Sentinel 3
SLSTR SST
Validation Report

A Match-up Data Base for S3A/SLSTR SST products validation

Federated activity between OSI SAF and EUMETSAT Secretariat
OSI_CAF_FA 13_02

Version : 1.0

Date : 11/04/2018

Gorm Dybkjaer, DMI
Anne Marsouin, METEO FRANCE
Steinar Eastwood, MET NORWAY
Jean-Francois PIOLLE, IFREMER
Jacob L. Høyer, DMI
Hervé Roquet, METEO FRANCE
Stephane Saux Picart, METEO FRANCE
Anne Ocarroll, EUMETSAT
Igor Tomazic, EUMETSAT

Document Change record

Document version	Data set version	Date	Change description
V 0.1		20171201	Draft version
V 0.2		20171204	Draft version / corrections
V 0.3		20180323	Inclusion of isar data
V 1.0		20180411	Final and approved by authors

Table of contents

1. Introduction

At the GHRSST XIII Science Team meeting in Tokyo (June 2012), some unofficial discussions took place between EUMETSAT and some OSI SAF SST team members, on how the OSI SAF could contribute to S3/SLSTR SST development. It was agreed that in the short term (within CDOP-2), the most reasonable OSI SAF contribution could be in the area of S3A/SLSTR cal/val activities for SST. After some further discussions between OSI SAF and EUMETSAT based on a preliminary proposal prepared by the OSI SAF SST team, it was decided to go for a Federate Activity (FA) proposal between OSI SAF and EUMETSAT Central Facility (CAF), to be approved by the OSI SAF Steering Group [RD-1].
This project and the team behind it, became part of ESA and EUMETSAT's Sentinel 3 Validation team for SST products from the SLSTR instrument.

1.1. Scope

The project includes the following activities:

- The building and delivery of a dedicated S3A/SLSTR MDB for SST cal/val activities, by IFREMER. Since January 2013, IFREMER is funded by ESA to implement an open-source software based on the High-Resolution Diagnostic Dataset concept (HR-DDS) but extending it with MDB and Multi-sensor Match-up Data base (MMDB) capability, similar to the one developed by the ESA CCI SST project.

- A special efford to validate Arctic and North Atlantic SST against Fiducial reference SST measurements (FRM) is performed using Radiometric SST data from the DMI-ISAR instrument, provided by DMI. The collocation of ISAR and S3A/SLSTR skin SST measurements during the S3A/SLSTR cal/val phase is performed by IFREMER in collaboration with DMI. The DMI-ISAR radiometer was planned deployed on a container ship from Royal Arctic Lines, operating between DK and GL, but delays and other obstacles caused a chance of deployment plans. The DMI-ISAR was subsequently deployed on Danish Research Vessel Dana and Passenger ship Norröna, from Smyril Line of the Faroe Islands, operating between Hirtshals (DK), Torshavn (FO) and Seyðisfjörður (IS).

- Norweginan Meteorological Institute, Met Norway, and Meteo France, MF, provide detailed SST cal/val results from the S3A/SLSTR SST MDB built and delivered by IFREMER. Consistently with the current share of expertise in the OSI SAF consortium, the activity at Met Norway focus on the High Latitudes, where satellite SST retrievals have to deal with problems related to presence of sea ice, specific illumination and atmospheric conditions, and the activity at MF focus on the Low and Mid-Latitudes.

This federated activity was approved by the OSI SAF Steering Group on 4 October 2013 to start in 2014. The kick-off of the activity was held on 18/11/2014 in Toulouse. The launch of S3A was delayed until 16/02/2016. First S3A/SLSTR data were available in November 2016 together with reprocessed data from July 2016).
First results were presented at the EUMETSAT conference in October 2017, final results are available in this report.

1.2. Report structure

The report is divided into 5 chapters. Chapter 2 describes the applied in situ data used for match up. The validation methodology is described in chapter 3. Results are presented in chapter 4 in sub sections that reflect the validation areas: 4.1 Mid and Low Latidudes vs Coriolis data; 4.2 High Latidude validation vs Coriolis data; and 4.3 High Latidude validation vs ISAR radiometer data. Conclusions are in chapter 5

1.3. Reference documents

[RD-1] Federated activity proposal, between OSI SAF and CAF (EUMETSAT Secretariat). A Match-up Data Base for S3A/SLSTR SST products validationOSI_CAF_FA13_02, 2013.

Project Wiki, including project interactions, documents, meeting minutes and time line:
https://wiki.met.no/osisaf-pt/fa-sentinel-3-calval/start#content

(User credentials are administered by Steinar Eastwood, <steinare@met.no>)

1.4. Definitions, acronyms and abbreviations

CMEMS	Copernicus - Marine environment monitoring service
DB	Data Base
DMI	Danish Meteorological Institute
ECMWF	European Centre for Medium range Weather Forecast
FTP	File Transfer Protocol
ISAR	Infrared Sea Surface Temperature Autonomous Radiometer
MET	Norwegian Meteorological Institute
MIZ	Marginal Ice Zone
MF	Meteo France
MDB	Martch-Up Data Base
MU	Match-Up
NWP	Numerical Weather Prediction
OSI SAF	Ocean and Sea Ice SAF
PB	Production Baseline
SAF	Satellite Application Facility
SLSTR	Sea and Land Surface Temperature Radiometer
SST	Sea Surface Temperature
TCWV	Total Column WV
WCT SST	SLSTR SST product (N2, N3, D2, D3) where the best at anytime constitute the user SST. Not distributed to users.
WST SST	SLSTR SST product distributed to users
WV	Water Vapour

2. Match-Up data

The collocation of the Sentinel SLSTR SST data with in situ data, i.e. reference data, include relevant data from both the in situ data stream and from the satellite data stream (see below). Other data sets are applied where collocation criterias are met. The ancillary data sets are other satellite SST data sets, data from Numerical Weather Prediction models and ice concentration data.
This section describes the reference data sets, the Sentinel SST product to be evaluated and the ancillary data sets.

2.1. In situ data

CMEMS (operated by Coriolis) and IQUAM datasets were both considered as good candidates for this analysis. CMEMS was preferred due to Copernicus links. The CMEMS data applied are drifters, tropical moored buoys, and Argo floats.
Beside the CMEMS data set, radiometric data from the DMI-ISAR are applied for a dedicated skin-to-skin validation analysis for High Latitudes.

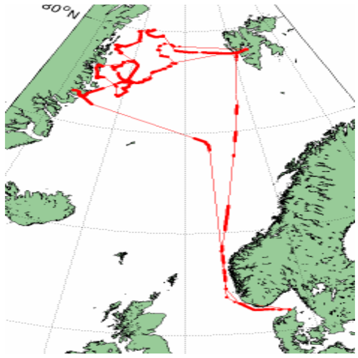

Illustration 1: Track of RV Dana, carrying DMI-ISAR during 6 weeks cruise in Greenland Sea and North Atlantic, mid August to end of September.

2.1.1. CMEMS

For all CMEMS in situ data, the five first valid values closest to surface are provided (fill values are used when there are no profile data). The in situ variables provided are:

- lat, lon, depth and time of all measurements
- platform_id, the WMO identifier of the platform
- water_temperature,
- pressure, the pressure depth in decibar for Argo floats
- solar_zenith_angle, in degrees
- climatology_water_temperature, the climatology SST value from Casey climatology
- closest_to_surface, indice of the closest measurement to surface in the variables for each in situ profile
- quality_level, an integrated quality flag ranging from 1 to 5 and combining all Coriolis flags, MetOffice and Meteo-France blacklists and additional QC tests into a single simpler flag. Failed QC tests are reported in a rejection flag.
- Water Vapour.

2.1.2. DMI ISAR

The ISAR radiometer (ISAR) is catagorized as a fiducial reference measurement instrument (FRM) that is able to provide accurate skin SST measurements with an accuracy of 0.1 K, and can be referenced to traceable standards. FRM measurements at high latitudes are extremely useful, since other in situ SST measurements are very scarce in these areas, and at the same time satellite SST retrievals are particularly challenging.

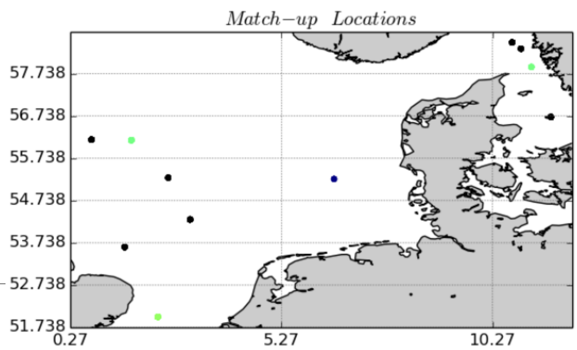

Illustration 2: Dots indicate the positions of valid DMI-ISAR data from RV Dana, during cruises in the North Sea and Skagerak, summer 2017.

The original plan was to deploy DMI-ISAR on Irena Arctica from Royal Arctic Line, during the early stage of Sentinal-3 operations. Delays of Sentinel-3 launch, logistical and technical issues with the ISAR resulted in an alternative deployment plan for the instrument, namely deployment on Danish Research Vessel Dana on cruises in the North Sea, Skagerak, and the Greenland Sea and later on Passenger Ship Norröna, operating between Denmark and Iceland. Tracks and positions of the DMI-ISAR, on-board Dana in the North Sea, Skagerak and in Greenland Sea, are illustrated in Illustration 1 and Illustration 2. The positions of Norröna in the North Sea are not shown.

The full DMI-ISAR record collected for this validation project consist of following cruises in 2017 :

- Skagerak and Kattegat; June 23 to August 15 on Dana.
- Greenland Sea and North Atlantic; August 17 to October First on Dana
- North Atlantic ; December 2017 on Norröna.

