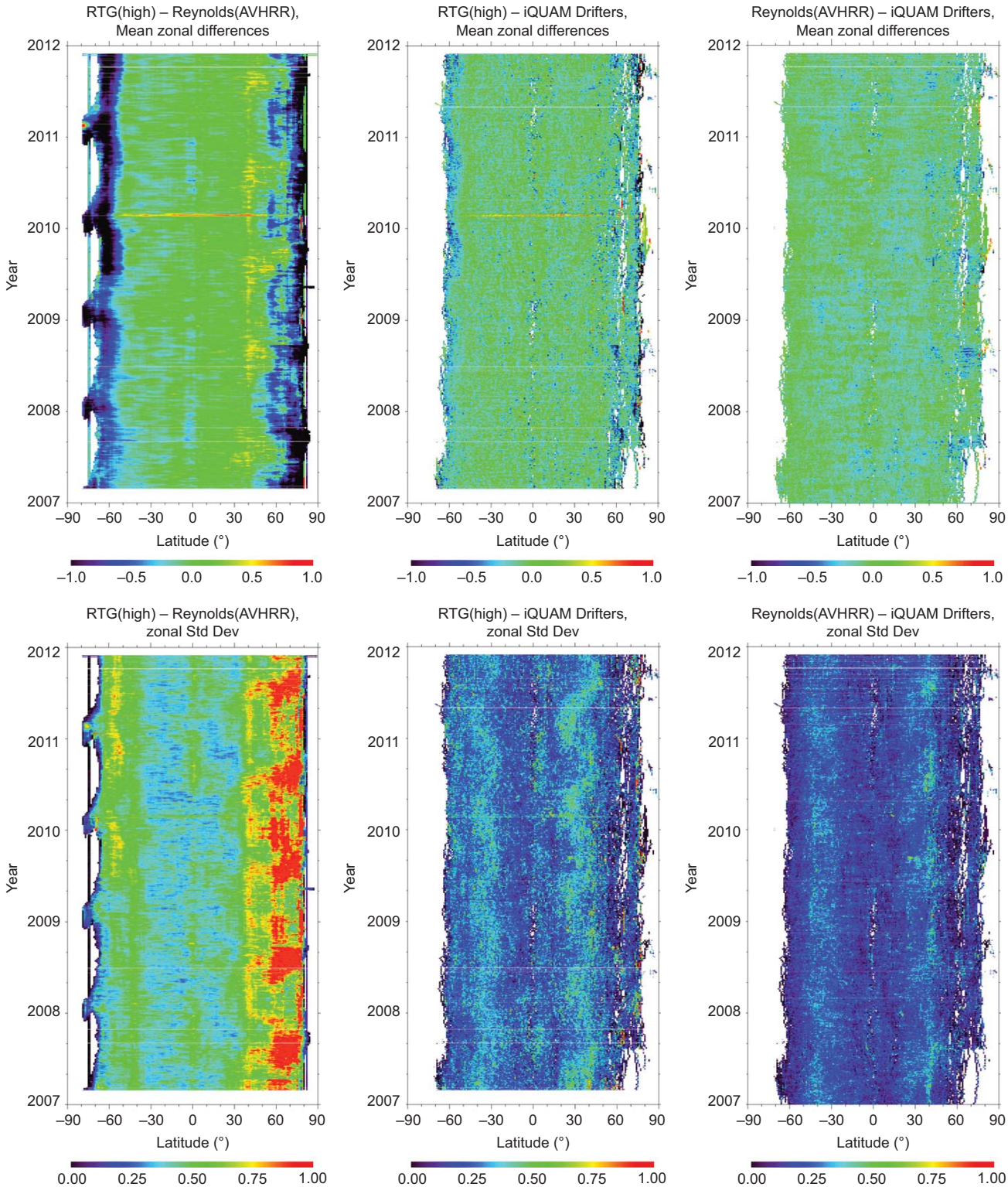

Fig. 4. Hovmöller diagrams of average zonal differences: *First column*: RTG (high)– Reynolds, ice excluded; *Second column*: RTG (high) – Drifters; *Third column*: Reynolds (AVHRR) – Drifters; Top-panels: mean differences; Bottom-panels: standard deviations.

