The EUMETSAT
Network of
Satellite Application
Facilities

OSI SAF
Ocean and Sea Ice

Validation Report for the OSI SAF High Latitude L2 Sea and Sea Ice Surface Temperature

OSI-205
Version 1.1

*Gorm Dybkjær and Steinar Eastwood**

*Danish Meteorological Institute and Norwegian Meteorological Institute**

18. April 2016

This page is intentionally left blank.

Document Change Record

Version	Date	Change	Description	Responsible
1.0	17-12-2015		First version for review.	Gorm Dybkjaer
1.1	18-04-2016		Amended after ORR RID's	Gorm Dybkjaer

Table of Contents

1 Introduction

1.1 Scope

This validation report presents the High Latitude L2 Sea and Sea Ice Surface Temperature product, OSI-205, from the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF). The focus of the report is to present the overall accuracy and precision of the product. This report is the product quality baseline for subsequent half-yearly validation reports that must comply with the quality framework presented here.

1.2 Overview

The EUMETSAT OSI SAF is producing a range of operational air-sea interface products, namely: wind, sea ice characteristics, surface temperatures and radiative fluxes. More details on the products and OSI SAF project are available at http://www.osi-saf.org.

Sea and sea Ice Surface Temperature (SST, IST), sea ice concentration, type, edge and emissivity, surface solar irradiation and downward longwave irradiance products from the OSI SAF High Latitude (HL) center are produced using geostationary and polar orbiting satellites. They are available in level 2 and level 3 formats, with different timeliness depending on the production setup and satellite sensor.

This product is partly a stand-alone HL level-2 Surface Temperature product and it partly provides data for the level 3 Northern High Latitude (NHL) SST and IST product (OSI-203-a) that will be produced by OSI SAF covering the High Latitudes north of 50N. The OSI-205 is an integrated surface temperature (ST) product consisting of SST and IST data. In addition, it contains unvalidated Land Ice Surface Temperature (LIST) values for the Greenland and Antarctic ice sheets.

The OSI-205 covers the sea and ice areas polewards of latitudes 50ºN and 50ºS in 3 minute segment data in level 2. Data are being produced continuously throughout the day, as they become available. The product consists of approximately 110 3-minute segments per day. The production use AVHRR data from Metop-B from EUMETCast and cloud mask data calculated by the PPS software from the NoW-Casting Satellite Application Facility (NWC-SAF).

This report is separated in chapters describing the in situ validation data and the results obtained and a sensitivity study to illustrate special issues concerning IST validation. The OSI-205 product is described in details in the Product User Manual (PUM, [RD.1]) and in the Algorithm Theoretical Basis Document (ATBD, [RD.3]).

1.3 Glossary

Acronym	Description
AVHRR	Advanced Very High Resolution Radiometer

Acronym	Description
ATBD	Algorithm Theoretical Basis Document
DMI	Danish Meteorological Institute
HL	High Latitude
ISAR	Infrared Sea Surface Temperature Autonomous Radiometer
IST	Ice Surface Temperature
LIST	Land Ice Surface Temperature
MET	Norwegian Meteorological Institute
NOAA	National Oceanic and Atmospheric Administration
NHL	Northern HL
NWC	Nowcasting
PUM	Product User Manual
SAF	Satellite Application Facility
SMHI	Swedish Meteorological and Hydrological Institute
SST	Sea Surface Temperature
ST	Surface Temperature

1.4 Applicable documents

[RD.1] OSI SAF High Latitude L2 Sea and Sea Ice Surface Temperature Product User Manual, v1.0.

[RD.2] OSI SAF CDOP-2 Product Requirement Document, v3.4.

[RD.3] Algorithm theoretical basis document for the OSI SAF High Latitude L2 Sea and Sea Ice Surface Temperature processing chain.

2 Validation data and strategy

Most of the validation of the integrated ST product is based on data match-up between satellite estimated surface temperatures and observation data from ships and drifting buoys. In addition, this validation contains a 4 month sensitivity analysis of IST data, because there are special issues to take into account when dealing with validation of IST.

Water and ice surfaces have quit similar longwave radiative properties, but their nature makes cloud detection over the 2 surfaces rather different. From the top of the atmosphere, cloud and ice surfaces can be hard to distinguish, both spectrally and texturally, whereas water and clouds most often can be separated from their thermal signature. Therefore proper cloud classification over ice in night time fails more frequently than cloud classification over water. For this reason alone, the accuracy and precision of IST estimates is worse than the quality of SST estimates. Still, IST retrieval has the potential of being more precise than SST retrieval since the atmosphere is drier over sea ice than over open ocean.

It has also been shown that standard SST algorithms show higher uncertainties in the North Atlantic than in the Southern Ocean. In LeBorgne et al (2011) it is shown that positive bias and increased standard deviation occur during summer months in situations with high humidity at low layers, associated with a temperature inversion.

This validation report includes a range of validation and comparison exercises in order to encompass an appropriate validation period, to deal with the special IST validation issues, to cope with the validation of 2 surface types and, finally, to split the validation into the 2 hemispheres, North and South of 50ºN and 50ºS, respectively. Unfortunately we have not been able to cover all corners of this validation with equally statistical robustness.

2.1 Match-up data sets

The OSI-205 algorithms that are evaluated here have been operating in test mode since the beginning of August 2015. Similar IST/SST algorithms, with slightly different calibrations, have been operating in MyOcean since 2011, for the area North of 50ºN. Based on algorithm validation similarities during 4 month of algorithm inter-comparison (August to November, 2015) the MyOcean data will subsequently act as an extension of the OSI-205 validation back to October 2014, thus providing a proxy for a 14 month validation period.

- ***MyOcean*** IST/SST data set covers 14 month from October 2014 to November 2015. This data set is only available for the Northern Hemisphere. It is based on Metop-A data (see section 2.3).
- ***OSI-205*** IST/SST data set is available for 4 month from August to November 2015, for both Northern and Southern hemispheres (see 2.2. This product is based on Metop-B data.

The algorithms used for MyOcean and OSI-205 are conceptual identical, but they use slightly different calibration coefficients. However, OSI-205 is based on a newer PPS cloud masking software and it comes with additional data fields, such as quality levels and probability of ice and sea.

The products performances are evaluated by the STandard Deviation (STD) and Bias between satellite and in situ data sets.

The available in situ data sources for sea and ice surface temperatures are drifting buoys, moored buoys and ships. These are received locally at DMI through its Global Telecommunication System (GTS) network. Drifting buoys have shown to be the most accurate source for SST validation, whereas standard ships observations are excluded from SST validations. Observations from both buoys and ships (air temperatures) are used for IST validation.

In situ data sets are:

- ***DMI-GTS*** data stream is the DMI-operated Near Real Time (NRT) in situ data base. This is an observation data stream from The Worlds Meteorological Organization (WMO) in situ network. This data set is providing the bulk part of in situ observations used in this report and it will be the primary in situ data source for coming half-yearly validation reports of this product.
- ***DMI-AWS*** data is a 4 month multi sensor and high temporal resolution in situ data set that is recorded by a mobile DMI Automatic Weather Station (AWS) deployed on level sea ice in North West Greenland. This data set contains radiometric surface temperature measurements and traditional air temperature measurements at 1 and 2 meters and 4 sub snow temperature sensors. Results from a match up between the DMI-AWS data and the MyOcean IST data are included here to illustrate special issues to pay attention to when comparing satellite IST measurements with traditional buoy data.

The number of available drifting buoys varies a lot at Atlantic high latitudes and in particular in areas with sea ice. During special campaigns there can be many buoys available, while in “normal” years quite few buoys are operating.

