Funded by the European Union

Sea-ice Surface Temperature Retrieval and Validation for Copernicus Sentinel-3 Sea and Land Surface Temperature Radiometer

Product validation and evolution report

Document Reference Number: EUM/OPS-COPER/21/1213711

Authors: Gorm Dybkjær, Wiebke M. Kolbe and Emma Dodd (University of Leicester)

Version 1.2

PVRv2

Deliverable 13 (PVR v2)

EUMETSAT ITT No. 215580; Contract No. EUM/CO/18/4600002129/AOC

The Danish Meteorological Institute

08-02-2021

Scope of this report

This Product Validation and evolution Report (PVRv2) describes the performance of the two algorithms implemented in the Copernicus Sentinel-3 Sea and Land Surface Temperature (SLSTR) sea-Ice Surface Temperature (IST) prototype processor. The selection of these two algorithms, and the associated cloud screening procedures, is based on an earlier evaluation of 15 IST algorithms and cloud screening products in PVRv1 [AD-8].
Most algorithms tested in PVRv1 are described in the first version of the ATBDv1 [AD-4] while the new ideas emerging during the evaluation are documented in [AD-8]. The IST product requirements and a review of earlier and present IST algorithms for both single view and dual view sensors are available in the Requirement Baseline Document [AD-1]. The validation data sets and the validation strategy used in this report, are described in the Product Validation Plan [AD-2]

Document change record

Version	Date	Author	Description
D10 V1.0	June 2020	Gorm Dybkjaer	Final version of the D10 validation report
v1.0	September 2020	Gorm Dybkjaer	Final validation report
v1.1	December 2020	Gorm Dybkjaer, Wiebke Kolbe and Emma Dodd	RID revision
v1.2	February 2021	Gorm Dybkjaer	Mainly editorial updated and future FRM considerations

Applicable Documents:

[AD-1] Requirement Baseline Document (RB, D4), Ref. No. EUM/OPS-COPER/19/1065840
[AD-2] Product Validation Plan (PVP, D5), Ref. No. EUM/OPS-COPER/19/1065836
[AD-3] Input Output Data Definition Document (IODD, D6.1), Ref. No. EUM/OPS-COPER/19/1083003
[AD-3.2] Input Output Data Definition Document (IODDv2, D6.2, v1.3), Ref. No. EUM/OPS-COPER/19/1083003
[AD-4] Algorithm Theoretical Basis Document – working paper (ATBDv1, D7.1)
[AD-4.2] Algorithm Theoretical Basis Document (ATBDv2, D7, v1.1), Ref. No EUM/OPS-COPER/20/1205462
[AD-5] Project Proposal (internal document available at EUMETSAT)
[AD-6] QA4EO documentation. (http://qa4eo.org/documentation.html)
[AD-7] Statement of Work (SoW)
[AD-8] Product Validation and evolution Report (PVR v.1, D10)

AD-1 and AD-2 are publicly available at: https://www.eumetsat.int/S3-SLSTR-SIST.

Table of content

Acronyms and abbreviations

AVHRR	Advanced Very High Resolution Radiometer
AWS	Automatic Weather Station
BAS	British Antarctic Survey
CCI	Climate Change Initiative
CDR	Climate Data Record
CEOS	Committee on Earth Observation Satellites
CRREL	Cold Regions Research and Engineering Lab
ESA	European Space Agency
EUMETCast	EUMETSAT's primary dissemination mechanism
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FRM	Fiducial Reference Measurements
GEO	Group on Earth Observations
GDS	GHRSST Data Specification
GHRSST	The Group for High Resolution Sea Surface Temperature
IMB	Ice Mass-balance Buoy
IR	Infrared
IST	Ice Surface Temperature
L2, L3, L4	Level-2, Level-3, Level-4
LST	Land Surface Temperature
Metop	Meteorological Operational satellite (EUMETSAT)
MIZ	Marginal Ice Zone
MIZT	Marginal Ice Zone Temperature
MODIS	Moderate Resolution Imaging Spectroradiometer
MUDB	Match-Up Data Base
NH	Northern Hemisphere
NTC	Non Time Critical
NWP	Numerical Weather Prediction
OSISAF	EUMETSAT Ocean and Sea Ice Satellite Application Facility
OSI-205	OSISAF operational L2 IST product based on Metop AVHRR and VIIRS data
PROMICE	Programme for Monitoring of the Greenland Ice Sheet
QC	Quality Control
RBT	Radiances and Brightness Temperature
RTTOV	Radiative Transfer for TOVS (TIROS Operational Vertical Sounder)
SAMS	Scottish Association for Marine Science
SH	Southern Hemisphere
SLSTR	Sea and Land Surface Temperature Radiometer
SOW	Statement of Work
SST	Sea Surface Temperature
STD	Standard Deviation
TB	Brightness Temperature
TCWV	Total Column Water vapour
TOA	Top of atmosphere
VIIRS	Visible Infrared Imaging Radiometer Suite
VIS	Visible
WBS	Work Breakdown Structure
WCT	Several SLSTR SST products that are disseminated to users (Basis for WST).
WST	SLSTR SST product that is disseminated to users

1 Introduction

Fifteen IST algorithms have been evaluated along with several cloud-screening products [AD-8]. The evaluation criteria used were the mean monthly performance in terms of the standard deviation and bias over one year, from August 2016 to July 2017 and the performance stability through time. From the previous evaluation report [AD-8], two IST algorithms were selected for implementation in the SLSTR IST prototype processor. The first is a split window algorithm using nadir view data only (IST2) and the second is a dual view single channel algorithm, (IST12).

Also from AD-8 a composite cloud screening strategy was chosen, namely to use the Liberti cloud mask during sunlit hours (Liberti, 2017) and the Basic cloud mask during twilight and night time, i.e. for sun elevation greater than 80 degrees.

This validation report summarises validation results for the two algorithms selected, and the Liberti/Basic composite cloud screening procedure, following the Product Validation plan [AD-2]. The validation activities cover the period from August 2016 to July 2017, and consists of monthly bias and standard deviation (std) of difference between SLSTR and in situ in high latitude regions. Southern Hemisphere (SH) validation data are provided by example, because in-depth SH performance analysis is excluded due to poor coverage of high quality in situ data in that region.

The validation was originally intended to be stratified according to the type of in situ observation data because of the varying quality of these data, e.g. traditional drifting buoy temperature measurements on sea ice are often erroneous, mainly due to snow covered instruments. However since the amount of high-quality validation data was limited, it was decided to focus the validation on surface temperatures from the high quality AWS data set from PROMICE (promice.dk, [AD-2]). An example of IST validation against traditional buoy measurements is presented here to demonstrate that such measurements are inadequate for a complete algorithm performance evaluation.

The evaluation of both SLSTR IST algorithms against a small, high quality, sea ice in situ data set showed excellent performance. Although this data set cannot provide thorough and stratified validation results, it confirms the usefulness of the SLSTR IST product for sea ice monitoring. The focus on land AWS data does, of course, limit the representativeness of this validation, particularly in the Marginal Ice Zone (MIZ), where we have no performance estimates. The validation also includes comparisons against other satellite data and Numerical Weather Prediction model data, to evaluate the SLSTR against current IST products and to illustrate differences between model and observed temperatures.

A comparison between IST derived from SLSTR-A and SLSTR-B data is also carried out, based on 1 month of data. This relative comparison was performed at closed sea ice locations across the Arctic Ocean, with no ground observations as absolute reference.