Author's personal copy

approximately ranges from 0.35 to 0.8 °C for AVHRR_OI and CMC, 0.4 to 1.0 °C for the NESDIS Multi-SST, 0.38–1.5 °C for G1SST, and 0.6–1.4 °C for RTG, K10, GAMSSA, ODYSSEA and GMPE. There was an increase in the Std Dev for OSTIA on 28 July 2009. Prior to this date, the Std Dev ranged between 0.2 and 0.6 °C, but after this date, the Std Dev ranges from 0.4 to 1.6 °C. The reasons for these differences are not fully apparent at this stage and should be further investigated.

Although some pairs of products show a close to zero global mean difference, the large standard deviation suggests significant regional differences. These are further analyzed in L4-SQUAM 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 (two left panels) shows example of Hovmöller diagrams of ice-free mean biases and standard deviations for "RTG_HR *minus* AVHRR_OI".

On average, the two 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 standard deviations are small in the sub-tropical oceans, increasing in the Inter-Tropical Convergence Zone (ITCZ) and the mid-latitudes, and reaching 0.75–1 °C from 40°N to 75°N. The cause of these differences is not fully clear. Recall that AVHRR_OI uses the NAVOCEANO L2 SST as input, whereas RTG_HR 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 these differences, mean biases and standard deviations of "RTG_HR *minus* Drifters" and "AVHRR_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* AVHRR_OI" mean biases and standard deviations are not captured in the *in situ* validation, suggesting that they exist in the areas not covered by *in situ* data. On the other hand, in areas where *in situ* data are present, both L4s agree well with them. This suggests that drifters are assimilated in both L4s with a relatively large weight and illustrates that using the same *in situ* data to validate an L4 as were ingested by the L4 will yield an unreliable assessment of the true global performance of the product.

Another interesting observation includes warmer biases in GAMSSA over the Southern Ocean and colder biases over the Arctic Ocean compared to most other products (not shown). Over the Arctic Ocean, in fact, most of the products show distinctive seasonal biases with respect to 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 L2 pre-processed (L2P) products. For example, the GAMSSA system removes biases in the input L2P SSTs on a global basis by applying the bias corrections provided by the L2P producers (Cayula et al., 2004). In contrast, the Met Office uses regional AATSR and buoy SSTs to estimate and remove the biases of the L2P inputs (Stark et al., 2007). The Reynolds OISST and CMC systems adjust all satellite inputs for bias on a regional basis using both buoy and ship SSTs (Reynolds et al., 2007; Brasnett, 2008). Besides differences in bias corrections, 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 AVHRR_AMSR_OI SST. Noticeably, NAVOCEANO NOAA-17 and -18 AVHRR SSTs are warmer over the Southern Ocean. Similar patterns are also seen from comparisons

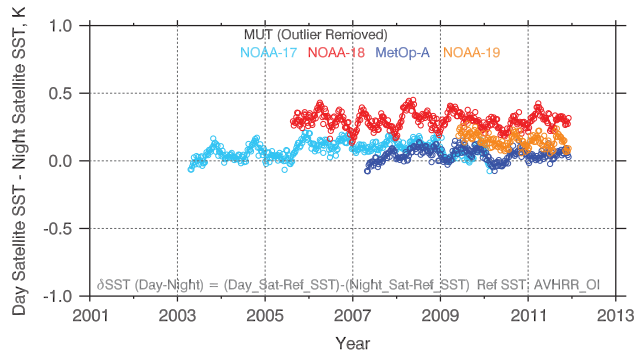

Fig. 5. Average "Day *minus* Night" SST differences estimated employing double differencing (DD) technique, with daily Reynolds OISST as the transfer standard.

between NAVOCEANO AVHRR GAC and AVHRR_OI SSTs (*cf.*, Hovmöller diagrams at http://www.star.nesdis.noaa.gov/sod/sst/squam/NAVO/). 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). Ideally, this should form a goal for the GHRSST L2P data providers. 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 and towards the true values.