NetCDF files for the ISAR data records are produced for North Atlantic, Skagerak and Kattegat cruises. The ISAR GPS failed during most of the East Greenland cruise and the data record is matched up with ship position records. All additional ISAR information like observation uncertaintyis therefore not available for that perod.

2.2. Sentinel-3 SLSTR SST

The first Copernicus Sentinel-3 satellite, Sentinel-3A was launched on 16th February 2016 from Plesetsk, Russia. The mission is to provide a consistent, long-term collection of marine and land data for operational ocean analysis, forecasting and service provision (Donlon et al, 2012). EUMETSAT processes Sentinel-3 marine data and products at its Sentinel-3 Marine Centre, for real time delivery to end-users (Bonekamp et al, 2016). Operational Sea Surface Temperature (SST) products based on measurements from the Sea and Land Surface Temperature Radiometer (SLSTR) (Coppo et al, 2013) on board Sentinel-3A are processed and delivered by EUMETSAT within the marine centre. The dual-view design and calibration characteristics of SLSTR aim for highly accurate surface temperature retrievals, with an increased swath width and global coverage over its predecessor (ENVISAT-AATSR).

The first SLSTR level 1 (L1b) data was released to expert and validation users on the 14th June 2016, with the level 2 (L2) data released shortly after on the 21st June 2016. A successful commissioning review was held on the 12th July 2016, and the SLSTR L2 SST products were released publically on an operational basis from 5th July 2017 onwards. Routine operations of Sentinel-3A were confirmed from 16th October 2017.

SST from SLSTR provides increased global coverage than AATSR due to an increased swath width (up to 1400km) for both nadir only and dual (740km) view scans. An example of the daily SST global coverage from Sentinel-3A for one day is shown in Illustration 3.

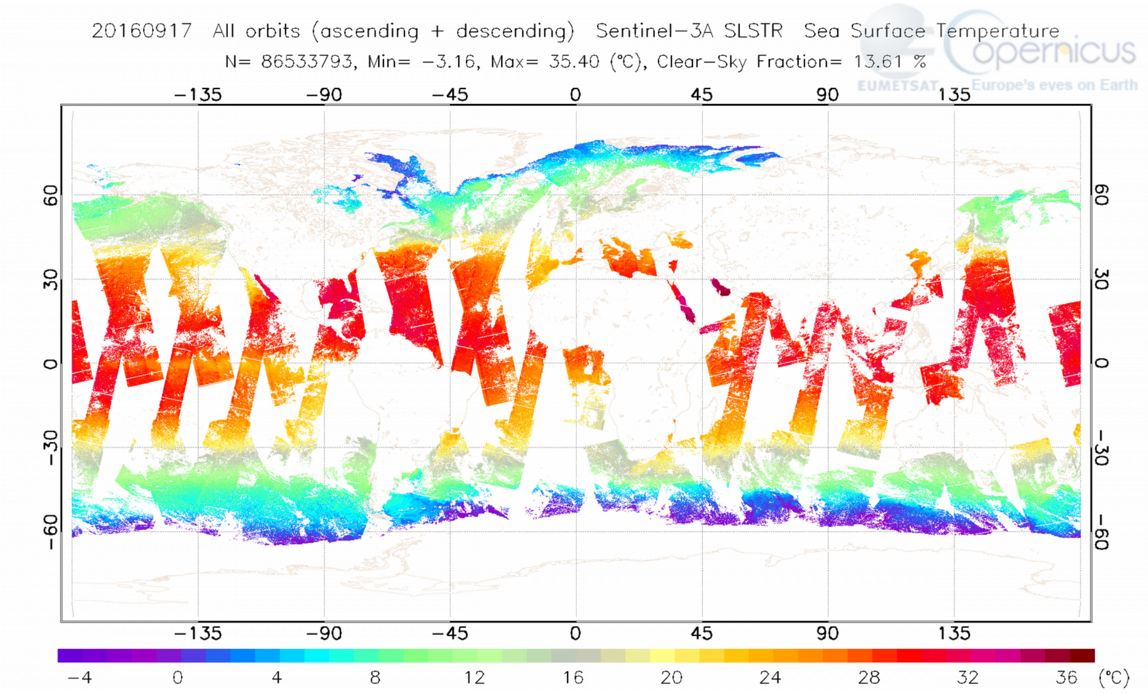

Illustration 3: Global map of Sentinel-3A SLSTR Sea Surface Temperature (day and night-time) for 17th September 2016.

The SLSTR SST level 2 product (WST) follows the Group for High Resolution SST (GHRSST) Data Specification (GDS2r5) L2P, in NetCDF4 format. It is a L2 swath product containing a preferred choice algorithm skin SST. The product also contains auxiliary data including ECMWF wind-speed, sea-ice fraction and background SST (*SLSTR PDFS*). The product is designed to contain uncertainty estimates of both Sensor Specific Error Statistics (SSES), following GHRSST specifications, and pixel-level theoretical uncertainty (*SLSTR ATBD*) which are provided as experimental fields. The quality levels are currently based on the values of theoretical uncertainty.
The SST retrieval (*SLSTR ATBD*) is performed for five different SST algorithms (*O'Carroll et al, 2015*). The preferred choice algorithm for WST is based on the time of day, view and aerosol conditions. These are referred to as a dual-view 3-channel retrieval (D3), a dual-view 2 channel retrieval (D2), a nadir-view only 3-channel retrieval (N3), a nadir-view only 2-channel retrieval (N2) and an aerosol robust nadir-view only 3-channel retrieval (N3R). The 3-channel retrieval utilise the SLSTR channels S7, S8, S9 (*Donlon et al, 2012*); and the 2-channel retrievals utilise S8 and S9. The SST is derived from a weighted combination of brightness temperatures measured in both the nadir and oblique (when available) views by the thermal channels. The coefficients are weighted as functions of viewing geometry and water vapour loading. The offset adjustments in the algorithms are designed to be tuned following validation assessments to a common reference algorithm.

Pixel-level theoretical uncertainty is estimated from a combination of measurement noise to retrieved SST; uncertainty from water vapour loading; and uncertainty from proximity to land and cloud. The values are derived separately for each SLSTR SST retrieval type (D3, D2, N3, N2, N3R). Further information can be found in the SLSTR Level-2 ATBD (*SLSTR ATBD*).

3. Methodology

All analysis of the Sentinal-3 SLSTR SST performance are based on stratified and filtered data of SLSTR SST MDB.

The validation work consists of following measures and procedures:

- Standard and robust statistics (bias, std). Nearest pixel only is applied for selected analysis (see results)
- Data screening against climatology is done. Match Up data with higher difference than 5K to climatilogy will be rejected for the PB 2.13 data sets.
- There is no corretion of buoy temperatures to skin temperature – neither for day nor night data. Issues regarding diffeerences between observed water temperatures versus satellite based skin temperature measurements are known and standard measures are often taken for night time measurements (Fairall et al., 1996).
- Sampling issues related to diurnal warming are not delt with comprehensively, but minimized through short term match-up constraints

The MUDB consists of daily files in netcdf4 format. Data are acquired through Jean Francois Piolle (IFREMER).

3.1. Match-Up Method

The collocation between SLSTR SST and in situ data are constraint by a maximum distance of 5 km and a maximum time difference of 2 hours. The satellite data consist of one square box of pixels centered on the matching pixel (the nearest cloud free pixel in the box at a maximum distance of 5 km). Because there are different interpretations on what should be the closest «valid» pixel, this choice is left to the user and only the boxes are provided.

The box size are 21 pixels for SLSTR products on the 1 km grid (similar to the OSI SAF AVHRR MDBs) and 11 pixels for those on the 2 km grid.

The distance and time difference between the closest satellite pixel and the in situ position are included in the MDB, to allow further refinement in the analysis. The match-ups also provide the history of the in situ measurements within ± 6 hours.

The area covered by Low and Mid latitudes is between -50S to 50N, however the 50N-60N band is processed also, to allow comparisons between LML and HL processing. SST High Latitude covers the area North of 60 N and South of 50 S.

Duplicate matchups, e.g. with a satellite pixel box containing more than one in situ collocation, may occur. The closest in time will be chosen and, if needed, the closest in distance. In the (unlikely) case where two in situ devices are collocated with the same satellite pixel box, both match-ups are kept.

3.2. SST validation data and procedures for mid and low latitudes

About 8 months of MDB files, produced from Near Real Time (NRT) and reprocessed data have been used in the Low and Mid-Latitude (LML) validation (Table 1). The NRT processing was not satisfactory before May 5, 2017, especially in terms of cloud clearing. So the NRT data have been used only after this date, corresponding to two versions of the Processing Baseline (PB): 2.13 and 2.18. SLSTR products, from November to March 2017, have been reprocessed with PB 2.13, then MDB files have produced.
Data set 1 and 3 (see Table 1), both obtained with PB 2.13, have shown consistent results and they have been merged into a so-called set 4, which has been widely used in section 5.1.

	version	processing mode	start date- end date	duration
Set 1	PB 2.13	NRT	5 May – 4 Jul 2017 (1)	2 months
Set 2	PB 2.18	NRT	25 Jul – 24 Aug 2017 (2)	1 month
Set 3	PB 2.13	reprocessed	1 Nov 2016 – 31 Mar 2017	5 months

Table 1: SLSTR data used in the LML validation. (1) The SLSTR annual black body cross over test was performed on 13-14 June 2017, these 2 days should have been excluded (Gary Corlett personal communication, 5 October 2017). (2) PB 2.18 started on 5 July, but a 0.5K cut-off to OSTIA was applied until 24 July (Igor Tomazic, personal communication). The cut-off period has been eliminated.