DMI is in possession of an ISAR radiometer (ISAR) that is acknowledged to provide state of the art ground based validation measurements of IST and SST. The DMI-ISAR has not been deployed during the period of OSI-205 coverage, hence, there are no validation references to the ISAR instrument in this report. For future validation of OSI-205, we will apply radiometer data from the DMI-ISAR for SST validation, mainly. The DMI-ISAR is anticipated to be deployed for 3-6 month per year, on the commercial Royal-Arctic Line route between Aalborg, Denmark and Nuuk, Greenland. On occasional fields campaigns, the instrument will be deployed on sea ice and these data will be included in future OSI-205 IST validation as well. The first DMI-ISAR deployment on sea ice will be in Qaanaaq (NW Greenland) in the first week of April, 2016. As soon as possible hereafter the instrument will be deployed on Irena-Arctica, Royal Arctic Line (Irena-Arctica).

2.2 Brief description of *OSI-205* data

The product covers High Latitude Sea and Sea Ice areas polewards of 50ºN and 50ºS. The algorithm inputs are thermal infrared data from the Advanced Very High Resolution Radiometer (AVHRR) on-board the EUMETSAT Metop satellites. These data are nearly identical to those from one of the longest existing satellite records, the series of NOAA-AVHRR instruments.

The OSI-205 production has been running in test mode since August 5, 2015, based on Metop-B data. Cloud mask data are produced by the PPS software version2014 package from the NWCSAF (PPScloud). The associated ATBD [RD.3] and PUM [RD.1] describes in details the OSI SAF surface temperatures product (OSI-205) from thermal infrared satellite sensors.

2.3 Brief description of *MyOcean* data set

As mentioned above, this data set is evaluated here as a means to extent the validation of OSI-205 beyond it time of existence. The MyOcean IST and SST algorithms are conceptual identical to those of OSI-205 and the output is therefore assumed comparable. However, the MyOcean data do not provide the same level of additional data fields like probability of water and ice and quality levels and the soon-to-come uncertainty fields. Cloud mask data are produced by the PPS software version2010 package from the NWCSAF (PPScloud2010). More information of the MyOcean production can be found in Dybkjaer et al. (2012).

2.4 Match-up criteria

Satellite observations and in situ data are matched up and saved in Match up Data Bases (MDB), when satellite and in situ observations are within a certain range of each other and if the data is recorded within a certain period. For both IST and SST data the maximal allowed separation from in situ data is 5km and the maximal temporal separation is 30 minutes for both IST and for SST. Other variables like scan angle and sun-zenith angles and vicinity of clouds and quality of cloud mask are used to group the OSI-205 observations into quality levels (see [RD.1]). These quality levels are evaluated below. The MyOcean data set does not include quality levels, so the filtering of these data are done subsequently in order to make them comparable with the OSI-205 quality level 4 and 5 products (See appendix C).

The data collected through the DMI-GTS network are not quality controlled and an on-the-fly gross check is performed to filter out obviously bad data. The buoy temperature shall be within 3K of the ECMWF 2m Temperature. Likewise, the satellite SST estimate must be within 5K of the ECMWF 2m Temperature estimate, otherwise the data is rejected for suspicion of cloud. The corresponding ice buoy data must be within 5K of the ECMWF 2mT. The satellite IST estimate must be within 8K of the ECMWF 2mT, otherwise it is rejected for suspicion of cloud. The reason for using different gross check threshold values for SST and IST, is that model representation of IST is less accurate than the model representation of SST.

The choice of match-up criteria is a balance between retaining a proper volume of validation data, getting rid of outliers and not to remove real geophysical values. When a given match-up criteria gets too 'narrow' and real data are filtered out, the number of data pairs drops without gaining improvements in the product performance. In addition, when filtering observation data based on the difference NWP values we have to be extra careful, because the NWP data most likely are used to buid the 2mt NWP analysis field. The filters based on NWP temperatures are therefor set relatively 'broad'.

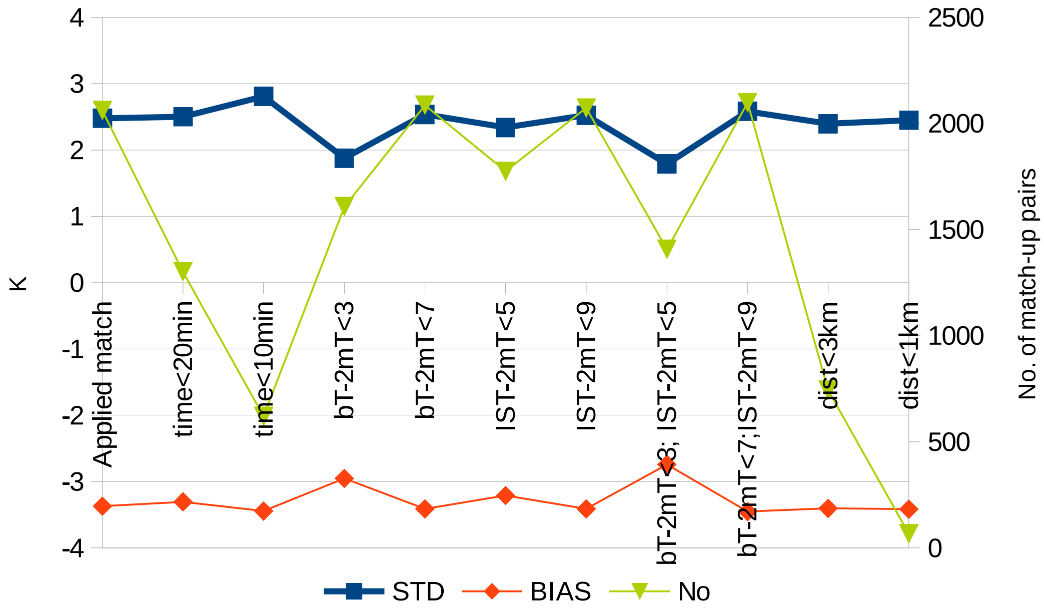

Figure 1: Sensitivity analysis of the effect of changing match-up criteria for OSI-205 IST. Far left are values of STD, Bias and No. of values for the weighted and average (4 month) IST in quality 5, using the applied match-up criteria (se text). The consequtive values are obtained by changing 1 critiria after the other. The criterias are (from left to right): 1) Applied criterias from figure 5, 2) Time difference less than 20 minutes, 3) time difference less than 10 minutes, 4) Buoy temperature minus NWP 2m Temperature less than 3 K, 5) Buoy temperature minus NWP 2m Temperature less than 7 K, 6) OSI-205 IST minus NWP 2m Temperature less than 5 K, 7) OSI-205 IST minus NWP 2m Temperature less than 9 K, 8) Criteria 4+6 combined, 9) Criterias 5+7 combined, 10) Distance less than 3 km, 11) Distance less than 1 km.

The performance of IST and SST is plotted in figures **1** and 2, respectively, for a range of match-up criteria thresholds, where one threshold is changed at a time. As expected, the number of excluded data match-up pairs drops drastically when narrowing down time and space thresholds. This does not have a significant improvement on OSI-205 performance. The figures **1** and **2** also show that data filtering using NWP temperatures has the potential to improve OSI-205 surface temperature performance, particular for IST. However, we are reluctant to sharpen these thresholds because of a likely inter-dependency with the in situ observations and the NWP temperature fields. It is not clear why the SST/IST performance does not improve with more strict spatial and temporal thresholds. This may be related to limited accuracy of the buoy data and positions.