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 updated daily, and do not resolve the diurnal cycle. Some L4 developers have started generating diurnal cycle resolved L4 products (*e.g.*, BoM and NCODA produce 3-hourly experimental and 6-hourly operational products, 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). Note that the original Global Ocean Data Assimilation Experiment (GODAE) requirements for GHRSST L4 products are 10 km resolution and 0.5 °C accuracy, available every 6 h (*cf.*, Donlon et al., 2009). Therefore, 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. For example, to compare L4 products of different update cycles, such as 24-hourly products against 6-hourly products, one way is to calculate a daily average of the product with shorter update cycle. Comparison of products with similar update cycles is achievable with the current L4-SQUAM framework.

Analyses by Dash et al. (2010) suggest that one could validate the DV models, by combining satellite L2 products with L4 SSTs. Toward 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

Author's personal copy

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 (by including even those points where either day or night L2 SST is unavailable) 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 AVHRR_OI as the transfer standard.

The DN values are mostly positive, because day SST is warmer than night SST. In calculation of DDs, the AVHRR L2 SST product is subject to diurnal changes but AVHRR_OI SST field has one daily value per grid which cancels out. 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. (A systematic residual offset between NOAA-18 and -19 of ~0.10 °C 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 the model for removing the diurnal variations from L2 SSTs, or by adding DV amounts 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 geophysical 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".

Both MUR and CMC are foundation SST products. Comparisons within L4-SQUAM indicate that 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 0.1 °C, whereas at high winds it is warmer by 0.1 °C. 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. Later versions of MUR will use *in situ* data.), which reduces with increasing wind speed. The corresponding standard deviations are largest at low winds (~0.5 °C) and decrease towards larger winds reaching ~0.35–0.40 °C. The CMC product shows similar trends but with lesser magnitudes. Including such dependencies in SQUAM and verifying over a longer time series will help to better understand the cause of these residual biases.

4.3. Correlograms

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

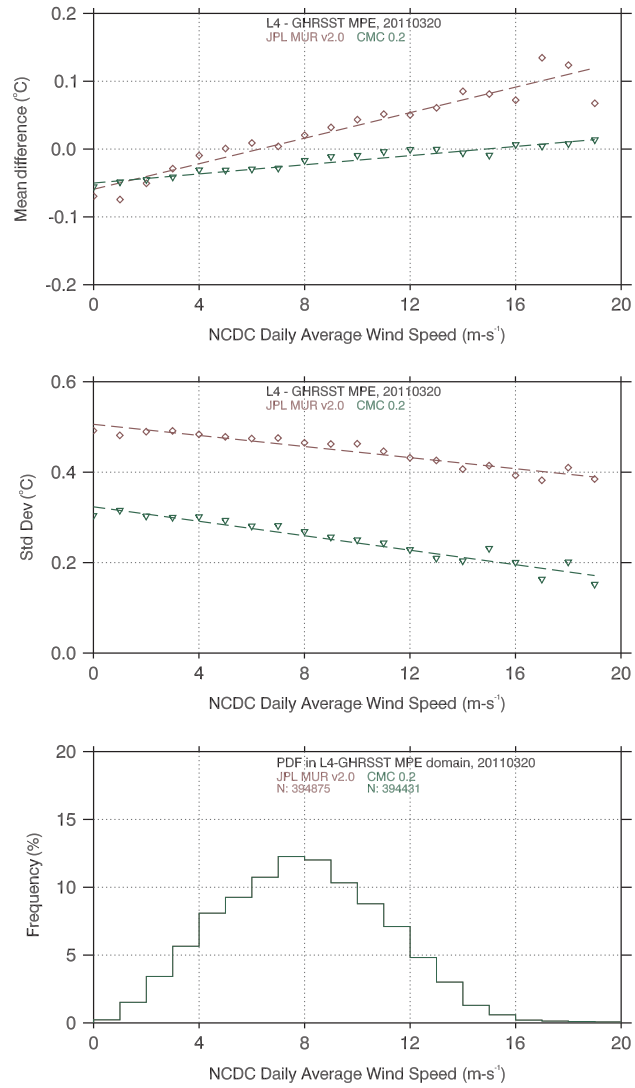

Fig. 6. Dependence of "JPL MUR – GMPE" and "CMC 0.2 – GMPE" ΔT_S on wind-speed. Top-panel: dependence of mean ΔT_S; Middle-panel: dependences of ΔT_S standard deviations; Bottom-panel: Distribution of wind-speed to check where distributions are statistically relevant.