The MDB data have processed according to the following principles:

a) SLSTR skin SST is compared to in situ bulk SST, without any correction,
b) nighttime and daytime data are considered separately,
c) the SLSTR SST at the box center is compared to the in situ measurement,
d) the time difference between SLSTR and in situ SST is less than 1 hour, this delay is extended to 3 hours for argo floats,
e) the in situ measurement quality level is equal to 4 or 5,
f) the SST difference between in situ and climatology is lower than 5K; this test uses the MDB climatology value, which is an interpolation of Casey pentad climatology (Casey, 1999),
g) the SLSTR box center is cloud free , i.e. WST quality level is higher than 1,
h) the SLSTR quality level (QL) takes the following values:
   - PB 2.18 nighttime, QL equal to 5. Daytime, QL equal to 4 or 5
   - PB 2.13 nighttime, QL equal to 5. Daytime, QL equal to 3, 4 or 5

Test f) has successfully detected a few anomalous in situ measurements, but this was exceptional. Concerning test h), the Product Notice issued on 05/07/2017 has recommended the QL values to be used with PB 2.18. No recommendation was given for PB 2.13, so the selected QL values have been chosen empirically from statistics per QL. For daytime, the cases with QL equal to 2 gather correct SST and obviously wrong ones, so we chose to eliminate these cases. But they correspond to 50% of the total daytime population. As a result the number of cases with PB.13 will be much smaller for daytime than for nighttime

Illustration 4 shows the geographical distribution of the three types of in situ measurement (drifters, moored buoys and argo floats). The maps show the measurement locations that correspond to the matching and selection criteria a) to h). The maps, which concern nighttime cases over 2 months, are representative of the whole data set. The drifters have a rather good distribution, as expected. The moored buoys do not include the tropical buoys and contain many coastal buoys, not recommended for satellite validation, so the moored buoys have not been used in the LML validation. The argo floats have an acceptable distribution, if the time difference with SLSTR is extended to 3 hours, but with a very low number of matches (compared to the drifters). The argo floats have have been used only to calculate some nighttime statistics.

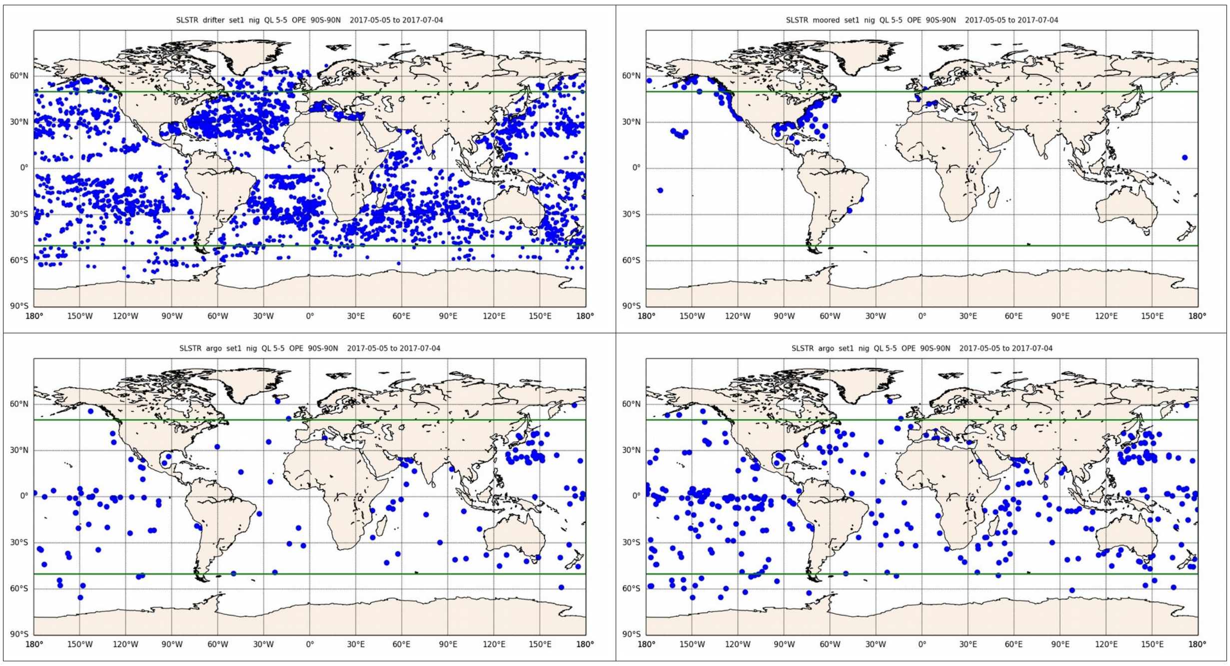

Illustration 4 Locations of the matches, nighttime, over 2 months, for drifters (top left), moored buoys (top right), argo floats with |dt| < 1h (bottom left) and argo floats with |dt| < 3h (bottom light), dt being the time difference between SLSTR and in situ SST.

3.3. SST validation data and procedures for high latitudes

The dedicated High Latitude SLSTR SST validation is mainly performed for North Atlantic ocean, the Greenland Sea and Barants Sea. The validation is further stratified into validation against traditional water temperature measurements from the Coriolic DB, and radiometric measurements from the DMI-ISAR radiometer.

3.3.1. Validation against drifting Buoy measurements

The validation of the high latitude (HL) areas (poleward of 50S and 60N) using drifting buoys has been based on the same match-up dataset as the LML validation, except some additional filtering and extended period for 2.18 as described below. The description above therefore also applies for the HL validation, to a large extent.

The MDB data used for HL validation have been processed according to the principles described in a) – h) in the list above. In addition, the following principles have been applied:

i) all matchups with sea ice concentration > 10% are excluded,
j) all matchups with SSTinsitu < -1.8C are excluded,
k) all matchups between 13th and 14th June 2017 are excluded.

Test i) and test j) are used to remove in situ observations done by drifting buoys on/in the sea ice. The MDB do not distinguish between in situ observations in the sea and observations on the ice. Test k) is due to the note in Table 1, which produced several very wrong match-ups.

For the validation of data from PB 2.18 of the SLSTR SST product, the period for high latitude validation has been extended compared to the LML validation period listed in Table 1. This is due to the few drifters available at high latitudes. The period used for PB 2.18 validation is 25th July to 31st October 2017, with a few holes in the period.

Illustration 14 and Illustration 15 show the geographical distribution of the in situ observations at high latitudes.

3.3.2. Validation against radiometer measurements

The validation of the high latitude (HL) areas (poleward of 50S and 60N) against DMI-ISAR radiometer is based on the reference dataset described in 2.1.2. This validation is based on a much smaller data set than validations against Coriolis data, due the fact that only one observation platform is applied.

The MDB data set used for HL validation against DMI-ISAR has been processed slightly differently from the Coriolis based MDB. Here the temporal and spatial constraints are 1 hour and 1 km respectively, and the data from the Greenland Sea/Fram Strait are validated seperately from the data from the North Atlantic, North Sea and Skagerak. The reason for this is to separate pure «CLEAN» SST data from MIZ SST in East Greenland. Here special challenges are present, like steep ocean temperature gradients, ice infested pixels and clouds hampering the SST performance.

The errors between SLSTR SST and radiometric in situ SST is expected to be smaller than to buoy SST, because skin temperature estimates from satellite is directly comparable with in situ radiometer skin measurements, i.e. a skin-to-skin temperature intercomparison.

4. Results

The validation results in this chapter is split in three parts: validation of low/mid latitudes SST against drifting buoys, high latitude validation against drifting buoys and high latitude validation against in situ radiometer SST. In the comparison with the drifting buoys, the SLSTR SST skin temperature products are compared directly with the drifting buoy bulk temperature, without any skin-to-bulk temperature correction. Therefore some of the observed bias can be explained by skin-bulk differences. In general, the skin is expected to be slightly cooler than the bulk during night time (about 0.2C in difference on average) due to surface cooling. During daytime in calm and clear conditions, the skin can be warmer than the bulk due to diurnal warming, with different amplitude depending on wind speed, solar heating, temperature stratification etc.

4.1. SLSTR SST vs Coriolis SST at low and mid latitudes

The SLSTR SST product studied in sections 4.1.1 to 4.1.4, is the product provided to users, i.e. the WST SST, where the word WST is often omitted. The SST difference or errors refer to (SST_{SLSTR} – SST_{insitu}). All results have been obtained at low and mid-latitudes (50N to 50S) and, for most of them, on data set 4, which includes 7 months of data.

4.1.1. Cloud screening

The cloud screening has been problematic since the beginning of SLSTR data. The processing has been improved, but the PB 2.13 SST data are not yet satisfactory. The SLSTR SST versus in situ SST plot (Illustration 5) do show erroneous pixels, where SLSTR SST is much colder than in situ SST. These pixels are actually clouds and they are more frequent for nighttime (left) than for daytime (right). The few erroneous hot pixels occurred on 13-14 June 2017, they are not physically significant. The cloudy pixels do not correspond to a specific geographical region but are distributed all over the ocean (Illustration 7).
To cope with the cloud screening problem, a filter has been applied, keeping only the data having a SST difference between SLSTR and climatology lower than 5K (Illustration 6). A similar filter is included in PB 2.18 but using the SST difference between SLSTR and OSTIA analysis. Indeed such filters reduce the number of cloudy pixels but do not eliminate them completely.