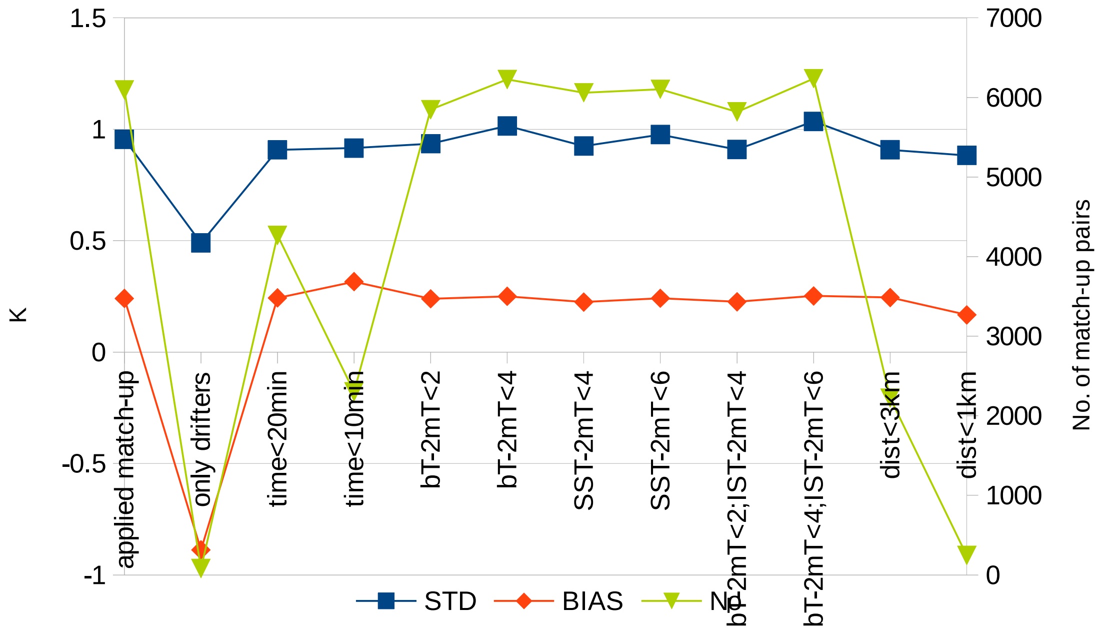

Figure 2: Sensitivity analysis of the effect of changing match-up criteria thresholds for OSI-205 SST. Far left set of STD, Bias and No. is the average (4 month) night time QL-5 SST, using the applied match-up criteria (se text). The consequtive values are obtained by changing 1 critiria after the other. The criterias are (from left to right): 1) Applied criterias (The mean ql-5 values from Figure XX2, 2) Buoy drifters only, 3) Time difference less than 20 minutes, 4) time difference less than 10 minutes, 5) Buoy temperature minus NWP 2mTemperature less than 2 K, 6) Buoy temperature minus NWP 2m Temperature less than 4 K, 7) OSI-205 SST minus NWP 2m Temperature less than 4 K, 8) OSI-205 SST minus NWP 2m Temperature less than 6 K, 9) Criteria 5+7 combined, 10) Criterias 6+8 combined, 11) Distance less than 3 km, 12) Distance less than 1 km.

The SST performance improves significantly when using drifting buoys only, but the number of data pairs is largely reduced. The reason for this is not clear, but there may be a bias towards fewer quality level 5 data at open sea, where the free drifters are available.

The subsequent validations are based on all pairs of satellite/observations data that comply to the spatio-temporal criteria mentioned above. This is to maximize the statistical basis. This validation procedure will give a slightly larger STD error than validation procedures that average all satellite measurements that comply with one in situ measurement, which is the procedure elsewhere in the SST community. The reason for higher error statistics from pixel to buoy comparison rather that average pixel to buoy comparison is that the random and unbiased instrument error for SST and IST algorithms is around 0.2K for temperatures larger than 240K and a little less for colder temperatures. This is shown in figure 3 where ST's from several Metop-A segments are calculated using perturbed Tb values. The Tb perturbations are within the instrument random uncertainty noise, the so called NEdT values. The unperturbed ST's are subtracted the ST values that are calculated from perturbed Tb values (St*perturbed*). The random uncertainty error is the STD of all pairs of ST-ST*perturbed* values. Moreover, partly random and partly locally correlated SST algorithm uncertainty values for Metop-A and B are between 0.2K and 0.3K and the corresponding uncertainty for the IST algorithm is around 0.1K (EUSTACE, 2015). These are the STD values obtained when tuning the algorithms to the algorithm tuning data set based on NWP and RTM data, see the ATBD [RD.3].

The uncorrelated and partly random uncertainties (the instrument and algorithm uncertainties, respectively), equalize themselves when averaged over a number of pixels. The root of the summed squares of these two uncertainty is the theoretical maximum STD value that one can expect the non-averaged STD is higher that the averaged ST estimate, i.e. 0.36K and 0.22K for SST and IST, respectively.

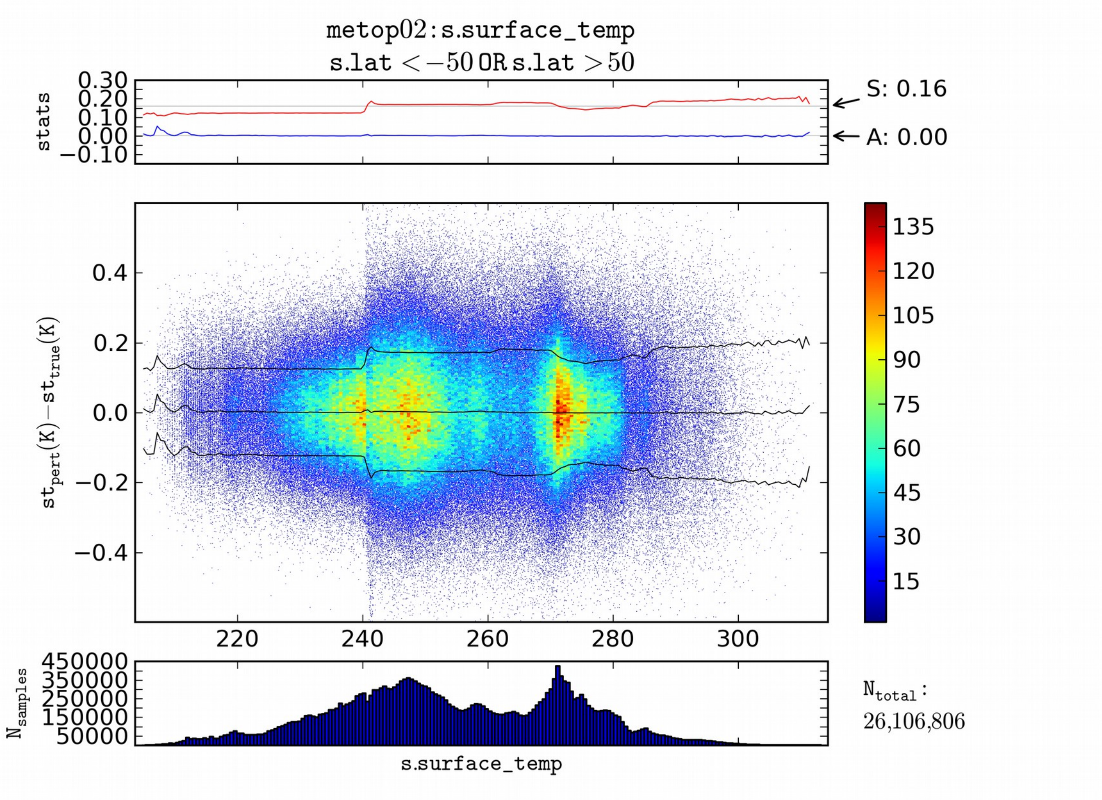

Figure 3: Difference between OSI-205 ST algorithm and corresponding surface temperature calculated from perturbed Tb's. All Tb perturbations are within the instrument random noise level (NEdT value). Top plot is the STD and Bias (S and A) of ST-STperturbed as a function of ST. Middle plot shows the density plot of all ST-STperturbated values. For Metop—A data.