Theoretical uncertainty estimates are compared with actual performance, i.e. the calculated standard deviations are compared against theoretical uncertainties.

The central elements of the report are algorithm description (Chapter 2), a description of the validation data (Chapter 4-5), a review of the IST requirements (Chapter 7), and results in Chapter 8. Conclusions drawn from this report are given in Chapter 9 and suggestions for future works in Chapter 10.

2 SLSTR IST Algorithms

The algorithms selected for validation (IST2 and IST12) can be written as (see also ATBD, [AD-4.2]):

$$IST2 = a_0 + a_1 Tb_{11nadir} + a_2 Tb_{12nadir} + a_3\left((Tb_{11nadir} - Tb_{12nadir})(sec\,\theta - 1)\right)$$

$$IST12 = a_0 + a_1 Tb_{11nadir} + a_2 Tb_{11oblique}$$

All a_x are calibration coefficients, Tb is the satellite brightness temperature and the subscripts 11 and 12 represent the centre wavelength of the given satellite channels (in microns) nadir and oblique represent the sensor view. Sec θ is the secant of the satellite view angle.

IST2 is a split window algorithm based on nadir view data only. This algorithm is identical to the current operational algorithm (OSI-205) implemented by the OSI SAF with the main difference being that IST2 uses one set of coefficients for the full temperature range, whereas OSI-205 uses three sets of coefficients, representing warm, middle and cold temperature intervals. Validation results from PVRv1 revealed no performance improvements from using distinct temperature ranges

IST12 is a dual view algorithm using a single channel. It was selected because of its good general performance, the assumed small sensitivity to variation in snow properties (Bamber and Harris, 1994), as well as superior performance when applied to a small, high quality, sea ice data set (see Results below). These qualities are essential applied to a wide range of snow and ice properties across the Arctic and Antarctic sea ice covered areas.

3 Validation metrics and definitions

The following metrics are used to assess the performance of the algorithms (AD-2):

Discrepancy:	The difference between the result and the validation value.
Bias:	The mean value of the discrepancy.
Standard deviation of differences (std):	The standard deviation of differences between the satellite and reference.

The following definitions are used throughout the document:

Ice Surface Temperature (IST): The temperature measured by an infrared radiometer that represents the skin of the snow and sea ice temperature. In the subsequent validation, IST is referring to land ice surface temperature, because of the focus on PROMICE AWS data from the Greenland ice cap, except where sea ice is mentioned explicitly. The focus on ice cap based in situ observations is unfortunate, because the goal here is to evaluate sea ice surface temperatures. However, the choice was necessary as only this data set has the required quality and volume for the validation carried out here and in PVRv1 [AD-8].

Validation: The process of assessing, by independent means, the quality of the data products (the results) derived from the system outputs.

3.1 Header labels

Most validation/evaluation plots below are associated with a code-like header that describes filters and other conditions applied to the data shown on a given plot. This information is explained in the figure captions, but not always in detail. For reference, the header syntax is described in detail here:

Example:

IST12 2017-03-01 to 2017-03-31. Post/pre filter: 128/286. STD: 2.258427 BIAS: -2.421128 ql 4-5 dist_2000_time_1800_tcwv_400_promcloudx_10.5_ecmwfcloudx_10.5_liberticloudx_10.5_BayesCloudx_2 _basiccl_1_tdiffnwp_100.0_sunzenn_0.0_sunzenx_180.0_satzen_0.0_satzenx_165.0. stationfilter: ist_aws_istP*/U/P

IST12	Algorithm
2017-03-01 to 2017-03-31	Start period End period
Post/pre filter: 128/286	Number of data post and pre filtering
STD:2.258427 BIAS: -2.421128	Std and bias
ql 4-5	Quality levels applied
Dist_2000	Maximum distance in meters between satellite centre and in situ
Time_1800	Maximum time lag in seconds between satellite centre and in situ
Tcwv_400	Maximum Total column water vapour (kg/m2)
Promcloudx_10.5	Maximum Promice cloud area index [0:1] larger than 1 means no filter
Ecmwfcloudx_10.5	Maximum ECMWF NWP cloud area index [0:1] larger than 1 means no filter
Liberticloudx_10.5	Maximum Liberti cloud probability [0:1] larger than 1 means no filter
BayesCloudx_2	Bayes cloud maximum value. 0 is cloud free, 2 means both cloud and no cloud data.
Basiccl_1	Basic cloud maximum value. 0 is cloud free, 1 means both cloud and no cloud data.
Tdiffnwp_100.0	Maximum absolute temperature difference to NWP skin temperature
Sunzenx_180.0	Sun-Zenith minimum angle and Sun-Zenith maximum angle
Satzen_0.0_satzenx_165.0	Sat-Zenith minimum angle and Sat-Zenith maximum angle
Stationfilter: ist_aws_istP*/U/P	Station filters (string match): current example includes all “Upper” AWS data from PROMICE plus EASTGRIP.

4 In Situ observations

Fiducial Reference Measurements (FRM) – see https://earth.esa.int/web/sppa/activities/frm) - are defined as independent, fully characterized, and traceable ground measurements that follow the guidelines outlined by the GEO/CEOS Quality Assurance framework for Earth Observation (QA4EO, [AD-6]). FRM delivers the required confidence in data products, in the form of independent validation results and estimates of uncertainty in the satellite measurements, for entire duration of a satellite mission.

The IST in situ observations acquired for this project come from various operational data streams and research projects. An automatic quality control procedure has been developed specifically for IST applications within the ESA FRM4STS project (Høyer et al., 2018). This has since been further developed and implemented by the OSI SAF. The quality control (QC) procedure includes 16 quality tests, with sanity, self-consistency and checks against other in situ platforms in the vicinity.

The observation types within the in situ observation database are:

- **Conventional air/ice temperatures from GTS drifters (DMI and ECMWF)** covering global sea ice**.**
- **Thermometric surface temperatures from IMB (SAMS, BAS, CRREL)** covering Arctic Ocean, Fram Strait and few from the Southern Ocean.

- **Radiometric surface temperatures** from ship and aerial campaigns (IceBridge, DMI/LOMROG, POLARSTERN/AWI, Tara and MET Norway, 4 winters of DMI AWS deployments on the fiord of Qaanaaq, NW Greenland).
- **Blended Surface and air temperature observations** from various scientific campaigns, for example, SST observations from drifting buoys and Argo floats, from the Coriolis archive.
- **Polar Automatic Weather Stations** (AWS), some only recording 2 m temperatures, others measuring radiation balance, including estimated skin temperatures and fractional cloud cover.

Note that not all types are available within the agreed SLSTR IST validation period. Further information, access and links to the data sources, can be obtained from the OSI SAF team that operates the IST data archive and conversion process (https://osisaf-hl.met.no/).

All in situ temperature observations have been processed into a common NetCDF format. Along with the data processing, an inventory list is maintained with all relevant information about the in situ observations, such as data policy, timeliness, data distribution, spatial and temporal sampling. The spatial and temporal distributions of the in situ data are show in Figure 1, where each 5th data point is plotted on a map (left column). The monthly data frequency is shown in histogram (right column) for Northern and Southern hemispheres, top and bottom row respectively.