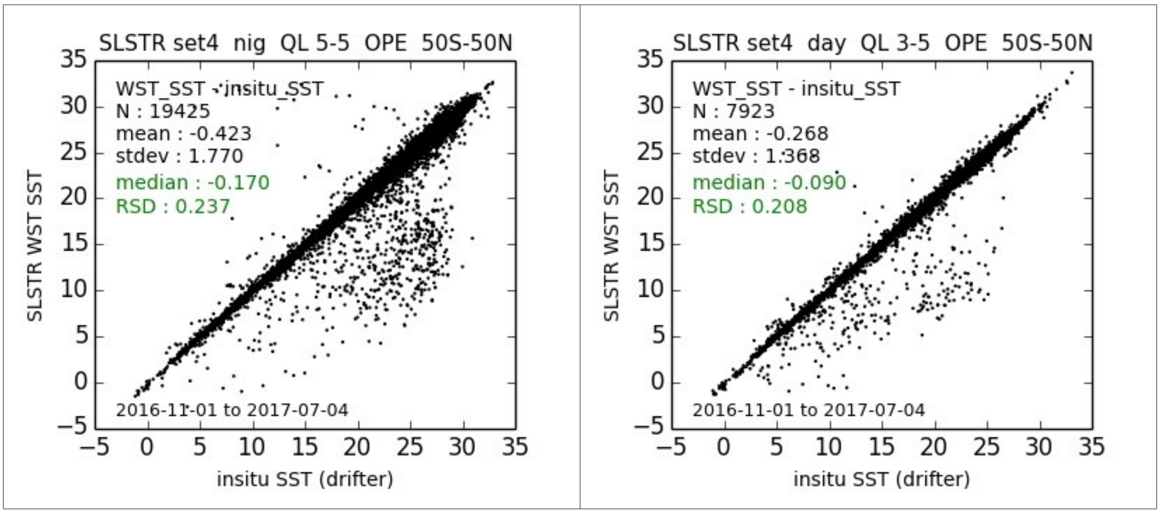

Illustration 5: SLSTR SST versus in situ SST, for nighttime (left) and daytime (right), on set 4 (8-months),

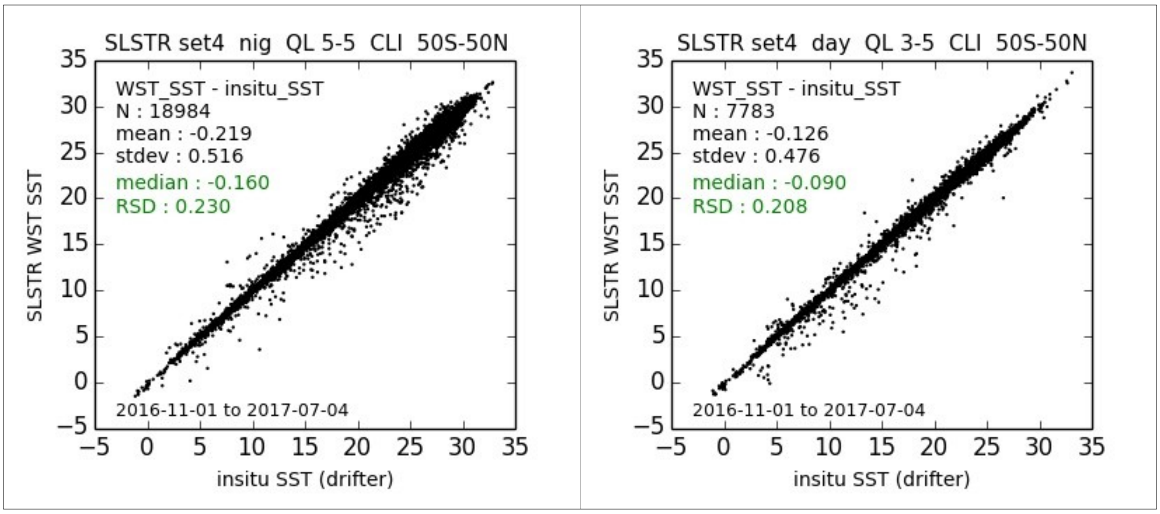

Illustration 6: Same as Illustration 5 with an additional filter $|SST_{SLSTR} - SST_{clim}| < 5K$

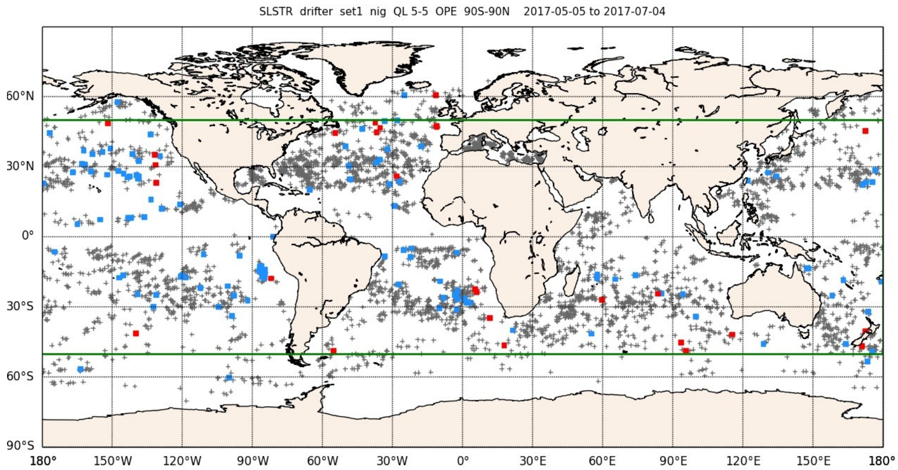

Illustration 7: Nighttime erroneous pixels. Cold pixels ($SST_{SLSTR} - SST_{clim}|$)< -5K are shown in blue and hot pixels ($SST_{SLSTR} - SST_{clim}$) < 5K in red.

4.1.2. SLSTR SST validation statistics

The WST SST validation statistic have been calculated mainly against drifters, in the area 50N-50S. The results are presented in Table 2 for operational versions 2.13 and 2.18, and for version 2.13+clim (2.13 with filter $|SST_{SLSTR} - SST_{clim}| < 5K$). Nighttime results with version 2.13+clim are also presented for argo floats, they are rather close to the corresponding drifter results.

The classical statistics of the operational versions represent the performances seen by the users:

- Version 2.13, operational in May-June 2017, has a mean difference of -0.42K and a standard deviation of 1.77K for nighttime, the corresponding daytime values being -0.27K and 1.37K. These bad results are due to the cloud contamination presented in 4.1.
- Version 2.18, operational since July 2017, has a mean difference of -0.25 K and a standard deviation of 0.51 K for nighttime, the corresponding daytime values being -0.12K and 0.48K. These better results are due to the 5K cut-off to OSTIA, included in PB 2.18 .

Robust statistics have been calculated systematically, as they are less sensitive to the erroneous cloudy pixels and better representative of the SST algorithm performances (Merchant, 1999). The nighttime median values, from -0.15 to -0.19, are consistent, as we compare SLSTR skin SST to buoy bulk SST; for nighttime, the skin minus bulk SST difference is -0.2K in average. The robust statistics of set 4 with 2.13+clim are probably the best summary for the SST algorithm performances:

night median = -0.16K RSD =0.23K day median = -0.09K RSD =0.21K

	data	version	mean	stdev	median	RSD	nb cases	SST
night	set 4	2.13	-0.42	1.77	-0.17	0.24	19425	22.94
night	set 4	2.13+clim	-0.22	0.52	-0.16	0.23	18984	22.95
night	set 4 A	2.13+clim	-0.19	0.47	-0.15	0.24	1483	24.06
night	set 2	2.18	-0.25	0.51	-0.19	0.24	1752	24.78
day	set 4	2.13	-0.27	1.37	-0.09	0.21	7923	19.23
day	set 4	2.13+clim	-0.13	0.48	-0.09	0.21	7783	19.27
day	set 2	2.18	-0.12	0.48	-0.08	0.25	1328	23.04

Table 2: SLSTR SST validation statistics , 50N-50S, for several processing versions. All results concern drifters except "set 4 A", which concerns argo floats. The statistics are calculated on the difference between SLSTR SST and drifter SST. "mean" is the mean, "stdev" the standard deviation, "median' the median and "RSD" the Robust Standard Deviation. "nb cases" is the number of cases and "SST" the mean in situ SST.

4.1.3. SST difference dependencies

The nighttime and daytime SST differences, calculated with 2.13+clim, are displayed as a function of the satellite zenith angle (Illustration 8) and as the function of the SST (Illustration 9). Because of the remaining cloudy pixels, the "binned" curves are the median and robust standard deviation (instead of the mean and standard, as usual). No significant dependency is observed.

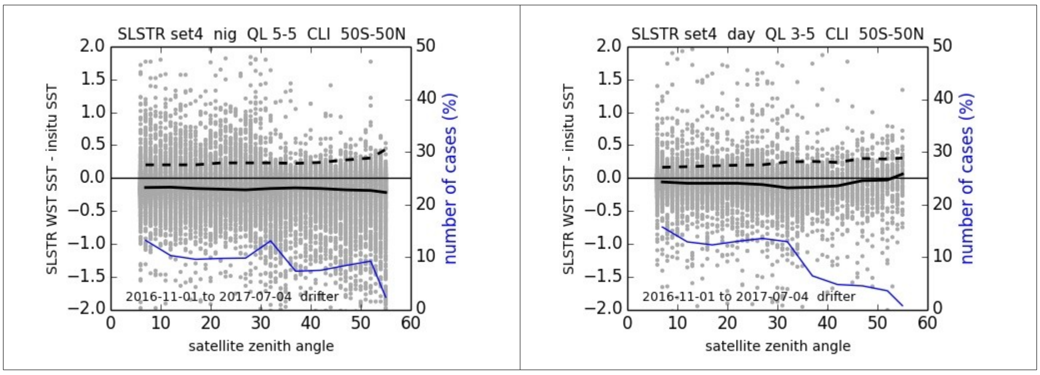

Illustration 8: Differences between WST SST and drifter SST as a function the satellite zenith angle, fo*r nighttime (left) and daytime (right). Each plot shows the individual cases (grey dots), the median (solid), the robust standard deviation (dash) and the number of cases (thin blue).*

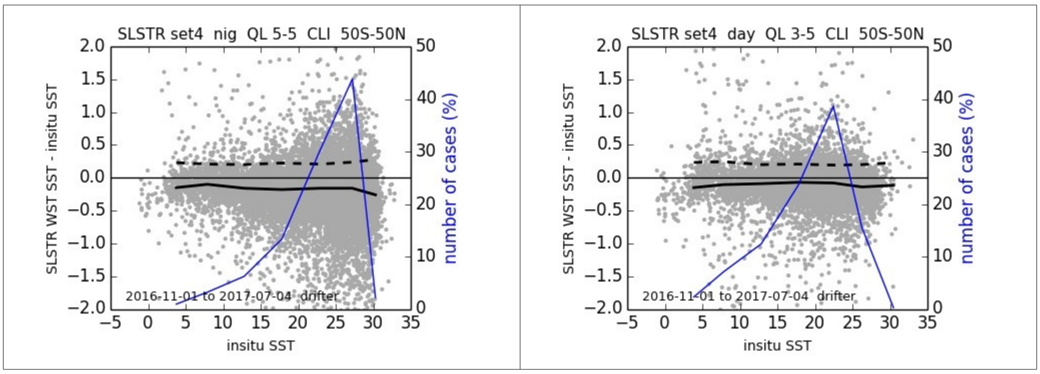

Illustration 9: similar to figure Illustration 8 but with in situ SST on x-axis.