However, in an analysis where we compared area mean surface temperatures with buoy measurements we did not find the anticipated improvement in performance. We believe that a more idealized data set would reveal an performance improvement by using area mean values for validation and that the GTB buoy data are too noisy too validate this statement. In the CDOP-3 proposal we suggest a work package that deals with compilation of quality ensured buoy temperature and position data for IST and Ice drift validation purposes.

2.5 Product accuracy requirements

The required accuracy of the products are defined as monthly mean bias and standard deviation of the surface temperatures values compared with in situ measurements. Three requirement levels are defined in the PCR v3.4 [RD.2]:

- *Threshold* – The model user community gain no improved model performance using data of worse quality than this.
- *Target* – This is an intermediate quality level, between the two extremes (Threshold and Optimal), at which the product quality aim at.
- *Optimal* – The model user community can not gain from improvements in the ST quality beyond this level.

The IST accuracy requirements are split in two: 1) requirements for validation against in situ IR radiometers, and 2) requirements for validation statistics against in situ buoy data. This is discussed in chapter 4. All threshold accuracies are given in table 1.

	Treshold, std/bias, K		Target, std/bias, K		Optimal, std/bias, K	
	Radiometer	Buoy	Radiometer	Buoy	Radiometer	Buoy
SST	--	1.5/1.5	--	1.0/0.7	--	0.3/0.1
IST	3.0/2.5	4.0/4.5	2.0/1.5	3.0/3.5	0.8/0.5	1.0/0.8

Table 1: SST and IST temperature quality requirements thresholds (from [RD.2]).

3 Validation results

This chapter is divided into 5 sub-sections; The OSI-205 SST and IST for both Northern and Southern hemispheres and the 14 month MyOcean data set covering Northern hemisphere IST. The most important results are presented here. All relevant and used validation results are presented in tables in appendix A, B and C, covering OSI-205 IST, OSI-205 SST and MyOcean results, respectively.

All validation results are calculated from all data pairs of in situ and satellite ST data that comply with the match-up criteria mentioned in section 2.4, i.e. no data averaging is performed. The validation of SST polewards of latitude 50 N as based on data from the SST night algorithm, which correspond to solar zenith angles greater than 110 degrees. This is the anticipated best performing time-of-day for the SST algorithm because the divergence between in situ buoy temperatures and the actual sea surface skin temperature is at a minimum. There was not sufficient in situ observations from the SH to perform the same stratification for the SH SST data. The IST validations are not stratified into algorithm, because no algorithm specific differences were found.

3.1 SST North of 50N

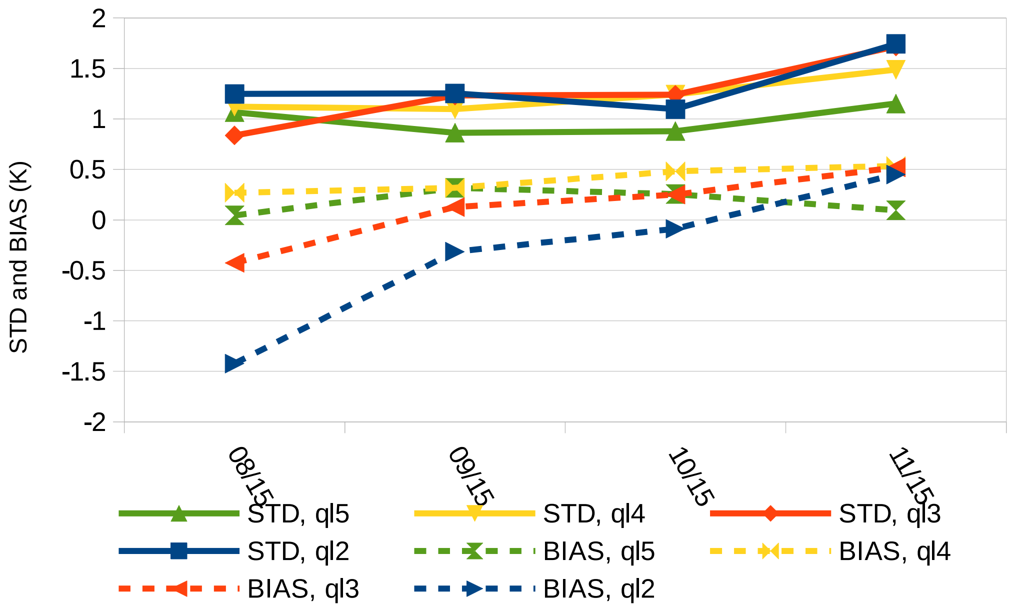

Figure 4: OSI-205 SST standard deviation and Bias for quality levels 2-5 for OSI-205 August to November, 2015, for the Northern hemisphere. Results are for night time data, and drifting and moored buoys only.

The performance of the SST night algorithm for quality levels 2 to 5 show slight improvements in STD and Bias towards 'best quality', as expected (figure 4). The STD level for ql 4+5 is approximately 1.1 K and the corresponding Bias is 0.3K. For daytime (solar zenith angle < 80) the STD and Bias for ql 4+5 is 1.39 K and -0.19 K, respectively. See appendix B for more validation statistics, both for night time and day time.

3.2 SST South of 50S

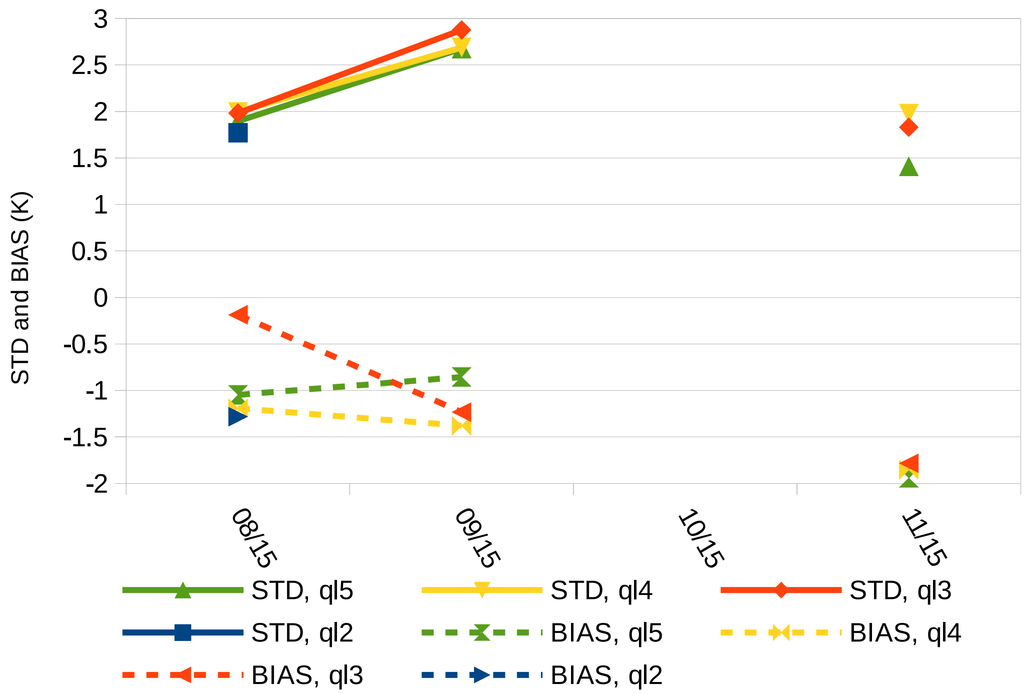

Figure 4: OSI-205 SST standard deviation and Bias for quality levels 2-5 for OSI-205 August to November, 2015, for the Sourthern hemisphere. Both Ship and buoy data from DMI-GTS data set are used to improve statistics.