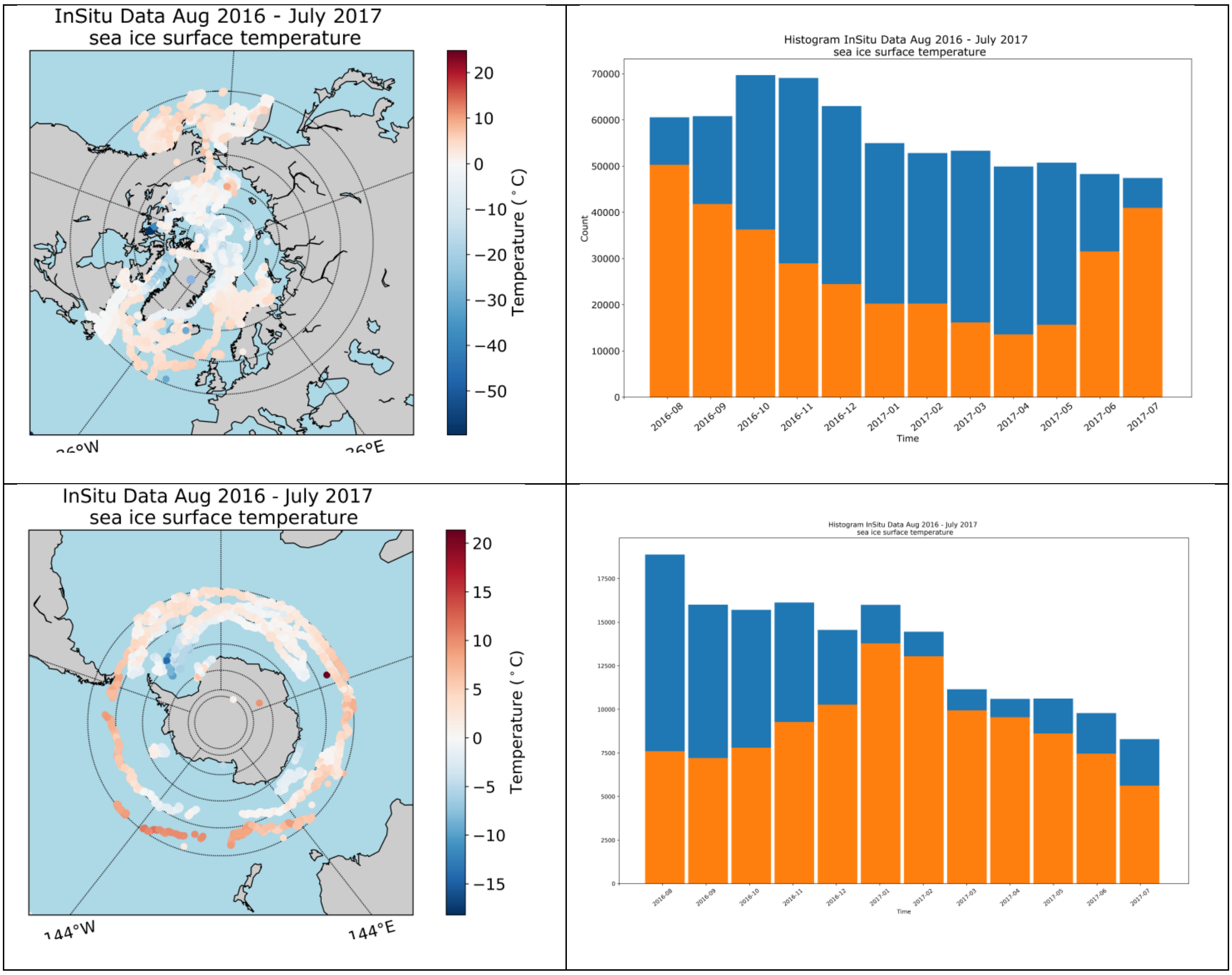


Figure 1 *The spatial and temporal coverage of the applied in situ observations for the period August 1st 2016 to July 31st 2017. Top row is spatial distribution of NH data (left) and monthly distribution (right), where blue columns are temperature data lower than 0 C and orange columns are observations warmer than 0 C. Bottom row is the corresponding data distribution for SH.*

Unfortunately, from this work and earlier IST validation studies with e.g. the half year reporting of the OSISAF IST product performance, it has become clear that traditional drifting buoy measurements, from e.g. isvp buoys, do not have the sufficient quality for validation and tuning of satellite IST products. Below (in Figure 2) it is shown by example that traditional buoy observation are generally too warm . We have concluded that this warm bias is caused by the fact that traditional buoy sensors after deployment most often will be covered by snow. The SLSTR IST project team has documented this from field campaigns. Consequently, present study has rejected vast parts of the match up data that was compiled for the study and almost entirely rely on ice cap observations from PROMICE weather station from the Greenland ice sheet.

The PROMICE data offer reliable observations of both surface and air temperatures in FRM quality. Furthermore, the PROMICE data are available from a suitable number of sites from homogeneous surfaces at a high sampling rate that enables for retrieving robust validations of IST. However, the PROMICE stations cannot be used to evaluate satellite IST products on sea ice with the special challenges of these environments, like changing ice concentrations and different atmospheric compositions. It is crucial for the IST production and user communities to develop means to retrieve fiducial surface temperature from the sea ice in the future. The project team has written some considerations and recommendations regarding future sea ice observation platforms that can comply with FRM recommendations (see section 10.1).

5 Match-Up Data Base

For this validation, a match-up database (MUDB) was generated, including the satellite brightness temperature observations matched up against in situ observations. The spatial and temporal criteria for matching an in situ observation with a satellite observation are 5 km and 3 hours, respectively.

Each Match-Up (MU) includes the surrounding 401x401 SLSTR pixels around SHIP observations and temporally averaged IceBridge data (averaged to 30-second observations). For drifting buoy observations, other drifting platforms and AWS locations, the corresponding surrounding SLSTR pixel matrix is 21x21 pixels.

From the MUDB files, the following information is available:

- In situ observations from Automatic Weather Stations (AWS), Ice Mass Balance Buoys (IMBs), drifting buoys, DMI in situ db and Operation Ice Bridge:
    - Temperature (Air temperature (usually T2m) and Surface Temperature, based on up- and downward long wave radiation, from radiometric sensors at e.g. PROMICE stations.
    - Wind speed
    - Humidity
    - Radiation (in/out, Long/Shortwave)
    - Cloud cover calculated at each PROMICE station, based on in situ measured downward long wave radiation and air temperature.
- SST observations from the Coriolis data archive pole ward of 60° South and 60° North.
- SLSTR Level 1 data (all RBT) and L2 (WCT and WST) either NTC or reprocessed data. Both Nadir view and oblique view data are included.
    - SLSTR channels (VIS and IR)
    - Native SLSTR L1 cloud parameters from basic and Bayesian cloud masks
    - any relevant noise/quality data that are available and relevant
    - sun, satellite and view geometry
    - SSTs from WST (including quality indicators and sun-satellite-view geometry)
    - NWP (25 layers from SLSTR data stream) for RTTOV processing

- SLSTR IST recommended from the requirement baseline document and ATBD v1 [AD-4] will subsequently be calculated
- Surface and air temperatures, TCWV, Cloud Area Fraction from the NWP date in the level 1 data stream

- Cloud masks based on the official SLSTR cloud-over-ice ATBD [Liberti, G.L, 2017] is calculated from MUDB variables and added.
- Alternative cloud-screening procedure developed at University of Leicester, based on SLSTR data and simulated TOA TBs (see section 5.5.6) [Ghent and Sembhi, 2017]
- Daily sea-ice concentration (Operational OSI SAF product OSI-401-b)
- Metop IST (operational OSI SAF product OSI-205)
    - IST ("surface_temperature")
    - IST uncertainty (3 variables)
    - L2p_flags (GHRSSTvariable)
    - Processing_flags
    - Quality_level
    - Probability_of water
    - Probability_of_ice
    - View and sun geometry.