4.1.4. SST difference maps

Using version 2.13+clim, the SST difference median has been calculated on a regular 5-degree grid over 7 months, separately for nighttime (Illustration 10) and daytime (Illustration 11). No significant regional bias is observed. The daytime coverage is lower than the nighttime one, because of the elimination of the cases QL=2, as explained in section 3.2.

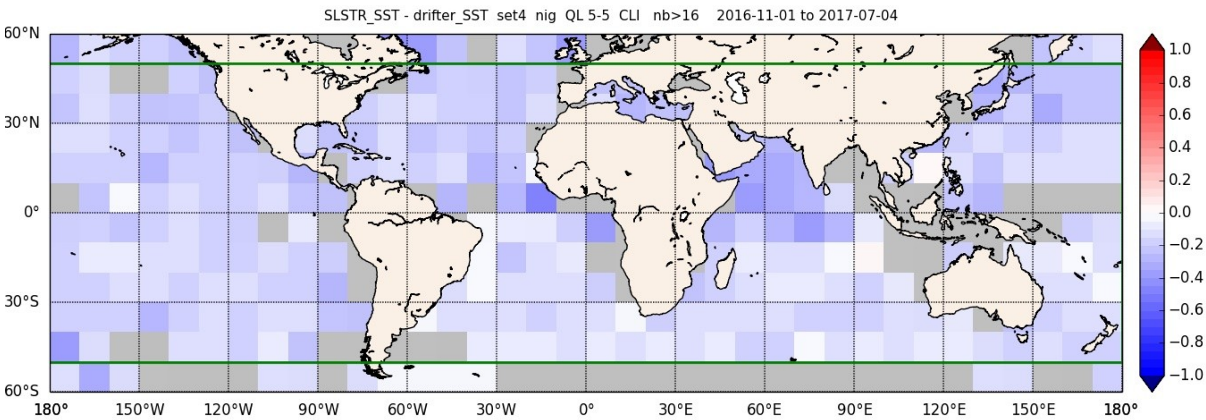

Illustration 10: Median of ($SST_{SLSTR} - SST_{insitu}$) *nighttime., on set 4 (8-months). Version 2.13+clim is used. A minimum number of 16 cases is required in each grid box.*

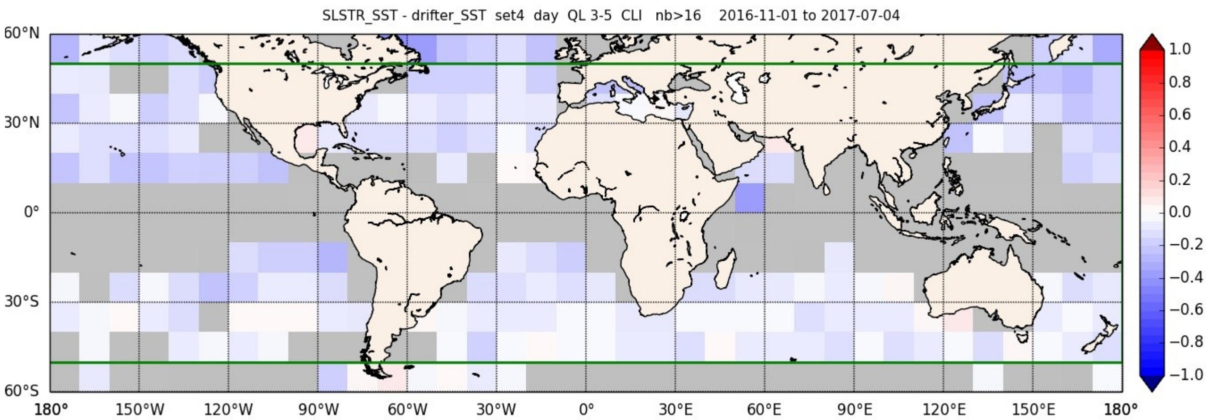

Illustration 11: Similar to Illustration 10 but for daytime.

4.1.5. Single and dual view SST

The single and dual view SST, called WCT SST, are produced, but not disseminated:

- D3, 3 channels and dual view (across and along track)
- N3, 3 channels and nadir single view
- D2, 2 channels and dual view (across and along track)
- N2, 2 channels and nadir single view

The 3 channels are (3.74, 10.85 and 12 µm), the 2 channels (10.85 and 12 µm).
The nadir view is available on the full swath (about 1400 km), the dual view on a limited part of the swath (about 700 km).

The WST SST product combines the four WCT SST products, choosing for every pixel with the following rules:

- 3-channel algorithms are used for nighttime, 2-channel algorithms for daytime,
- if available, the dual view is preferred to the single view

The WST SST product includes a variable giving the chosen algorithm for every pixel.

A detailed comparison of the WCT SST performances is not simple. The 4 WCT SST should be compared on a same data set. Regional data sets would be needed, as dual view performs better with high water vapor content, typically in the tropics). And the cloud screening may differ between the WCT SST.

We have made a user oriented comparison, simply considering the WCT values that have contributed to the WST SST. The filter $|SST_{SLSTR} - SST_{insitu}| < 5K$) has been applied. The statistics are presented in Table 3. Because of the remaining cloudy pixels, the robust statistics better represent the SST algorithm performances. The dual view algorithm performs better than its corresponding nadir view algorithm, as expected. For daytime, the RSD is significantly lower for D2 (0.22) than for N2 (0.27). For nighttime, the RSD is slightly lower for D3 (0.23) than for N3 (0.24).

	algorithm	mean	stdev	median	RSD	nb cases
night	D3	-0.13	0.45	-0.15	0.23	10147
night	N3	-0.30	0.62	-0.16	0.24	8836
day	D2	-0.11	0.46	-0.08	0.22	5361
day	N2	-0.17	0.52	-0.12	0.27	2421

Table 3: Single and dual view SST statistics against drifters, on set 4, 50N-50S,s. Each line shows the statistics calculated on the difference (WCT SST - drifter SST.), where WCT corresponds to the algorithm in column 2. “mean” is the mean, “stdev” the standard deviation, “median' the median and “RSD” the Robust Standard Deviation. “nb cases” is the number of cases.

4.2. SLSTR SST vs Coriolis SST at high latitudes

For high latitudes the validation of SLSTR SST has also been performed using drifting buoys from Coriolis. As for low and mid latitudes, the SLSTR SST investigated for HL is the product distributed to users, the WST SST. The SST difference refer to ($SST_{SLSTR} - SST_{insitu}$). All results have been obtained for southern and northern hemisphere separately, to illustrate possible differences at the two hemispheres.

4.2.1. Cloud and ice screening

The cloud screening has also been a problem at high latitudes, as discussed in section 4.1.1 for low/mid latitudes. The scatter plots in Illustration 12 show that there are a few match-ups with significant cold bias, typical for undetected clouds, even using the suggested quality flags only. This results in an overall cold bias and high standard deviation. There seems to be more undetected clouds on southern hemisphere than on northern hemisphere, but since the validation period is not a full year, this might be caused by seasonal differences.