In the southern hemisphere the in situ data coverage is significantly lower and the majority of the in situ data here are recorded from ships. So, the ship data can not be discarded in the SH due to extreme poor in situ data coverage. Hence, since using ship observations, the results presented can not be expected to meet the accuracy requirement. The performance of the SH SST night and day algorithms for quality levels 2 to 5 show slight improvements in STD towards 'best quality' in figure 5, whereas the Bias improvement, going towards better quality is less clear. The STD level for ql 4+5 is approximately 2.1 K and the corresponding Bias is -1.4K. See appendix B for more validation statistics.

3.3 IST North of 50N

The ice covered areas in the NH is relatively well covered by drifting buoy and ship observations that are reported to the GTS data archiving system. This is reflected in the relative large number of samples (data pairs) used in this NH IST validation (see appendix A). However, buoy data are of a dubious character, as is illustrated in chapter 4. Unfortunately there is no alternative in situ observation data source with a sufficient amount of data to perform monthly validations of IST products like OSI-205 and MyOcean date sets. The OSI-205 performance with respect to STD values is almost constant for all quality levels, with a mean over 4 month of 2.8 K. Only the bias value improves with increasing quality level, except for an outstanding bias value for ql-5 on November 2015 (see figure 5 and appendix A). The Bias improvement gained from ql-2 to ql-5 is -4.11 K to -3.45 as shown in table 2. The fact that the STD does not improve with increasing quality level suggests that the quality level algorithm can be improved, and, thus that quality levels 4 and 5 can improve from their current levels of error and bias. These values comply to the target requirements in the Product Requirement Document ([RD.2])

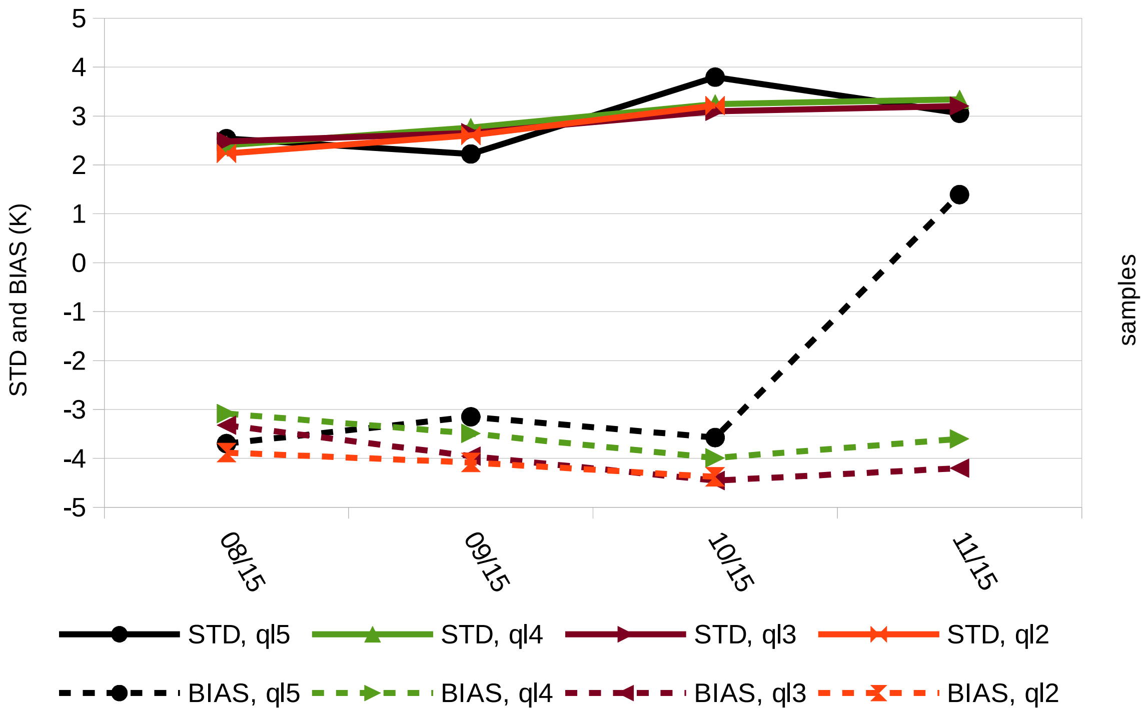

Figure 5: OSI-205 IST standard deviation and Bias for quality levels 2-5 for OSI-205 August to November, 2015, for the Northern hemisphere.

3.4 IST South of 50S

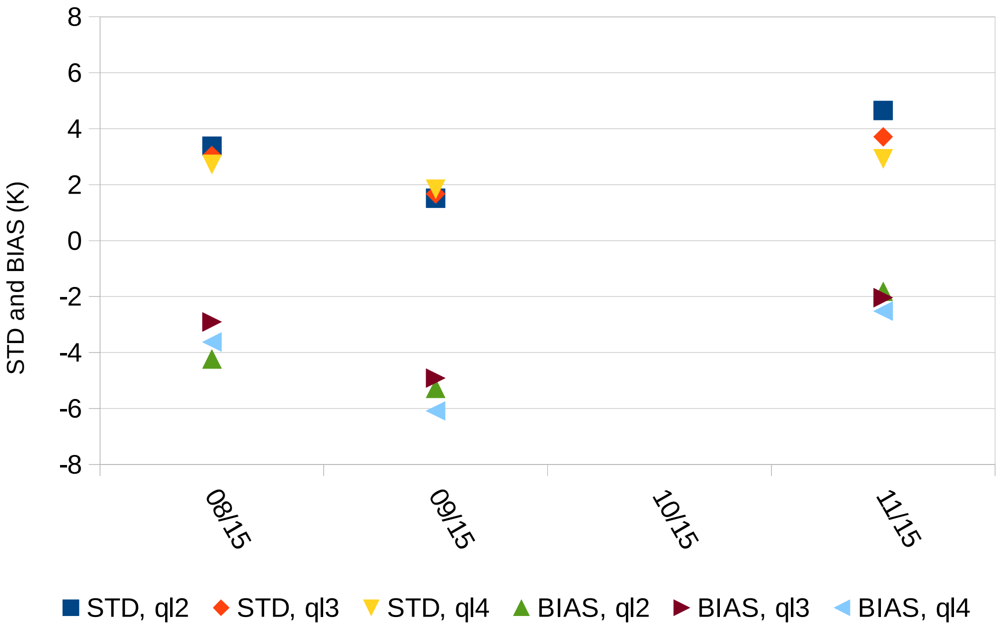

Figure 6: OSI-205 IST standard deviation and Bias for quality levels 2-4 for OSI-205 August to November, 2015, for the Southern hemisphere. There was not enough data to evaluate October.

It was not possible to find many in situ surface temperature observations for the Antarctic sea ice, only ~1300 satellite/in-situ data pairs in total. The results are plotted in figure 6. For the month of October there was not sufficient data to calculate proper statistics. Further, ql-5 is not represented in this SH data set, because there were too few high quality cloud mask data available (see quality level algorithm in PUM [RD.1]). However, there is no clear difference in performance between the data from the 3 remaining quality levels, showing average STD and Bias of 2.83 K and -3.7 K, respectively. Hence, the target STD ([RD.2]) is met with a margin whereas the 4 month mean bias is slightly larges than the target requirement for validation against buoy data. It remains to be shown that the annual mean bias will comply to the requirements.