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 $r = \sum(x_t - \overline{x})(x_{t+k} - \overline{x}) / \sum(x_t - \overline{x})^2$. The "$r$" *vs.* "$k$" for time series biases and standard deviations in "L4 *minus* drifters" are shown in Fig. 7 (upper and middle panels), respectively.

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 Figs. 7 upper and middle panels 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 Std Dev for a given product. For example, in Fig. 7 upper panel, OSTIA shows the lowest and RTG_HR shows the highest "randomness", whereas in Fig. 7 middle panel, AVHRR_OI and OSTIA_RAN show lowest and GAMSSA and ODYSSEA show highest "randomness". Fig. 7 upper panel suggests that the bias in some fields, *e.g.*, OSTIA, are rather

Author's personal copy

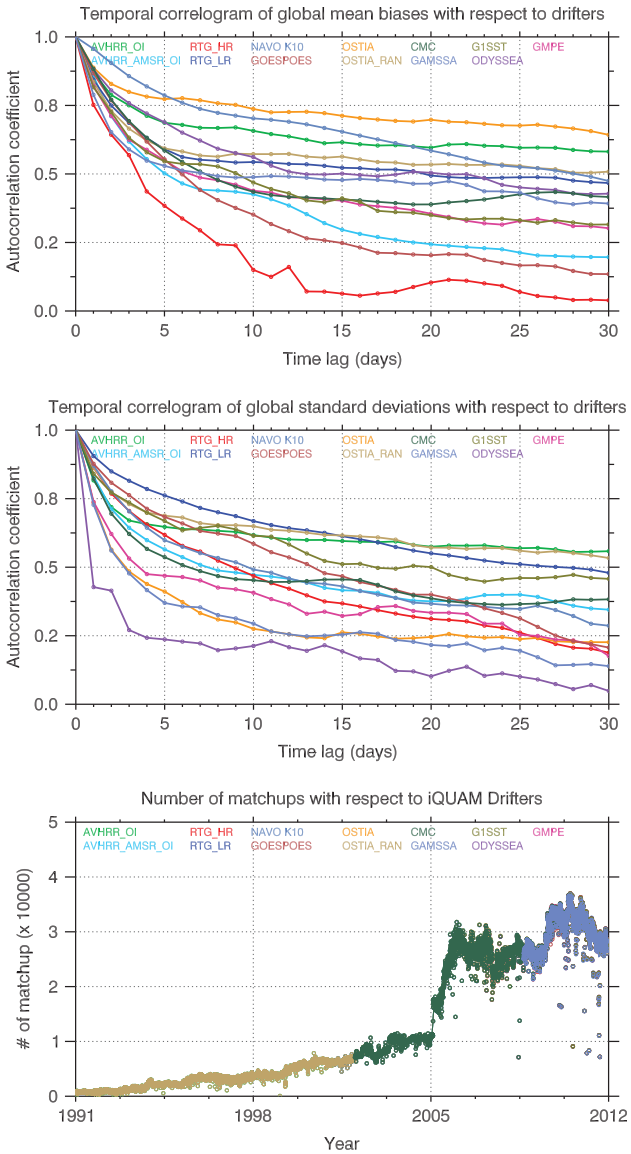

Fig. 7. Correlograms for daily time series data of "L4 *minus* Drifters". Top-panel: autocorrelation coefficients of mean biases; Middle-panel: same as in top-panel but for standard deviation; Bottom-panel: number of match-ups. For top and middle panels, *X*-axis shows time lag in days (k=0, 1, 2, …,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

smooth and consistent with respect to drifters whereas for some fields (*e.g.*, RTG_HR) they are noisier. When interpreted appropriately, the shape of such a correlogram can be used as an indicator of the closeness of the L4 SST to drifters for each product, and also of the level of persistence of the background field, globally. It should be noted that interpretation of such "preliminary" conceptual plots are best performed in conjunction with validation time series and spatial autocorrelation maps. For example, an L4 with high "*r*" but consistent low bias and Std Dev might indicate a higher quality as opposed to an L4 with low "*r*" but consistent high bias and Std Dev.