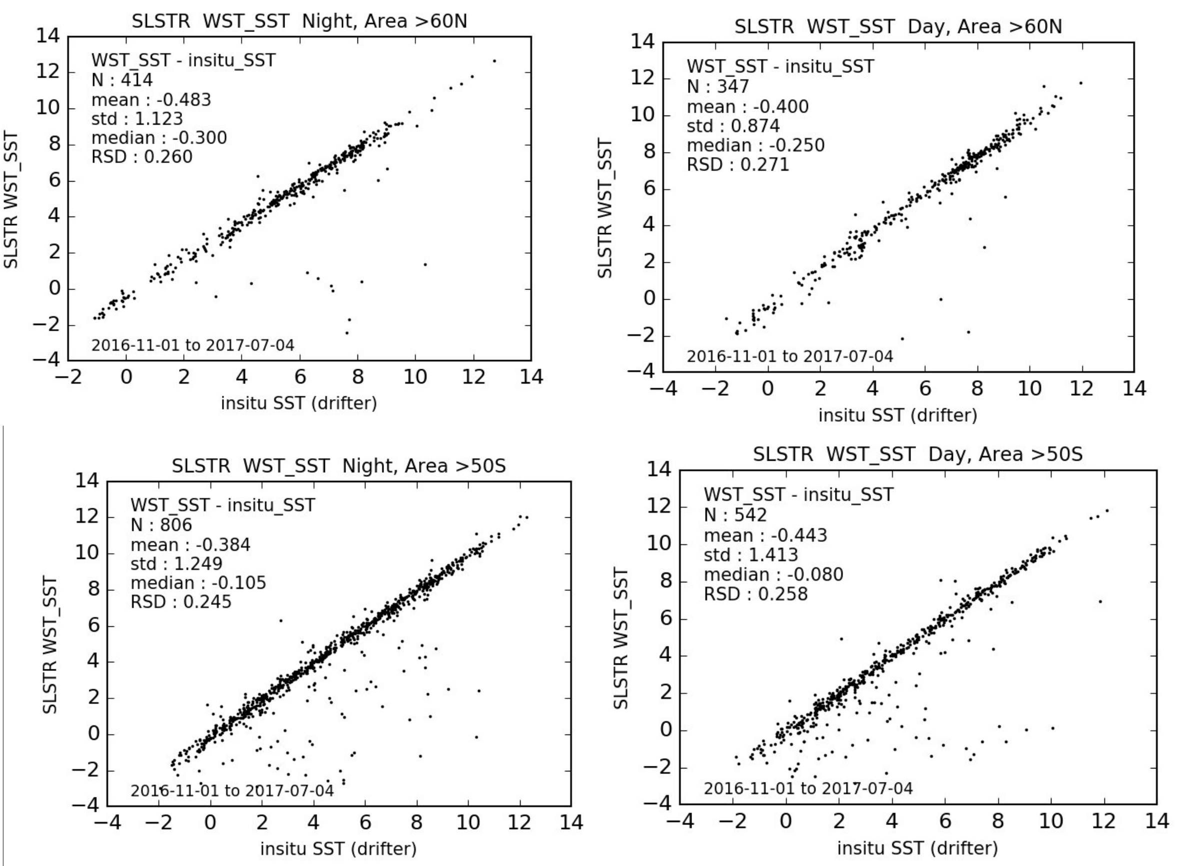

Illustration 12: SLSTR SST versus in situ SST, for nighttime (left) and daytime (right), northern hemisphere (upper) and southern hemisphere (lower), on set 4 (8 months).

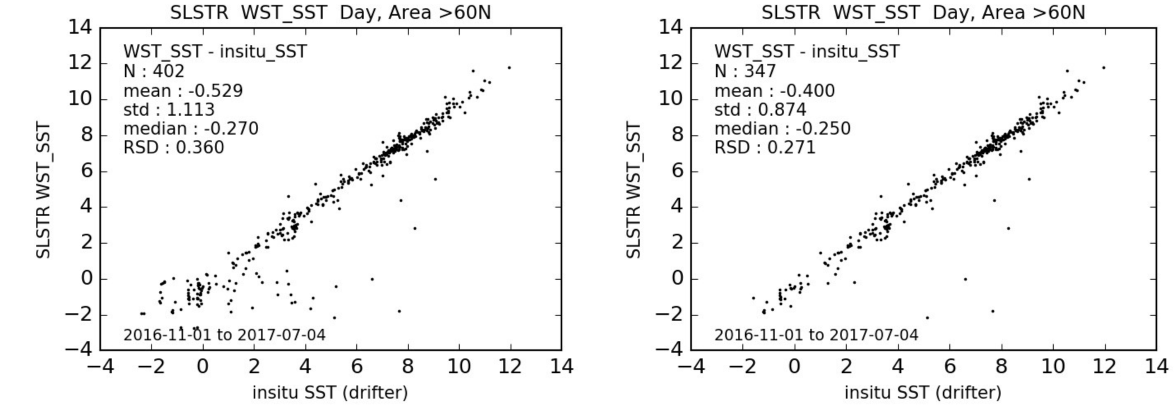

Illustration 13: SLSTR SST versus in situ SST for daytime on set 4, without (left) and with filtering (right) for sea ice concentration > 10%.

At high latitudes we also need to screen out areas with sea ice concentration when retrieving SST. This should be done similar as for screening out cloudy areas. As for cloud screening, the SLSTR data also have problems with sea ice screening. In Illustration 13 plots without and with filtering the MDB with sea ice concentration (SIC) > 10% are shown. The SIC data used is the OSI SAF sea ice concentration product. Without filtering, the MDB contains matchups with SIC up to 100%, and these matchups results in the higher variability for low temperatures in the left plot. The reason why there are matchups in the sea ice, is that drifting buoys are placed on the sea ice to report drift, temperature and pressure. Such buoys should be filtered out before validating SST, but in this case they also show that the SST product provided data with high confidence level over areas with sea ice.

4.2.2. SLSTR SST validation statistics at HL

The WST SST validation statistics for high latitudes has been calculated as for low-mid latitudes using drifting buoys. The results have been calculated for northern (NH) and southern hemisphere (SH) for both the PB 2.13 and 2.18 versions. For the 2.13 version results are also presented with an additional climatology filter ($|SST_{SLSTR} - SST_{clim}| < 5K$). The results are shown in Table 4 and Table 5.

The classical statistics that include all selected match-ups, shows the following results:

- Version 2.13 has a mean difference of -0.48K/-0.39K for nighttime and -0.40K/-0.44K for daytime, for NH and SH respectively. The standard deviations are 1.12K/1.25K for nighttime and 0.88K/1.42K for daytime. These quite bad results are due to cloud contamination, as discussed in 4.2.1. Most of the sea ice has been filtered out in this validation.
- Version 2.18 has a mean difference of -0.22K for NH at nighttime and -0.14K/-0.12K for daytime, for NH and SH respectively. The standard deviations are 0.40K for nighttime and 0.70/0.29K for daytime. There are too few observations for SH nighttime to report on. The improved results compared to 2.13 are due to the filtering done in 2.18, using OSTIA SST and imposing a 5K cut-off.

NH	data	version	mean	stdev	median	RSD	nb cases	SST
night	set 4	2.13	-0.48	1.12	-0.30	0.26	413	4.96
night	set 4	2.13+clim	-0.35	0.50	-0.29	0.25	406	5.05
night	set 2	2.18	-0.22	0.40	-0.19	0.18	123	8.64
day	set 4	2.13	-0.40	0.88	-0.25	0.26	344	5.53
day	set 4	2.13+clim	-0.33	0.53	-0.25	0.26	341	5.59
day	set 2	2.18	-0.14	0.70	-0.16	0.24	330	6.91

Table 4: SLSTR SST validation statistics for high latitudes at Northern Hemisphere (north of 60N), for several processing versions using drifting buoys only. The statistics are calculated on the difference between SLSTR SST and drifter SST. “mean” is the mean, “stdev” the standard deviation of errors, “median' the median and “RSD” the Robust Standard Deviation. “nb cases” is the number of cases and “SST” the mean in situ SST.

SH	data	version	mean	stdev	median	RSD	nb cases	SST
night	set 4	2.13	-0.39	1.25	-0.10	0.25	800	4.29
night	set 4	2.13+clim	-0.26	0.80	-0.10	0.24	785	4.39
night	set 2	2.18	-	-	-	-	*18*	-
day	set 4	2.13	-0.44	1.42	-0.08	0.26	538	3.72
day	set 4	2.13+clim	-0.26	0.84	-0.07	0.24	524	3.84
day	set 2	2.18	-0.12	0.29	-0.10	0.22	105	3.62

Table 5: SLSTR SST validation statistics for high latitudes at Southern Hemisphere (south of 50S), for several processing versions using drifting buoys only. The statistics are calculated on the difference between SLSTR SST and drifter SST. "mean" is the mean, "stdev" the standard deviation, "median' the median and "RSD" the Robust Standard Deviation. "nb cases" is the number of cases and "SST" the mean in situ SST.

The robust statistics have been calculated to get an impression of the algorithm performance, as these statistics are less sensitive to undetected clouds and sea ice. The median values shows bias between -0.30 and -0.07, with a bit colder bias at night compared to day, and colder bias at NH compared to SH. The robust standard deviation is similar for NH and SH, between 0.22 and 0.26. The night time set 2 validation for SH has too few match-up's to be evaluated (only 18).
The median values show a cold bias which is expected, due to the difference between satellite skin and in situ bulk temperature. Globally this difference is on average -0.17C at nighttime, and warmer on daytime, depending on the local diurnal warming. For this case, the nighttime median is –0.10C for SH and between -0.18C and -0.25C for NH. For daytime the median is a bit warmer, as can be expected due to daytime surface warming.

4.2.3. SST difference maps

To study the geographical distribution of drifting buoys and the distribution of differences, maps with color coded markers have been plotted for the version 2.13+clim. Such plots are shown in Illustration 14 and Illustration 15. At the northern hemisphere, the points with negative bias (blue) are found in all regions, which indicates that no regional biases are found within this areas for this data period. For the southern hemisphere, there seems to be some areas that are colder than others, but not consistent between day and night. So it is difficult to conclude on the cause of this.

4.2.4. Single and dual view SSTs

To compare the performance of single and dual view algorithms, using 2 or 3 channels, the WCT SST fields have been studied as described in section 4.1.5 for low/mid latitudes. The results are presented in Table 6 and Table 7 for NH and SH high latitudes. Using the robust statistic at high latitudes, the results are not conclusive regarding dual view being better than nadir view. Actually, nadir view is better than dual view for SH day and night, and NH day. At SH, using only two channels (D2) at night performs better than using three channels (D3). But at NH this is opposite, as would be expected. Some of these findings are opposite when the normal statistics are used. This makes it difficult to draw any firm conclusions concerning the preference of dual view or nadir view, two or three channels, at high latitudes. This could be caused by a quite limited data set and relatively few matchups, compared with the low/mid latitude results.