6 IST Validation

An essential part of infrared surface temperature monitoring from space is cloud screening. Several cloud screening products was tested and evaluated [AD-8]. Cloud products are included in the MUDB or calculated from variables in the MUDB. The tested cloud products were the native Basic and Bayesian/Probabilistic cloud masks based on both nadir and oblique view channels. The Bayesian mask is provided for water pixels and the Probabilistic mask is provided for land pixels, with the Basic mask provided for both. The Basic cloud mask is a bit mask with 14 tests and will indicate cloud if any of the 14 tests fail. The probabilistic Liberti sea-ice cloud mask is calculated from MUDB variables. Two Cloud Area Fraction products (CAF) were tested and used as reference, namely the ECMWF CAF from the NWP data stream in level 1 and the PROMICE CAF. The PROMICE CAF is used as a cloud reference, because it is considered as the best available measure of the real cloud cover. However, PROMICE CAF is not perfect and results from comparing satellite cloud masks with the CAF can only be used as indications.

Based on PVRv1 results [AD-8], it was decided that cloud screening shall be performed by a composite algorithm, using the binary Basic cloud mask for sun elevations lower than 80 degrees and Liberti cloud mask (cloud if probability-of-cloud > 50%) for sun elevation higher than or equal to 80 degrees. The Nadir-view-only IST algorithm, IST2, uses Basic-nadir and Liberti-nadir cloud masks and the dual view IST algorithm, IST12, uses both Basic-nadir and Liberti-nadir and Basic-oblique and Liberti-oblique cloud masks, i.e. data from IST12 are considered cloud contaminated if one of the nadir or oblique cloud masks are set.

Furthermore, an evaluation of the UoL cloud mask is also included and compared with the combined Liberti/Basic cloud mask. The UoL cloud mask is a refinement to the operational Probabilistic mask where the land cloud mask ADF has been refined for high latitudes.

6.1 Validation activities

The validation activities in this document include the following tasks:

- Evaluation of cloud screening (Basic and Liberti).

- SLSTR IST validation against in situ observations using the chosen cloud mask for day time, night time and twilight, respectively.
- Comparison of SLSTR IST with NWP and Metop IST products.
- Inter-comparison of IST retrieval estimates from SLSTR-A and SLSTR-B.
- Evaluation of the SLSTR total uncertainty product.
- Evaluation of the SLSTR IST quality level algorithm.
- Evaluation of the Leicester cloud mask (Ghent and Sembhi, 2017)

An overview of planned validation activities is available in the Product Validation Plan [AD-2].

7 IST product performance Requirements

The requirements for IST products from a range of stakeholders and users are discussed in the Requirement Baseline document [AD-1]. There is general agreement regarding the primary requirements for the target, breakthrough and threshold precision of 1 K, 1.5 K and 2 K, respectively, for daily or bi-daily products. The requirements for spatial resolution vary from 1-5 km to 100 km or even more, for daily products. Several stakeholders require sub daily temporal resolution to resolve diurnal variability. Other stakeholders require a seamless and stable temperature product, across various product domains, e.g. sun elevation, cloud screening procedures, season and sea ice concentrations.

The std and bias of IST retrievals relative to ground observations must comply with the product requirements, but the IST performance must be corrected for the uncertainties related to the ground measurement themselves in order to evaluate whether the requirements are met or not. Hence, performance of an IST retrieval cannot be better than the uncertainty of the in situ observation (see Section 8.2.9).

8 Validation Results

Early investigations in the preparations for PVRv1 indicated unreliable in situ data from most sea ice platforms, in particular traditional drifting buoys. In Figure 2, the bias and std for the validation of SLSTR IST against traditional drifters shows extremely poor performance, which is not suited for evaluating algorithm performances. Consequently, it was necessary to change the focus of the validation towards high quality land ice based observations from PROMICE even with the limitations this brings. A consequence is that some planned validation activities do no longer make sense or cannot be implemented, e.g. performance variability with total column water vapour makes less sense due to extreme dry atmosphere on the Greenland ice sheet and performance dependency on sea ice concentration can naturally not be carried out (see also Chapter 4).

Results presented in this chapter follow the validation activities outlined in Chapter 6, starting with evaluation of the cloud screening procedure.

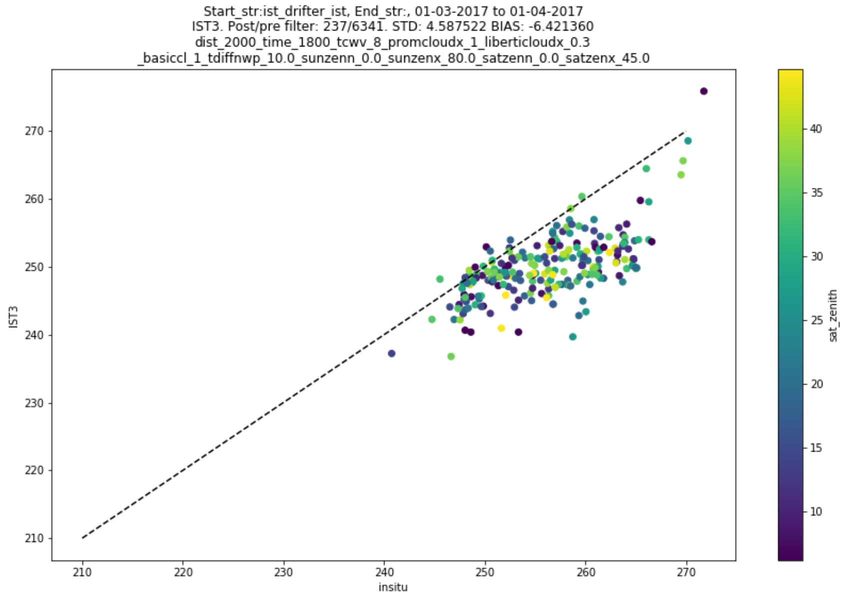


Figure 2 *Scatterplot of IST3 versus traditional sea ice drifter temperature data (day time, March 2017) [AD-8]. IST3 is conceptually identical with IST2 and the performance of IST2 and IST12 are qualitatively comparable with IST3 against current buoy observations. IST3 is used here as an example of inadequate in situ observation quality for IST retrieval evaluation.*

8.1 Cloud mask evaluation and inter-comparison

It is commonly known that undetected clouds are the single largest error for the satellite IST performance [Dybkjaer et al., 2012], [Hall et al., 2012]. Cloud screening over ice and snow is more complicated than over ocean because of the spectral and structural similarities between ice and snow covered surfaces and cold cloud tops.

Cloud screening procedures are, however, not the focus of this current activity and only a few basic characteristics of the applied cloud masks are shown, including a comparison with the PROMICE CAF, which is used as reference data. An assessment of the different combinations of cloud mask and IST retrieval algorithms is reported in PVRv1 [AD-8]. The conclusion of the analysis was a recommendation to use Liberti probabilistic cloud mask during sunlit hours and the Basic cloud mask during twilight and night time. Overviews of the two cloud screening procedures is presented in the following sections. In addition, the recommended cloud screening combination is also evaluated against the UoL cloud mask.