5. Summary and future work

The web-based L4 SST quality monitor (L4-SQUAM) has been 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., 2012) and the High Resolution Diagnostic Data Set (HR-DDS, Donlon et al., 2009). To date, thirteen daily L4 SSTs are monitored in L4-SQUAM, with two additions underway and four others planned.

L4-SQUAM metrics are based on analyses of "L4 *minus* L4" and "L4 *minus in situ*" ΔT_S. Maps and Hovmöller plots provide synoptic snapshots of similarities and differences between various products, histograms check for their proximity to a Gaussian shape, and time series assess relative stability of consistency statistics. To better interpret the effect of ice masks on these L4 products, analyses are performed in two ways: (a) including ice and (b) excluding ice, when the corresponding information to extract an ice-mask is available within the product. All processing is done automatically within 24 h of data availability to the system.

Our results show that foundation SSTs are more consistent with each other whereas some 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. Our analyses also emphasize the need for diurnally resolved L4 SSTs, and their global validation.

We note that an SST analysis is designed to produce the best estimate of SST for a given time and location, over a regular grid based on irregularly spaced sparse measurement datasets (*cf.*, Donlon et al., 2011). The specified grid resolution of a product defines the smallest possible SST features that could potentially be resolved by the analysis, but grid resolution is often not the same as the end-to-end analysis-system resolution. The length of analysis time window, during which the input data sets are considered "coincident", also varies among the L4 products. The design of analysis parameters and the variable characteristics of the inputs to the analysis thus determine the end-to-end resolution. An analysis may be smoother than the output grid resolution depending on the choice of background error covariance and correlation length scale. Such a product may be perfectly adequate for its intended application (*e.g.*, for numerical weather prediction systems where "noisy" features cause instabilities in the model) but inadequate for another application (*e.g.*, monitoring of SST frontal dynamics). Coarse resolution input data (*e.g.*, passive microwave SST at ~50 km resolution) cannot resolve fine mesoscale features unlike the infrared satellite sensors, and an analysis dominated by the former cannot be expected to deliver high resolution output. Thus high-resolution SST features within each L4 product are dynamic and are visible only when sufficient SST data are available for input: persistent lack of data results in the analysis system reverting to a smooth background climatological value, until new data are again ingested. Users of SST analysis products must be aware that the representation of reality of every analysis system on a given day is extremely dynamic (Donlon et al., 2011). The challenge for all analysis systems is to maximize the signal to noise ratio for a given output grid resolution while maintaining the highest feature resolution possible through the careful choice of analysis design for a given application.

The SQUAM system provides a tool that reveals differences between many operational analyses and is a significant step towards understanding such differences. Maps, histograms and time series plots for all combinations of "L4 *minus* L4" for all available dates are made available for users in an easy to use web-based interface at http://www.star.nesdis.noaa.gov/sod/sst/squam/L4/. Having L4 SSTs uniformly analyzed and compared to the same *in situ* data using a single interface allows L4 SQUAM to provide users and producers of L4 products with valuable information on availability, relative merit for particular applications and potential areas of improvement of

Author's personal copy

these products. We emphasize that it is not the purpose of this paper to determine which data set is the "best" or select "one" product suitable for all applications: it is a user choice to decide which product is better suited to their applications based on diagnostics from GMPE, L4-SQUAM and HR-DDS. Rather, we view the L4-SQUAM as a "best practice" for intercomparison, which does provide diagnostics to identify potential issues in any given product. For example, if one product deviates from the majority of the products for any given region, it is more likely that the problem is in the deviant product.

For all the L4 SST products, we make the following recommendations: (a) sea-ice information and corresponding masks that separate ice-covered grid cells from open water should be provided (ideally as sea ice concentration and sea ice edge); (b) independent reference datasets (*i.e.*, data not assimilated into L4 systems) should be maintained such as the surface SST values derived from Argo floats and un-ingested *in situ* data for consistent validation of products and (c) operational L4 SST data products should be reprocessed to provide consistent outputs for diagnostic and other scientific applications.