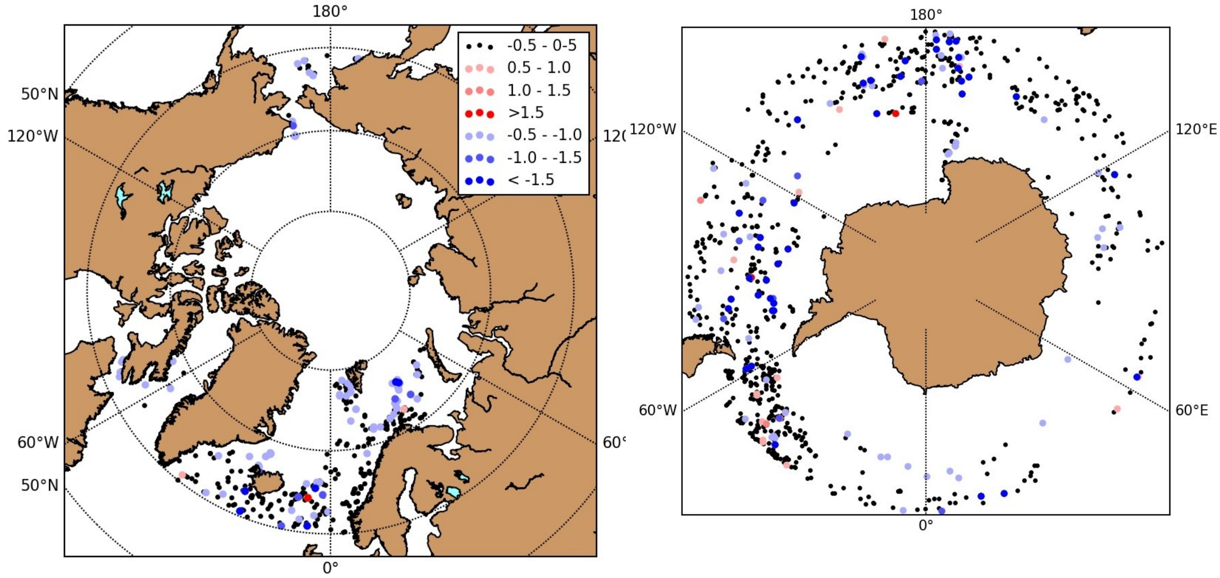

Illustration 14: SST_{SLSTR} $SST_{in\ situ}$ at nighttime, for v2.13 with climatology filter, on set 4.

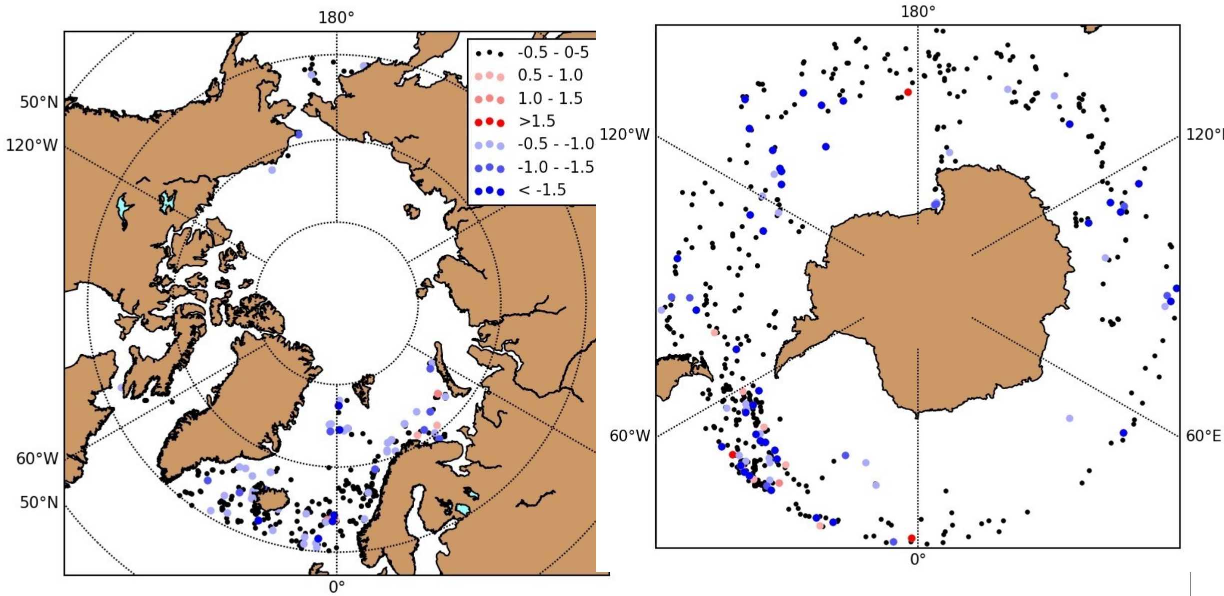

Illustration 15: As Illustration 14, but for daytime.

NH	algorithm	mean	stdev	median	RSD	nb cases
night	D2	-0.36	0.51	-0.33	0.30	233
night	D3	-0.33	0.44	-0.30	0.25	233
night	N3	-0.42	0.51	-0.32	0.27	406
day	D2	-0.33	0.40	-0.28	0.30	209
day	N2	-0.36	0.48	-0.28	0.25	341

Table 6: Single and dual view SST statistics against drifters, on set 4, >60N. Each line shows the statistics

calculated on the difference (WCT SST - drifter SST.), where WCT corresponds to the algorithm in column 2. “mean” is the mean, “stdev” the standard deviation, “median' the median and “RSD” the Robust Standard Deviation. “nb cases” is the number of cases.

SH	algorithm	mean	stdev	median	RSD	nb cases
night	D2	-0.19	0.51	-0.16	0.22	398
night	D3	-0.15	0.52	-0.13	0.25	399
night	N3	-0.24	0.79	-0.07	0.19	784
day	D2	-0.32	0.86	-0.14	0.24	295
day	N2	-0.23	0.89	-0.00	0.22	524

Table 7: Single and dual view SST statistics against drifters, on set 4, <50S. As Table 6.

4.3. SLSTR SST vs DMI-ISAR radiometer SST at high latitudes

As mentioned above, there are only few matches between SLSTR SST and DMI-ISAR, but, the available match-up data indicate very high performance of both the SLSTR WST product as well as for the WCT SST products, outside the MIZ.

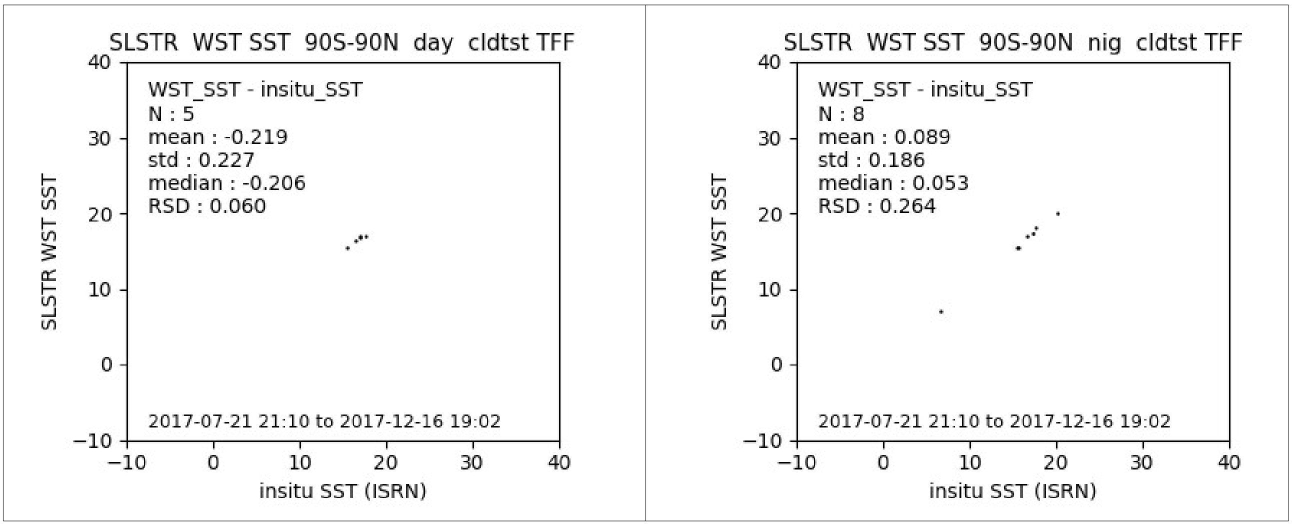

Illustration 16 Statistics for SLSTR WST high latitude validation against DMI-ISAR observations in The North Sea, North Atlantic and Skagerak (daytime left, nighttime right)

Five daytime data and eight nighttime matches are naturally too few data to perform thorough statistical analysis and in particular robust statistics, nevertheless, both statistics are presented here. The standard statistics (STD and bias) does indicate a very high performing SST product with STD around 0.2 K and bias of -0.22 during daytime and 0.09 K for nightime data. This is shown in Illustration 16 and in Table 8, where the latter also show the performance for the corresponding Nadir and Dual view algorithms.

SST algorithm	ToD	N	mean	std	median	RSD
WST	Day	5	0.219	0.227	-0.206	0.06
N2	Day	5	0.235	0.181	-0.127	0.108
D2	Day	2	0.172	0.046	-0.172	0.034
WST	Night	8	-0.089	0.186	0.053	0.264
N2	Night	8	-0.204	0.232	0.332	0.309
D2	Night	3	-0.217	0.155	0.269	0.137
N3	Night	8	-0.114	0.192	0.145	0.228
D3	Night	3	0.015	0.185	-0.111	0.156

Table 8: Full validation statistics of SLSTR WCT and WST algorithms, Algorithm (column-1), Time-Of-Day (c-2), Number of samples (c-3), Mean value of samples (c-4), Standard-Deviation of errors (c-5), Mediam value of errors (c-6) and Robust Standard Deviation (c-7).

The general picture from Table 8 is that the Dual view algorithms (D2 and D3) perform slightly better that the Nadir view algorithms (N2 and N3) and that the WST SST not necessarily performs best, as it ideally should. The validation numbers from the standard statistics and the robust statistics indicate that atmospheric contamination of the North Atlantic and North Sea data is neglectable, because the standard and robust statistics are almost equal. The results indicate very high «true» performance of the SLSTR SST performance, considering that this statistics is concerned with skin-to-skin temperature intercomparison, with minimized atmospheric interference.