8.1.1 Statistical performance

The statistical performance and distribution of the cloud masks is an important indicator of their ability to identify clouds. As reference for the 'true' cloud/no cloud distribution we use the PROMICE CAF product. The CAF is expected to provide realistic cloud/no cloud distribution, because PROMICE CAF is based on in situ radiation measurements. We do not expect each CAF value to be correct but we assume it is a good indicator for cloud coverage.

The histograms in Figure 3 and Figure 4 illustrate the cloud/no cloud distribution for the applied cloud screening techniques during day and night and for September 2016 and March 2017, respectively. The general distribution of the CAFs is approximately 50 % cloudy and 50 % not cloudy, using 50 % CAF as cloud/no-cloud threshold for both night and day and for both September and March.

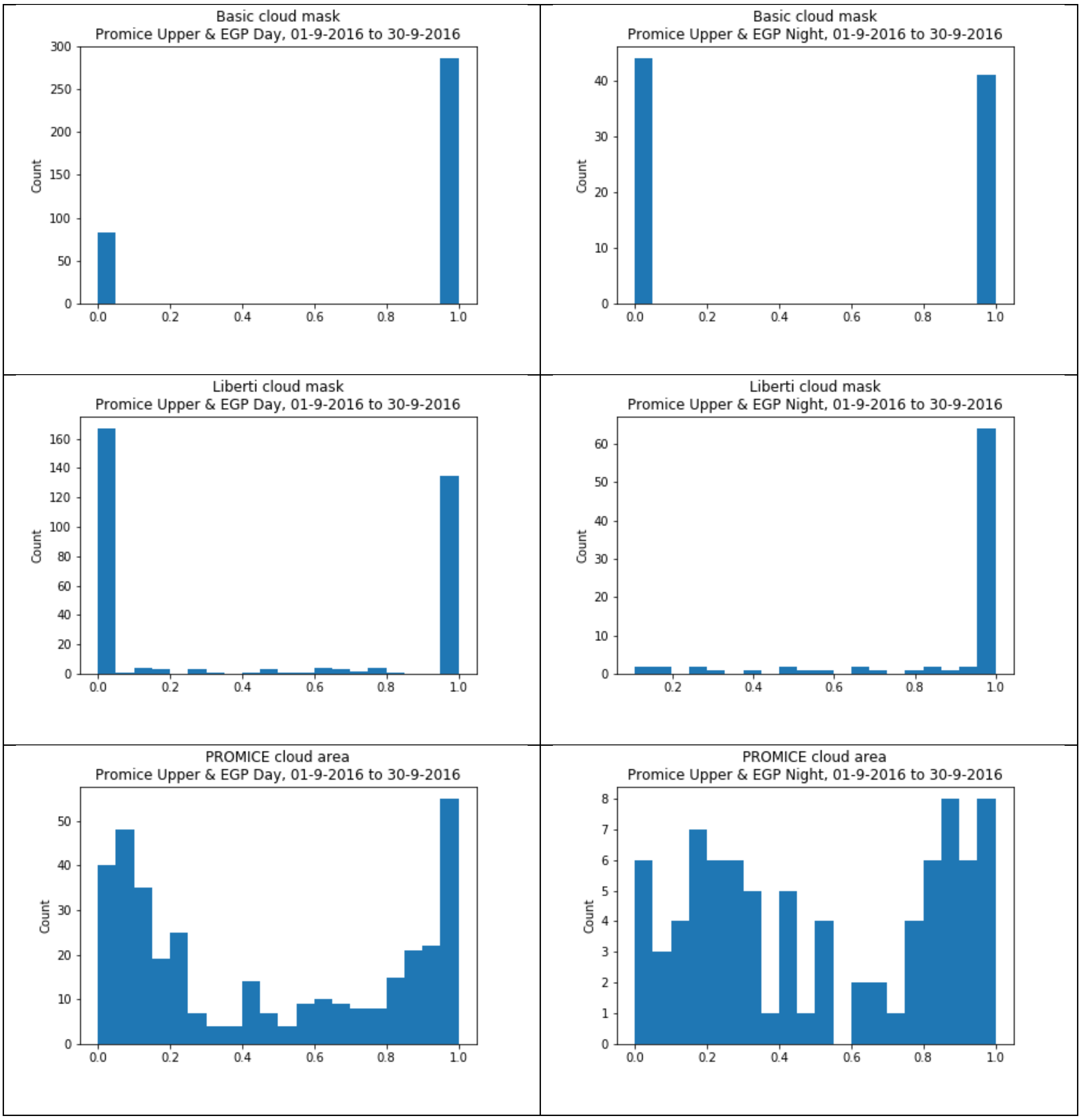


Figure 3 *Histograms of applied cloud screening techniques for September 2016. Day time (left column) and night time (right column). Cloud screening procedures are, from top row, Basic, Liberti and PROMICE CAF. The primary axis is the fraction (Promice and NWP)/probalbility (Liberti and Basic) of cloud.*