3.5 IST North of 50N – 14 month validation

This section covers the validation of 14 month of MyOcean IST data. As mentioned, this product uses a similar algorithm to the OSI-205 IST product. It is included here to evaluate whether the IST data sets show inter-annual dependencies with respect to performance. The MyOcean data set overlaps the OSI-205 test data set by 4 month and it is intended to use the MyOcean validation to 'extrapolate' the OSI-205 validation beyond the 4 month, if the MyOcean and the OSI-205 data perform comparable during the overlapping period.

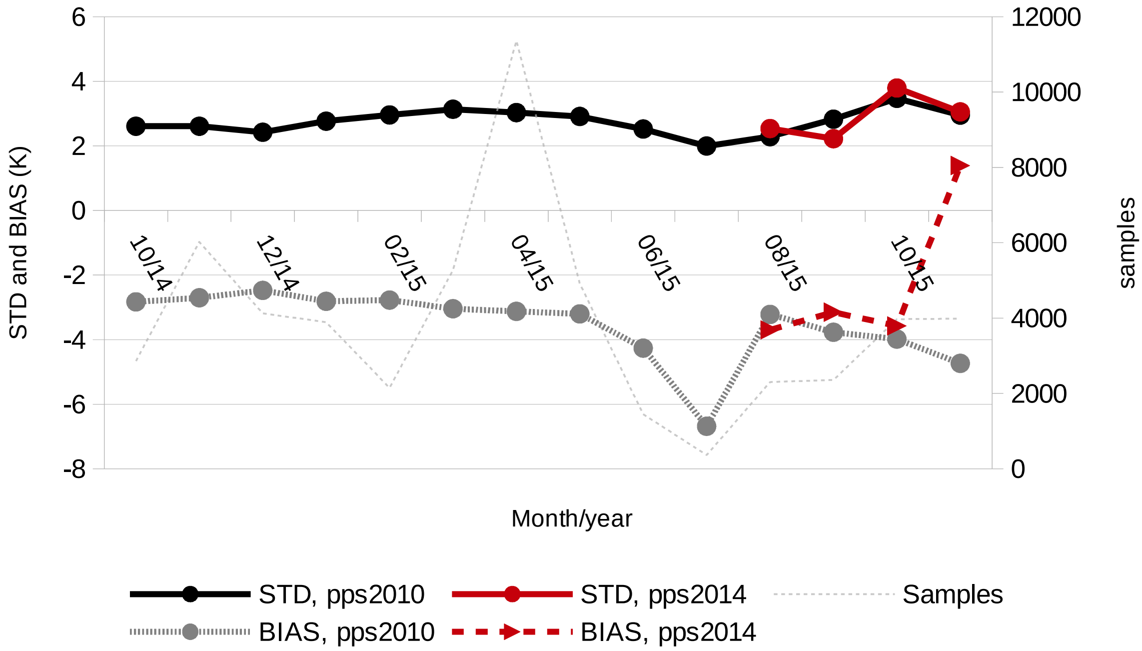

Figure 7: Validation of the MyOcean setup for IST - over 14 month. OSI-205 ql-5 validation for the overlapping period is added.

The MyOcean data set STD is 2.75K and BIAS is -3.54 as a mean of all 14 month, hence, the STD comply to the requirements with a relative large margin whereas the annual bias is on the limit . See full statistics in appendix C.

There seems to be no distinct inter-annual variability in product performance, apart from Bias dip in July 2015, and the overlap between the MyOcean and OSI-205 data sets show almost identical STD and Bias, as illustrated in figure 7. Hence, we consider the period August-November to be representative for a full year and the MyOcean data set to be representative for the OSI-205 product quality. In fact, the OSI-205 IST performs slightly better in Bias, probably due to better cloud detection in OSI-205 production chain. The jump in November Bias (1.39 K) of the OSI-205 data is most probably caused by erroneous observation and poor statisatics (37 data pairs), because a positive bias between satellite data and ground observations is not credible.

4 IST validation issues

In particular 2 circumstances make the quality assessment of satellite based IST worse that that of SST:

1) Failing cloud detection both with respect to false alarms and probability of detection.
2) Large vertical and diurnal temperature variation over sea ice occur in particular in spring and autumn.

There exist a pronounced higher frequency of false cloud classification over sea ice than over open water. Sea ice and clouds often show similar radiometric and structural properties and cloud classification fail therefore relatively frequent. On the other hand, the radiometric contrast between open ocean and clouds is normally large.

The consequense of falsely detected clouds over ice, is a reduced data set and the consequence of lower probability of detection is a cold bias in the IST estimates, because the cloud tops usually are colder than the ice surface. This is a difficulte issue to cope with, because this cold bias is not a general bias, but an effect that either is there or not. A normal bias correction is therefore not a good solotion to this problem. The only solution seems to be a better cloud mask or by using physical based compensations in seaice models – both of these solutions are naturally out of the scope of this text.

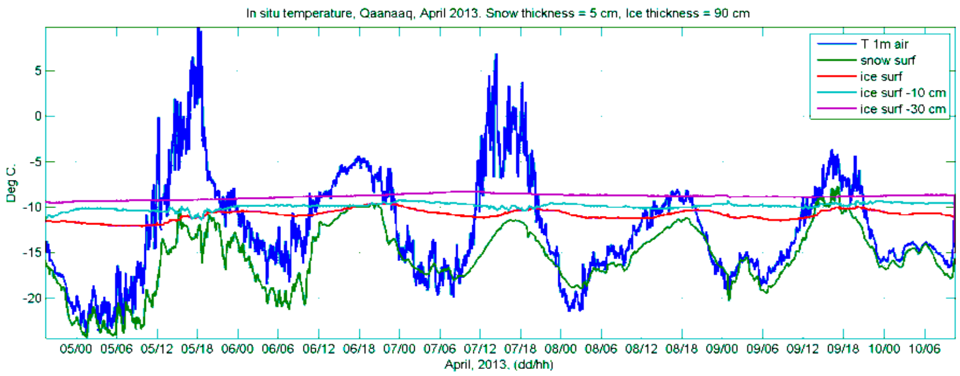

Figure 8: Diurnal temperature variation measured in situ from 1m air to 0.3m ice depth.

Extreme vertical and diurnal temperature variation in sea ice, snow and air makes IST comparison with traditional temperature observations ambiguous and dubious. Figure 8 is showing 5 days of temperature variations at 5 vertical levels, between 1m in the air and 30 cm ice into the ice, with a 5 cm snow cover. There are four important features to note from this short data set that should be dealt with when comparing IST with buoy measurements:

- The skin temperature rate can be as large as 2K per hour.

- The total diurnal skin temperature variation can be 15K between day and night.
- The air temperature variability can be much larger than the skin temperature, with occasional very high air temperature spikes if the sensor is exposed to direct sun.
- There hardly exist any diurnal temperature variability under a few cm of snow.

When dealing with ice tethered buoy measurements one can not know the exact position and state of the temperature sensor. The sensor can be buried under snow or exposed to sun or represent some kind of an air temperature. In either case, the ice skin temperature is rarely equal to any of the other temperature measurements in the data set in figure 8. The consequences are clear in the following section.