The SLSTR SST performance is different in the MIZ. The MIZ in the Fram Strait and in the Greenland Sea is characterized by fronts of large ocean temperature gradients and with local and pronounced cloud formation, where the dry Arctic air meets the warm North Atlantic water. These effects are anticipated to worsen the performance of satellite SST algorithms in general. The validation statistics from ISAR observations in the Greenland Sea supports this assumption. In Table 9 the standard error statistics of the DMI-ISAR matches with SLSTR SST is presented, showing a STD of errors around 0.7 K, i.e. approximately 3 times larger than outside the MIZ. Also the bias for WST and N2 SST's are significantly larger than for the North Atlantic SST matches,

Algorithm	STD	Bias	N
WST SST	0.735	-0.630	13
N2	0.657	-0.745	13
D2	0.576	-0.085	5

Table 9: Standard validation statistics of SLSTR WCT and WST (D2 and N2) algorithms, The columns from left to right show Standard-Deviation of errors, Bias and Number of samples.

5. Conclusions

The WST and WCT SST retrival algorithms using SLSTR data have been validated against buoy and radiometer data. Concluding remarks are given below, for each validation area and in situ data type.

General conclusions to be drawn from this work are:

1) The overall performance is high and within the pre-launch performance requirements.
2) The cloud screening procedures are not working satisfactory, in particular not for the PB 2.13. data sets, but the quality of PB2.13 improves significantly if a climate filter is imposed.
3) The cloud screening for PB 2.18 is better than for the PB2.13. data sets,
4) After applying a climate filter for the PB 2.13 data, the overall performance of the WST and WCT algorithms are good, with no clear geographical dependencies and the HL performance is comparable with performance at mid and low latitudes.
5) The performance of SST retrievals from dual view sensors is generally better than algorithms using Nadir sensors only, despite the relative cource resolution of the dual view sensors.
6) There is no clear difference in performance between SH and NH high latitudes SST performance.
7) The performance of SLSTR SST retrievals against radiometric SST at high latitudes indicate that SST retrieval from SLSTR SST algorithms is of extrodinary high quality.
8) Validation results from the MIZ indicate that special measures are needed in order to improve SST performance there. This was anticipated.

Generally, and for the latter point (above) in particular, the performance of the up-coming baysian and probabilistic cloud mask screening algorithm is awaited with exitement. This cloud screening procedure will be implemented from April 2018, and a reprocessing of the full SLSTR Level 1 and 2 data sets will follow shortly after (Personal communication, O'Carrol 2018).

5.1. Low and mid Latitudes vs Corioles

SLSTR SST has been validated against drifting buoys at low and mid-latitudes (50N-50S), from November 2016 to August 2017. The validation statistics of the WST SST disseminated to users are the following:
PB 2.13 night mean = -0.42K, stdev = 1.77K; Day mean = -0.27K, stdev =1.37K
PB 2.18 night mean = -0.25K, stdev =0.51K; Day mean = -0.12K, stdev =0.48K
Version 2.13, operational in May-June 2017, shows bad results due to cloud contamination, version 2.18, operational since July 2017, shows improved results because of the 5K cut-off to OSTIA; which is included in the processing.

The cloud contamination is a major problem. It has been partly solved by applying a 5K cut-off to the climatology, which reduce the number of cloudy pixels, and by using robust statistics for most of the validation. Indeed the robust statistics are better representative of the SST algorithm performances.

The robust statistics of the difference between SLSTR SST and drifter SST, calculated in the area 50N-50S over 7 months are:
Night median = -0.16K, RSD =0.23K; Day median = -0.09K, RSD =0.21K

The nighttime median is consistent, as we compare SLSTR skin SST to buoy bulk SST

No dependencies in SST and satellite zenith angle and no regional bias have been observed.
The single and dual view SST have also been validated against drifters. The dual view algorithm performs better than its corresponding nadir view algorithm, as expected. D2 performs significantly better than N2, while. D3 performs slightly better than N3.

5.2. High Latitudes vs Coriolis

For high latitudes similar validation statistics of the WST SST is done.

For the Northern Hemisphere:
PB 2.13 night mean = -0.48K stdev =1.12K day mean = -0.40K stdev =0.88K
PB 2.18 night mean = -0.22K stdev =0.40K day mean = -0.14K stdev =0.70K

and for the Southern Hemisphere:
PB 2.13 night mean = -0.39K stdev =1.25K day mean = -0.44K stdev =1.42K
PB 2.18 night mean = NA stdev = NA day mean = -0.12K stdev =0.29K

The PB 2.13 results are not good, due to cloud masking problems. This problem has been improved with the inclusion of the 5K cut-off to OSTIA climatology, which was introduced in PB 2.18. There are also problems with the sea ice masking, and improvements are needed to improve the SLSTR SST product quality.

The robust statistics at high latitudes shows similar results as for low/mid latitudes over the 7 months period covered by PB 2.13:
NH: night median = -0.30K RSD =0.26K day median = -0.25K RSD =0.26K
SH: night median = -0.10K RSD =0.25K day median = -0.08K RSD =0.26K

The main difference is that the median at NH is lower than at SH, and robust standard deviation is a bit higher.

The difference between single view and dual view, and between two channels and three channels algorithm does not give the same results as for low/mid latitudes. Dual view do not show better results than single view, and the three channels algorithm does not show conclusively better results than the two channels algorithm. More complete MDB data and further studies are needed to conclude on this issue for high latitudes. Stratification of the validation into total column Water Vapour (TCWV) dependency most likely would reveal that the Nadir view algorithms will out-perform the Dual view algorithms under drier atmospheric conditions. In dry atmospheres the correction for water vapour is small compared to the increased uncertainty from using the coarser resolution data from the Dual view sensors. Spatial resolution of the Dual view is around 4 km whereas the spatial resolution of the Nadir view is 1 km at Nadir. This analysis is not done here.

5.3. High Latitudes vs Radiometer

The validation of SLSTR SST against in situ SST from the DMI-ISAR instrument differ from validation against buoy SST, because this is a skin to skin temperature intercomparison. Due to a very low number of samples in this analysis the results are only indicative, but none the less relatively consistent. Outside the MIZ the SST performance is extraordinary good with STD of errors of 0.1-0.2 K and bias between -0.2 to 0.2 K.

In the MIZ the errors are significantly higher, as expected, due to special challenges with large temperature gradients, mixed water and ice pixels and pronounced formation of local clouds. This is confirmed in the HL analysis against buoys in Illustration 13. The STD of errors of the SLSTR SST for water temperature around freezing point in the MIZ is 0.73 K, i.e. 3-4 times larger than the error estimate away from the MIZ.

In this analysis it is indicated that the Dual View algorithm perform better tha Nadir view algorithms, which is contradictionary to the assumption that Nadir view algorithms (N2 and N3) perform better for dry atmospheres, due to true higher spatial resolution of the Nadir sensors compared with the oblique view sensors. However, more work is needed to evaluated the algorithms dependency to WV.

6. Acknowledgement

Thanks to Technical University of Denmark for giving access to RV Dana for the deployment of the DMI-ISAR instrument and their provision of excellent logistical assistance.

7. References

Bonekamp, H., F. Montagner, V. Santacesaria, C. Nogueira Loddo, S. Wannop, I. Tomazic, A. O'Carroll, E. Kwiatkowska, R. Scharroo and H. Wilson, (2016) Core operational Sentinel-3 marine data product services as part of the Copernicus Space Component, Ocean Sci., **12**, pp 787–795.

Casey, K., & Cornillon, P. (1999). A comparison of satellite and in situ-based sea surface temperature climatologies. Journal of Climate, 12, 1848–1863. http://dx.doi.org/10.1175/1520-042(1999)012b1848:ACOSAIN2.0.CO;24.

Coppo, P., C. Mastrandrea, M. Stagi, L. Calamai, M. Barilli, J. Nieke, (2013) The Sea and Land Surface Temperature Radiometer (SLSTR) Detection Assembly design and performance, SPIE, 8889, doi: 10.1117/12.2029432

Dana. http://www.aqua.dtu.dk/english/About/Facilities/Research_vessel

Donlon, C.J., et al (2012) The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission, *RSE*, v120, pp 37-57.

Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B., & Young, G. S. (1996). Cool-skin and warm-layer effects on sea surface temperature. Journal of Geophysical Research, 101(C1), 1295–1308 URL. http://dx.doi.org/10.1029/95JC03190

FRM. http://www.frm4sts.org/

GHRSST Data Processing Specification 2.0 Revision 5 (GDS2r5) https://www.ghrsst.org/about-ghrsst/governance-documents/

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

O'Carroll, A.G., H. Bonekamp, F. Montagner, V. Santacesaria, I. Tomazic, (2015) Sea Surface Temperature from EUMETSAT including Sentinel-3 SLSTR, Proceedings of Sentinel-3 for Science Workshop, Venice, June 2015.

O'Carrol, Anne. Personal Communication, Sentinel 3 Validation Team workshop. EUMETSAT, March 13-15, 2018

Sentinel-3 Calibration and Validation Plan, S3-PL-ESA-SY-0265, Issue 2, 2014.

SLSTR ATBD: Sea Surface Temperature, S3-L2-SD-03-T02-SU-ATBD, Issue 2.5, 28-05-2015. (http://www.eumetsat.int/website/home/Data/CopernicusServices/Sentinel3Services/index.html).

SLSTR PDFS: Product Data Format Specification, SLSTR Level 1 & Level 2 Instrument Products, (http://www.eumetsat.int/website/home/Data/CopernicusServices/Sentinel3Services/index.html).