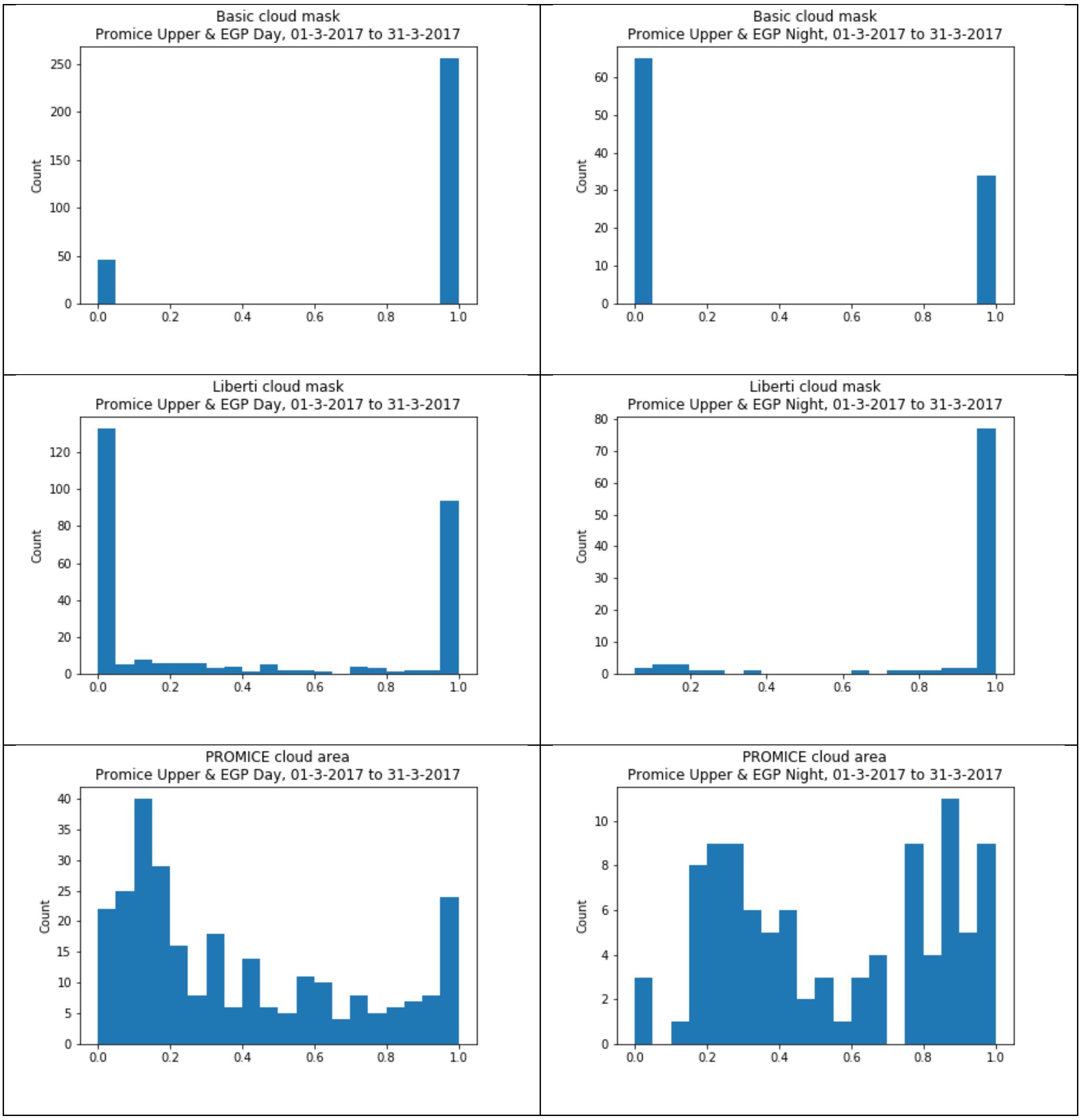


Figure 4 *Histograms of five cloud screening means for March 2017. Day time (left column) and night time (right column). Cloud screening procedures are, from top row, Basic, Liberti and PROMICE CAF. The primary axis is the fraction (Promice and NWP)/probalbility (Liberti and Basic) of cloud.*

As seen in the above scatterplots the Basic day time and Liberti night time performance is not as good during March or September.

8.1.2 Cloud classification skills

To quantify the cloud screening performances we have looked into the number of clouds correctly classified and those that were missed. Contingency tables with PROMICE CAF as reference data, along with the results from the Basic and Liberti masks are shown in the tables below, for September 2016 and March 2017. The PROMICE CAF products stratify cloud free and cloudy pixels using a threshold of 30 %.

The night time performance of the Basic cloud mask indicates 74 % correctly classified and 20 % misses in September, and 76 % correctly classified and 40 % misses in March. This is shown in contingency table (Table 1). The Basic day time performance is generally poor with many falsely detected clouds.

The Liberti day time performance has 88 % correctly classified in September and 76 % correctly classified in March (Table 2).

Table 1 *Contingency table for Basic cloud mask versus PROMICE CAF (threshold 30 %). Day time (top) and night-time (bottom) September 2016 (left) and March 2017 (right). Values are fractions of N.*

01-9-2016 to 30-9-2016

Promice Upper & EGP Day N.364	Basic No Cloud	Basic Cloud
Promice No Cloud	0.203297	0.274725
Promice Cloud (>30%)	0.0164835	0.505495

01-3-2017 to 31-3-2017

Promice Upper & EGP Day N.272	Basic No Cloud	Basic Cloud
Promice No Cloud	0.0992647	0.415441
Promice Cloud (>30%)	0.0220588	0.463235

01-9-2016 to 30-9-2016

Promice Upper & EGP Night N.85	Basic No Cloud	Basic Cloud
Promice No Cloud	0.317647	0.0588235
Promice Cloud (>30%)	0.2	0.423529

01-3-2017 to 31-3-2017

Promice Upper & EGP Night N.98	Basic No Cloud	Basic Cloud
Promice No Cloud	0.255102	0.0510204
Promice Cloud (>30%)	0.397959	0.295918

Table 2 *Contingency table for Liberti cloud mask versus PROMICE CAF (threshold 30 %). Day time (top) and night-time (bottom) September 2016 (left) and March 2017 (right), for day (top) and night (bottom). Values are fractions of N.*

01-9-2016 to 30-9-2016

Promice Upper & EGP Day N.330	Liberti No Cloud	Liberti Cloud (>30%)
Promice No Cloud	0.427273	0.0181818
Promice Cloud (>30%)	0.10303	0.451515

01-3-2017 to 31-3-2017

Promice Upper & EGP Day N.258	Liberti No Cloud	Liberti Cloud (>30%)
Promice No Cloud	0.418605	0.108527
Promice Cloud (>30%)	0.135659	0.337209

01-9-2016 to 30-9-2016

Promice Upper & EGP Night N.85	Liberti No Cloud	Liberti Cloud (>30%)
Promice No Cloud	0	0.376471
Promice Cloud (>30%)	0.0823529	0.541176

01-3-2017 to 31-3-2017

Promice Upper & EGP Night N.95	Liberti No Cloud	Liberti Cloud (>30%)
Promice No Cloud	0.0105263	0.294737
Promice Cloud (>30%)	0.0947368	0.6

8.1.3 University of Leicester – cloud mask performance

An evaluation of the updated Probabilistic cloud mask developed at the University of Leicester (UoL cloud mask) is included for potential replacement of the current cloud mask in a future development of the SLSTR IST processor. The results are based on in situ observations from UPPER PROMICE stations and EastGrip, because, as noted previously, these data are of fiducial quality and from homogeneous environments. The mask is only applicable to land ice surfaces and is only applicable for single view retrievals (although results to IST12 are still shown).

In situ measurements and SLSTR IST2 and IST12 algorithm retrievals were compared for cases where both cloud masks were available (UoL and Liberti/Basic) and for which the SLSTR IST was not more than 273.15 K. The observations were spatially and temporally matched within 5 km and 40 minutes. For all matchups considered together (Figure 5), UoL masked more SLSTR retrievals and had higher Median Differences (MD) and Robust Standard Deviation (RSD) than Liberti/Basic for both IST algorithms. When the matchups are split by day time and night time retrievals (Figure 6) again the MD and RSD are larger for UoL cloud masked data, but UoL is more successful at masking outliers. This is noticeable for cold outliers in night time data, which are likely to be unmasked cloud given the magnitude of their deviation from in situ data. Furthermore, UoL does not seem to mask daytime data to the same degree as Liberti/Basic, which seems to over-mask colder daytime values. It is also worth noting that while some tuning of the UoL cloud mask has been performed, more optimisation may reduce the median differences.

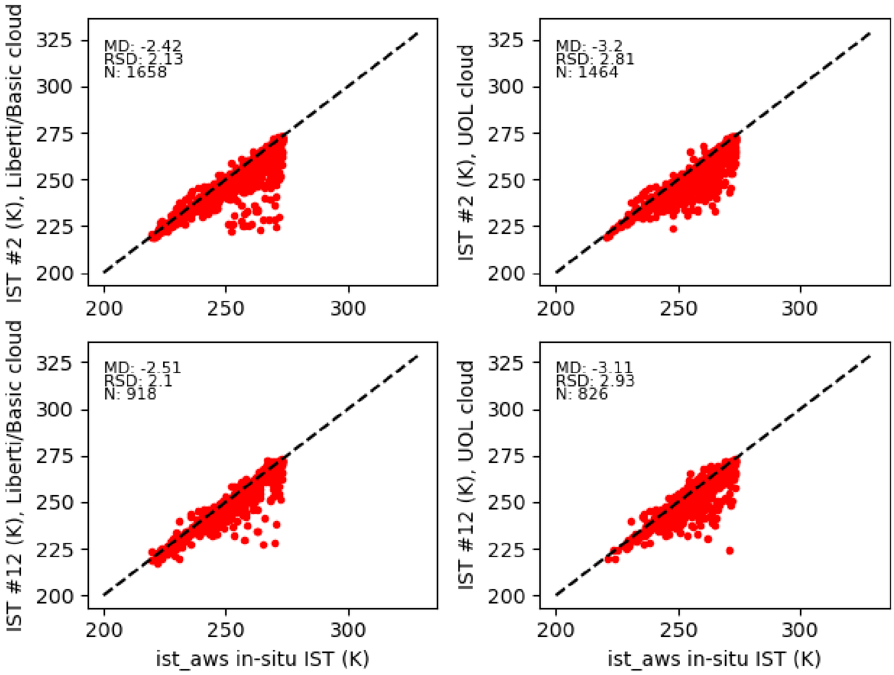


Figure 5 IST2 (top) and IST12 (bottom) SLSTR retrievals masked with the Liberti/Basic (left) and UoL (right) cloud masks compared to PROMICE Upper and EGP between August 2016 and July 2017.