4.1 An idealized IST validation study

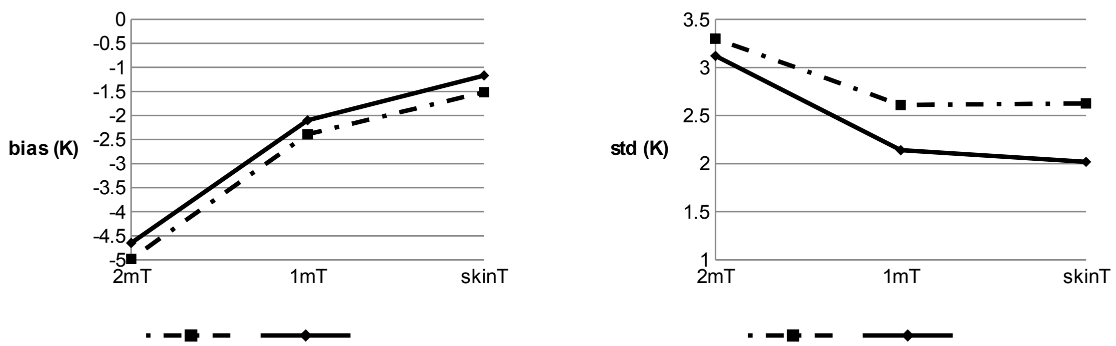

Figure 9: IST vs in Situ: STD and BIAS from comparing the MyOcean data set with 2m, 1m surface temperatures - within 10 min. (solid line) and 30 min. (dashed line).

The DMI-AWS data set is similar to the 5-day data record shown in figure 8, but it covers 4 month of winter and spring data from level ice in NW Greenland, thus making it suitable for closer analysis of temperature variability.

In figure 9 a comparison between MyOcean skin temperature data and in situ 2m temperatures, 1m temperatures and skin temperatures is shown. Also effects from changing the temporal match-up criteria from 30 minutes to 10 minutes is shown.

The bias improves from -5 K when compared with 2m temperature to -1.5 K when comparing MyOcean data with skin temperature. A smaller further improvement occur when changing the match-up period from 30 minutes to 10 minutes, namely from -1.5 to -1.3 K. The corresponding average bias for MyOcean data compared to DMI-GTS data is 3.04 K (Appendix C).

Likewise, the STD changes from 3.3K to 2.6 K. when changing the 2m temperature match-up to match-up against skin temperature. Further STD improvement from 2.6 K to 2.0 K occur when changing match-up period from 30 minutes to 10 minutes. The corresponding average STD for MyOcean data compared to DMI-GTS data is 3.01 K (See appendix C).

The DMI-AWS data are trustworthy and we assume these values to give a realistic picture of the real satellite performance, thus leaving the representativeness of buoy data for skin temperature doubtful. However, the buoy data are the only data that we can obtain in relatively large numbers.

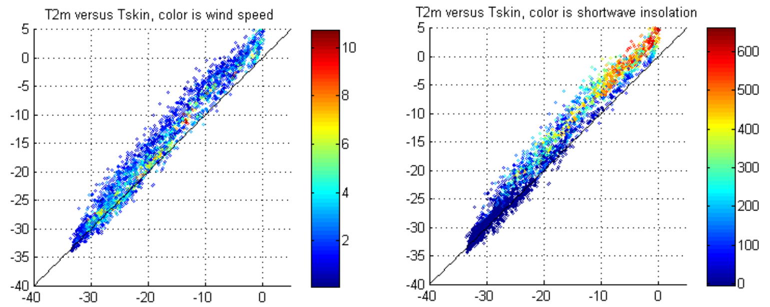

Figure 10: Scatter plots of skin temperature (first axis) vs 2m Temperature (second axis) with wind speed (left) and short wave insolation (right) using the colour scales.

In figure 10, we have illustrated some of the factors that influence the air/skin temperature divergence, namely wind speed (left) and shortwave insolation (right). The plot shows skin temperatures as a function of 2m temperature and wind speed and insolation plotted in the respective color codes. The data show that the temperature divergence is lowest at high winds and at low or no insolation.

5 Discussion

The main results are summarized in table 2, showing 4 month of mean STD, Bias and number of samples for all data set involved. The MyOcean validation results are compared with the OSI-205 performance in an attempt to extend the OSI-205 validation period beyond its 4 month of valid data. This is done by using the MyOcean validation data as proxy for the OSI-205 performance.
The IST product quality has not yet reached a proven and widely acknowledged set of requirement as is the case for SST.The values in table 1 are determined from Stammer et al. (2007) and from the literature (Dybkjaer et al., 2012, Hall et al. 2004).

IST STD for SH and NH varies between 2.68 and 3.12K which comply well with the product requirements for buoy validation, mentioned above. The product Bias lies between 3.07 and 3.78 K for all QL's and between 3.07 and 3.42K for ql 4 and 5, where the latter is within target *threshold*. In chapter 4 it was shown that the real STD and Bias probably is far less, due to a dubious skin representation by the buoy measurements in general. It is likely that the real STD and Bias of OSI-205 can be as low as 2.0 K and 1.3 K, respectively.

Domain	Quality level 2	Quality level 3	Quality level 4	Quality level 5	Mean all ql. Sample sum
IST, NH	2.68/-4.11/19	2.86/-3.98/158	2.94/-3.54/46	2.9/-3.47*/2	2.85/-3.78*/226
IST, SH	3.12/-3.78/.4	2.8/-3.28/.8	2.83/-3.07**/.1	---	2.95/-3.42/1.3
SST, NH	1.33/-0.34/1	1.26/0.12/13	1.24/0.40/44	0.96/0.24/6	1.20/-0.02/64
SST, SH	1.78/-1.28/.01	2.23/-1.07/.5	2.2/-1.47/2	1.99/-1.28/0.1	2.05/-1.37/3
IST, MyOcean			2.89/-3.93/13		

Table 2: STD/BIAS/Samples(x1000) error statistics (in K) of all data sets for August to November, 2015. In situ observations are the DMI-GTS ship and buoy data. In the SST-NH statistics only buoy data are used. *(Quality level 5 from November is excluded, due to positive bias that indicates erroneous observation data) **(Quality level 4 from September is excluded, due to too few samples(11 samples!)).

Validation of the SST algorithm has been performed using buoys only for NH and buoy and ship observations for the SH. For the NH area the STD for ql 4 and 5 are around 0.96-1.2 K and Bias close to 0.3 K. With respect to Bias, this is within the target accuracy of 0.8K and a little above with respect to STD (1.0K). In chapter 4 it was argued that the STD value correspond to a STD value of less that 1 K when compared to averaged STD values that are presented by other SST producers.
Taking this into account the level of NH STD of this OSI-205 SST product is just around the target of the product.
Southern hemisphere SST validation shows STD values around 2 K and corresponding Bias around -1.3 K. This is somewhat off the target of the product. However, is is likely that the real performance of the SH SST is similar to the NH performance. The SH performance is calculated for night and day algorithms jointly

and against mainly ship measurements, which degrades the comparison. The STD values for the night algorithm is approximately 0.4 K better than the validation of the joint algorithms and STD increased by approximately 0.2 K by including the ship observations in the validation data set.

6 Conclusion

The OSI-205 SST product performs slightly worse than expected at high latitudes, when comparing with previous validation community standards. However, high latitudes SST is also known to be the area of worse SST error statistics.

For the NH area, the product validation shows that the product is within target accuracy for the bias for all quality levels, and within target accuracy for standard deviation for the highest quality level (level 5). It is slightly outside the target accuracy for the lower quality levels. So for SST at NH the quality is just good enough for operational distribution.

Very few available observations for the Southern Hemisphere SST makes it difficult to draw conclusive remarks about the quality for this area. When using he few results that are available, which includes lower quality ship observations, the results suggest that the STD are not within the target accuracy requirements and that the Bias is too high. But more drifting buoy data are needed to conclude about the quality on the SH, and since the quality on NH is within requirements, we suggest to start distributing the product with an operational status.

The OSI-205 IST performs as expected with performance values generally within *Target, in both hemispheres* . The quality level tests are not functioning well, resulting in only marginal improvements between quality levels. In the future we will collect more validation statistics and we anticipate to improve the quality level algorithm and thereby improve quality for quality levels 4 and 5 data, at the expense of the low quality levels.