Future work to improve the L4-SQUAM system includes tools to estimate the individual contribution of a given product to the observed differences. This may be achieved by employing a three-way error analysis, recently applied by O'Carroll et al. (2008) allowing individual errors for a given combination of three datasets to be derived within SQUAM (assuming products have mutually independent errors required for a three-way error analysis). Product-specific spatial error fields may be derived, rather than a single global mean (*cf.*, Xu and Ignatov (2011), who explored derivation of error fields using Pathfinder SST, AVHRR_OI and *in situ* data). Time-averaged L4 SST differences, *e.g.*, monthly mean difference maps, may also be useful for identifying persistent and seasonal features, as has been suggested by some L4 producers. Potential use of dependence plots was also discussed in this study. As the SQUAM system matures, these features will be considered as part of the long-term effort to provide users and producers of L4 SST products with effective and useful diagnostic tools that facilitate improvement of these products and their applications.

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/CIRA) for setting up the *iQUAM* system and many NESDIS SST colleagues for helpful discussions. E. Armstrong, M. Chin and J. Vazquez acknowledge that the research was carried out, in part, by the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration. NAVO K10 product was provided courtesy of D. May and B. McKenzie (NAVO). We thank the two anonymous reviewers for their careful reviews and many helpful suggestions which improved this article. 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. Aust. Meteorol. Oceanogr. J. 61, 1–22.

Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, *CA*, 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, http://dx.doi.org/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.

Cayula, J., May, D., McKenzie, B., Olszewski, D., Willis, K., 2004. Reliability estimates for real-time sea surface temperature. Sea Technol. 45 (2), 67–73.

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, http://dx.doi.org/10.1175/2009JTECHO592.1.

Cummings, J.A., 2005. Operational multivariate ocean data assimilation. 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, http://dx.doi.org/10.1175/2010JTECHO756.1.

Donlon, C.J., Martin, M., Stark, J.D., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2011. The operational sea surface temperature and sea ice analysis (OSTIA) system. Rem. Sens. Environ. http://dx.doi.org/10.1016/j.rse.2010.10.017.

Donlon, C., Casey, K., Robinson, I., Gentemann, C., Reynolds, R., Barton, I., Arino, O., Stark, J., Rayner, N., Le Borgne, P., Poulter, D., Vazquez-Cuervo, J., Armstrong, E., Beggs, H., Llewellyn-Jones, D., Minnett, P., Merchant, C., Evans, R., 2009. The GODAE high-resolution sea surface temperature pilot project. Oceanography 22, 34–45.

Donlon, C., Robinson, I., Casey, K.S., Vázquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., Barton, I., Beggs, H., Poulter, D.J.S., Merchant, C.J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R.W., Kawamura, H., Rayner, N., 2007. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. Bull. Am. 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, http://dx.doi.org/10.1029/2002GL016291.

Høyer, J.L., She, J., 2007. Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea. J. Mar. Syst. 65 (1-4), 2007pp, 2007.

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, http://dx.doi.org/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 Bull. 73, 1–18. in Japanese.

Le Borgne, P., Legendre, G., Marsouin, A., 2007. Operational SST retrieval from MetOp/AVHRR. In: Proceedings of the 2007 EUMETSAT Conference, 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., 2012. Group for high resolution SST (GHRSST) analysis fields inter-comparisons: part 1. A GHRSST multi-product ensemble (GMPE). Deep-Sea Res. II 77–80, 21–30.

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. Am. 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.

Merchant, C.J., Harris, A.R., 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., 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.

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

Author's personal copy

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.In: Proceedings of the Oceans '07 IEEE, Marine Challenges: Coastline to Deep Sea, Ext. abs., 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. 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. Am. 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. Appl. Meteorol. 27, 115–124.

Wu, X., Menzel, P.W., Wade, G.S., 1999. Estimation of sea surface temperatures using GOES 8/9 radiance measurements. Bull. Am. 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, http://dx.doi.org/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〉.