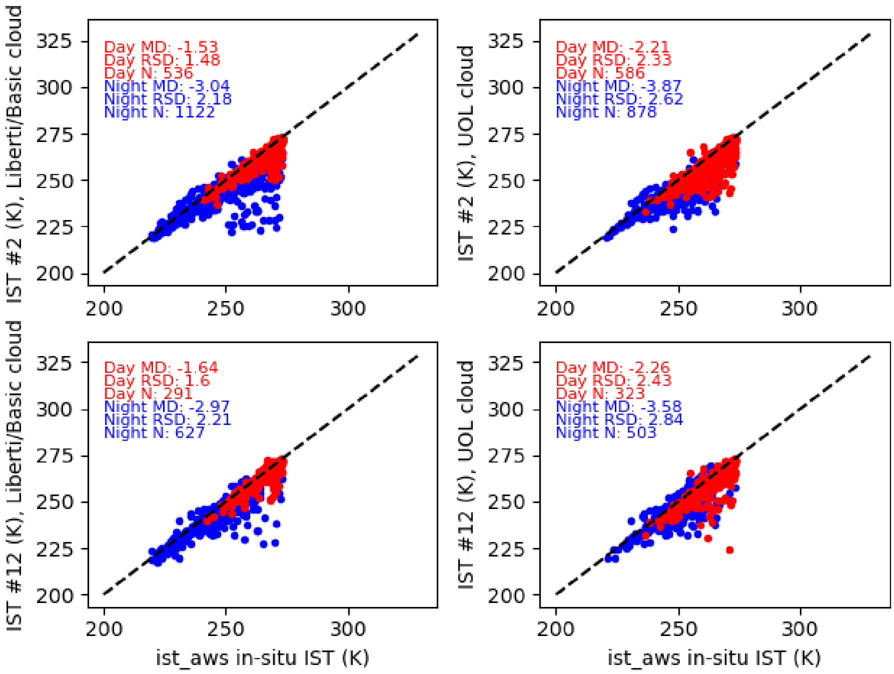


Figure 6 IST2 (top) and IST12 (bottom) SLSTR retrievals masked with the Liberti/Basic (left) and UoL (right) cloud masks compared to PROMICE Upper and EGP between August 2016 and July 2017 and split by illumination condition.

Similar to the Liberti/basic cloud mask (Table 1, Table 2), the UoL cloud masking was compared to PROMICE CAF. For the Liberti/Basic cloud mask compared to PROMICE CAF the agreement is between 55 and 74% for the matchups. Agreement is lower in March 2017 (55-56%) than in September 2016 (70-74%). For UoL compared to PROMICE CAF (**Table *3***), the cloud masks agree for 57-77% of matchups. Agreement is around 60% in March 2017 and the best agreement was for daytime in September 2016. This shows that the UoL cloud mask shows more agreement with PROMICE CAF than Liberti/Basic. The PROMICE CAF is applied as a cloud reference, as it is considered as the best available measure of the real cloud cover. PROMICE CAF is based on in situ radiation measurements, but it is not a perfect cloud mask so these results should be understood in that context.

Table 3 *UOL cloud mask vs PROMICE CAF for September 2016 and March 2017. Values are fractions of N.*

	Promice Upper and EGP Day N 183	UOL No Cloud	UOL Cloud
September 2016	Promice No Cloud	0.30	0.09
	Promice Cloud (>30%)	0.14	0.48
March 2017	Promice No Cloud	0.00	0.00
	Promice Cloud (>30%)	0.00	0.00

	Promice Upper and EGP Night N 7	UOL No Cloud	UOL Cloud
September 2016	Promice No Cloud	0.14	0.14
	Promice Cloud (>30%)	0.29	0.43
March 2017	Promice No Cloud	0.16	0.13
	Promice Cloud (>30%)	0.24	0.47
	Promice Upper and EGP Twilight N 67	UOL No Cloud	UOL Cloud
September 2016	Promice No Cloud	0.16	0.24
	Promice Cloud (>30%)	0.13	0.46
March 2017	Promice No Cloud	0.13	0.03
	Promice Cloud (>30%)	0.31	0.53

There is no clearly efficient cloud screening procedure (Liberti/Basic or UoL) from this initial analysis and further analysis and tuning is recommended. Future work should also consider other cloud screening methods such as the PPS cloud screening software (https://www.nwcsaf.org/).

8.2 SLSTR IST versus Surface temperature observations

The SLSTR IST2 and IST12 algorithms are validated in accordance with the general validation plan [AD-2]. The measures for performance are standard deviation of differences between IST and observations (std) and mean differences (bias), as well as qualitative estimates.

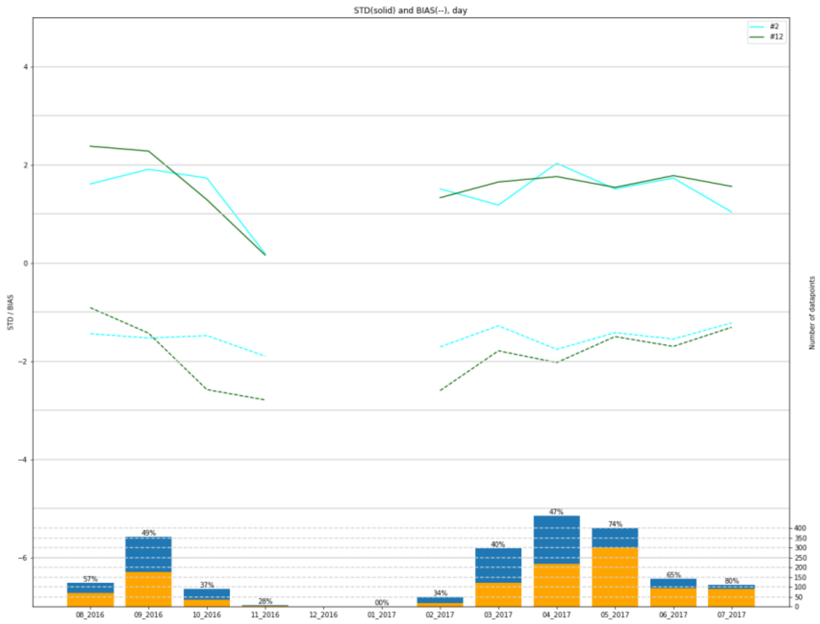
The algorithms are validated and compared using monthly mean performance metrics through 12 months. Cloud screening is done in accordance with cloud mask recommendation above, i.e. Basic cloud mask during night time and twilight and Liberti cloud mask during day time.