When the probability fields are incorporated in the quality level tests we anticipate both SST and IST statistics to improve for the highest quality levels.

7 References

Dybkjær, G., R. Tonboe, and J. L. Høyer: Arctic surface temperatures from Metop AVHRR compared to in situ ocean and land data. Ocean Sci., 8, 959–970, doi:10.5194/os-8-959-2012.

Dybkjær, G., J. Højer, R. Tonboe and S. Olsen. Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST. NACLIM Deliverable D32.28. http://www.naclim.eu/ ; 2014.

ERAint. http://www.ecmwf.int/en/forecasts/datasets/era-interim-dataset-january-1979-present, 2014-09.

François C., A Brisson, P. Le Borgne, A. Marsouin: Definition of a radiosounding database for sea surface brightness temperature simulations: Application to sea surface temperature retrieval algorithm determination. Remote Sensing of Environment, 81, 2–3, pp 309–326, 2002.

Hall, D. K., Key, J. R., Casey, K. A., Riggs, G. A., and Cavalieri, D. J.: Sea Ice Surface Temperature Product from MODIS, IEEE T. Geosci. Remote., 42, 1076–1087, 2004.

Irena-Arctica Royal Arctic Line. http://www.ral.dk/sejlplaner/skibe/irena-arctica/

ISAR Infrared Sea Surface Temperature Autonomous Radiometer. http://www.isar.org.uk/LeBorgne, P., S. Pere and H. Roquet, 2011. Errors and correction of Metop/AVHRR derived SST in Arctic conditions. 2011 EUMETSAT Meteorological Satellite Conference.

PPScloud2010. The Product User Manual for cloud processing software PPS version 2010 from NWCSAF (see PPScloud). Reference not found.

PPScloud. The Polar Platform System package, Product User Manual for "Cloud Products". http://www.nwcsaf.org/HD/MainNS.jsp, NWC/CDOP2/PPS/SMHI/SCI/UM/1, Issue 1, Rev. 0. September 2014.

PRD. OSI SAF Product Requirement Document, v3.0, 2014. http://www.osi-saf.org/biblio/docs/osi_cdop2_gen_prd_3_3.pdf#page=24&zoom=auto,126,-5

Stammer, D., Johanessen, J., LeTraon, P.-Y., Minnett, P., Roquet,H., and Srokosz, M.: Requirements for Ocean Observations Rel-evant to post-EPS, EUMETSAT Position Paper: AEG Ocean To-pography and Ocean Imaging, 10 January 2007, version 3, 2007.

Appendix A

Error statistics for the OSI-205 IST data set.

Match-up has been performed using following criteria:

Data must be within 30 minutes and 5 kilometers, buoy temperature and ECMWF 2mT must be within 5K and calculated IST and ECMWF 2mT must be within 8K of each other. IST must be colder than -3K. Validation results are calculated for quality levels 2-5. These are explained in the PUM [RD.1]

IST	2015	ql.2	ql.3	ql.4	ql.5
NH	Aug.	2.23/-3.88/.6	2.48/-3.32/11	2.40/-3.08/21	2.54/-3.70/1
	Sep.	2.60/-4.08/8	2.66/-3.95/57	2.76/-3.49/15	2.22/-3.15//.8
	Oct.	3.21/-4.38/10	3.10/-4.45/52	3.24/-3.99/6	3.79/-3.57/.1
	Nov.	---	3.20/-4.20/38	3.34/-3.60/4	3.05/1.39/.04
SH	Aug.	3.38/-4.23/.2	3.04/-2.90/.4	2.72/-3.63/.04	---
	Sep.	1.51/-5.28/.1	1.67/-4.91/.2	1.84/-6.09/.01	---
	Oct.	---	---	---	---
	Nov.	4.66/-1.81/.09	3.71/-2.04/.2	2.93/-2.52/.09	---

Table 3: Validation statistics for IST NH and SH (STD/BIAS/SAMPLES($x10^{-3}$)) against DMI-GTS data.

Appendix B

Error statistics for the OSI-205 SST data set.

Match-up has been performed using following criteria:

Data must be within 30 minutes and 5 kilometers, cloud mask must be of best quality and classified as "cloud free" or "ice contaminated", buoy temperature and ECMWF 2mT must be within 3K and calculated SST and ECMWF 2mT must be within 5K of each other. SST must be warmer than 0K.

SST Night	2015	ql.2	ql.3	ql.4	ql.5
NH	Aug.	1.25/-1.42/.04	0.84/-0.43/.4	1.12/0.27/.2	1.07/0.04/.02
	Sep.	1.25/-0.32/.7	1.23/0.13/6	1.10/0.32/19	0.86/0.32/3
	Oct.	1.10/-0.09/.04	1.24/0.25/.2	1.24/0.48/5	0.88/0.26/.7
	Nov.	1.74/0.45/.3	1.72/0.52/5	1.49/0.53/19	1.15/0.10/2
SH	Aug.	1.77/-1.28/.01	1.98/-0.19/.2	2.00/-1.19/.3	1.90/-1.05/.01
	Sep.	---	2.88/-1.23/.2	2.69/-1.38/1	2.68/-0.86/.03
	Oct.	---	---	---	---
	Nov.	---	1.83/-1.78/.1	1.98/-1.85/.6	1.41/-1.94/.07

Table 4: OSI-205 OSI-205 SST validation statistics (STD/BIAS/SAMPLES(x10^{-3})) against DMI-GTS data. Statististics from NH is SST night algorithm (sun-zenith angle > 110 degree) against buoys only (ship data excluded). Statistics for SH is both night and day algorithm compared against all DMI-GTS data (mainly ship observations).

SST Day	2015	ql.2	ql.3	ql.4	ql.5
NH	Aug.	1.37/-0.95/2	1.33/-0.71/17	1.33/-0.55/74	1.84/-1.22/1
	Sep.	1.48/-0.22/1	1.54/0.15/16	1.47/0.14/52	2.07/-0.18/.9
	Oct.	1.20/-0.61/.4381	1.17/-0.08/4	1.14/0.29/12	1.45/-0.43/.1
	Nov.	1.78/0.29/.2	1.50/0.26/4	1.49/0.11/14	2.23/-0.58/.2

Table 5: OSI-205 SST validation statistics (STD/BIAS/SAMPLES(x10^{-3})) against DMI-GTS data. Statistics from NH is SST day algorithm (sun-zenith angle < 80 degree) against buoys only (ship data excluded).

Appendix C

Error statistics for the MyOcean IST data set.

Match-up has been performed using following criteria:

Satellite and in situ observation data must be within 30 minutes and 5 kilometers of each other, cloud mask must be of best quality and classified as "cloud free" or "ice contaminated", scan angle is less than 60 degree, in situ temperature and ECMWF 2mT must be within 5K and MyOcean IST and ECMWF 2mT must be within 8K of each other. These criterias correspond best to the quality levels 4 and 5 of the OSI-205 data set. IST must be less than -3K

Month/year	STD (K)	Bias (K)	Samples
10/14	2.61	-2.83	2858
11/14	2.61	-2.70	6022
12/14	2.42	-2.48	4127
01/15	2.77	-2.81	3892
02/15	2.96	-2.78	2148
03/15	3.14	-3.04	5278
04/15	3.03	-3.12	11365
05/15	2.91	-3.20	4916
06/15	2.52	-4.26	1447
07/15	2.00	-6.68	364
08/15	2.29	-3.22	2302
09/15	2.83	-3.77	2359
10/15	3.47	-3.98	3972
11/15	2.95	-4.74	3986

Table 6: MyOcean IST validation statistics: STD, BIAS and SAMPLES. Validation is performed against DMI-GTS data.