Both the Liberti and Basic cloud masks have oblique view (io) and nadir view (in) cloud masks, Liberti_io/in and Basic_io/in, respectively. The nadir view only algorithm, IST2, is cloud masked using Liberti_in and Basic_in cloud masks. The dual view algorithm, IST12, is cloud masked if either or both *_io and *_in cloud masks indicate cloud.

In addition to cloud masking, any IST value more than 10 K different to the nearest NWP analysis IST is removed.

8.2.1 Full year performance of IST2 and IST12

Results of the SLSTR IST algorithms are plotted in Figure 7, divided into day time, night time and twilight from August 2016 to July 2017. Results of all 15 tested algorithms are listed in APPENDIX A, B and C, as monthly values for day time, twilight and night time validation.



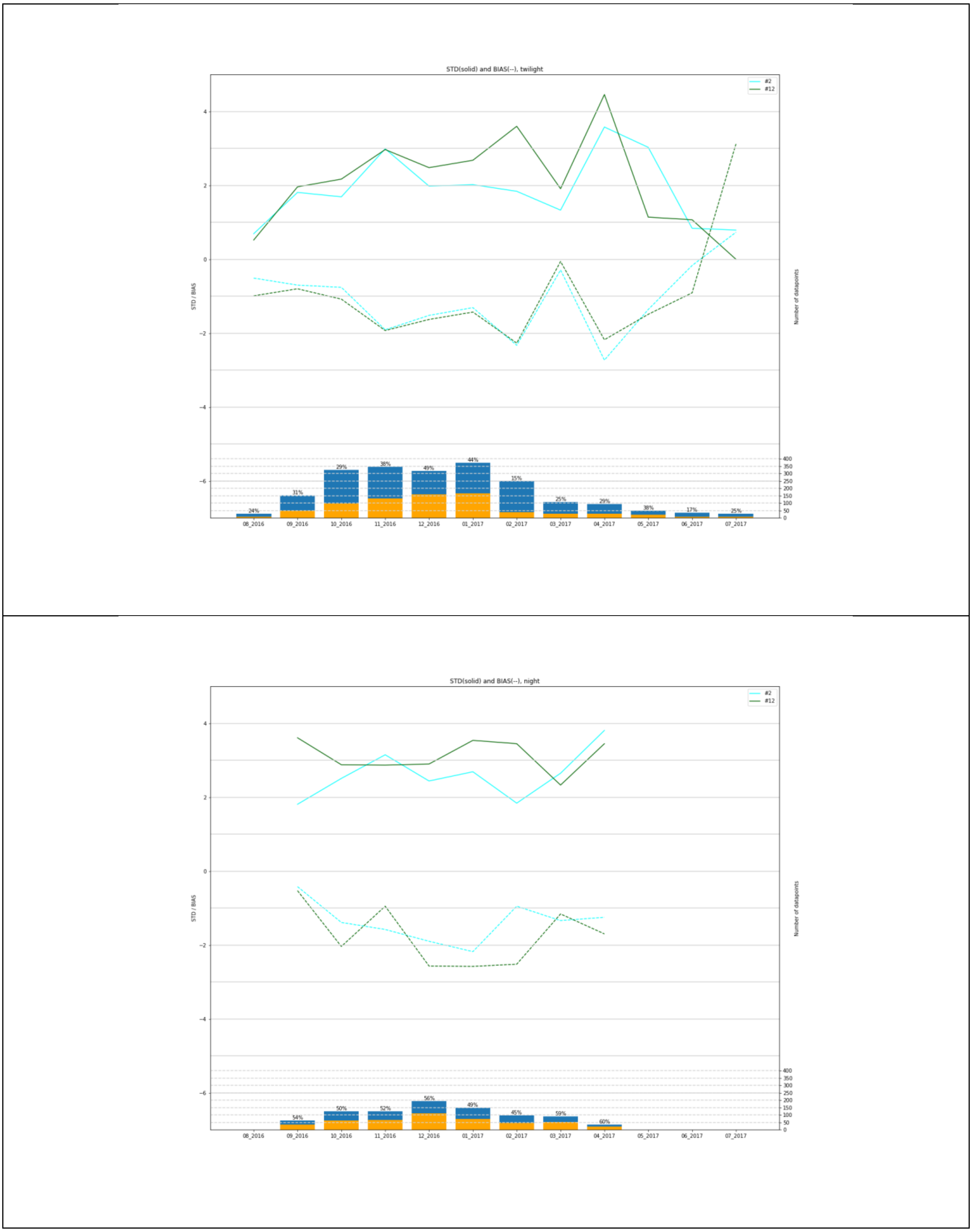


Figure 7 *Performance of IST 2 and IST12 (Solid lines are std and punctured lines are Bias). Day time (top panel), Twilight (middle panel) and night time (bottom panel). Bars indicate the number of data points before and after masking for IST2 (blue and yellow, respectively). Percentage data remaining after cloud masking is written on the top of the bars. The corresponding statistics for cloud screening for IST12 is similar to the IST2 statistics.*

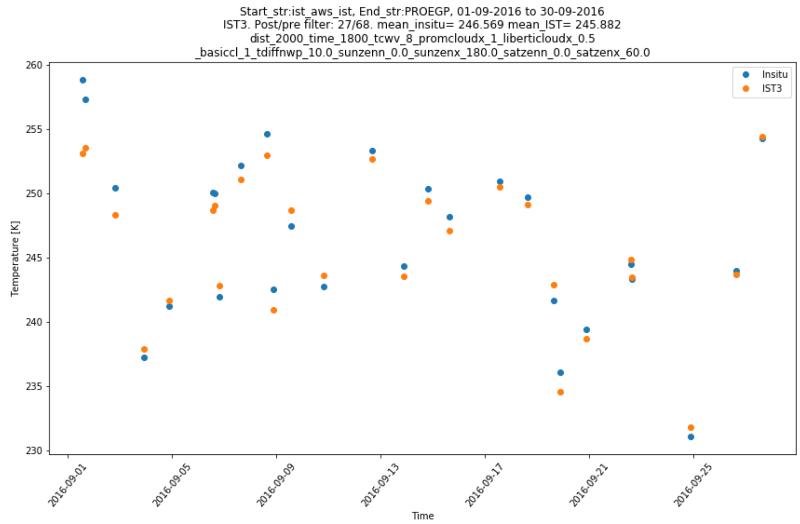
The annual mean statistics for IST2 are approximately, std: 1.4 K, 1.9 K and 2.6 K, for day, twilight and night time, and corresponding bias are -1.5 K, -1.1 K and -1.4 K.

The annual mean statistics for IST12 are approximately, std: 1.6 K, 2.3 K and 3.1 K, for day, twilight and night time, and corresponding biases are -1.9 K, -1.3 K and -1.8 K.

All algorithms seem stabile throughout the year within the three sun-elevation domains (except for months with low match-up numbers). See also APPENDIX A, APPENDIX B and APPENDIX C for the full monthly statistics.

8.2.2 Time series

A qualitative way to look at algorithm performance is to evaluate a time series of match-ups as a time series reveal temporal data gaps and other irregularities like systematic errors. IST3, which has a performance comparable to IST2 [AD-8], is plotted against in situ surface temperatures from the PROMICE EASTGIP data in Figure 8, for September 2016 (top) and March 2017 (bottom). The figure shows an overall bias well below 1 K in both months. Match-ups are well distributed across both months with no clear systematic error, indicating temporal consistency.



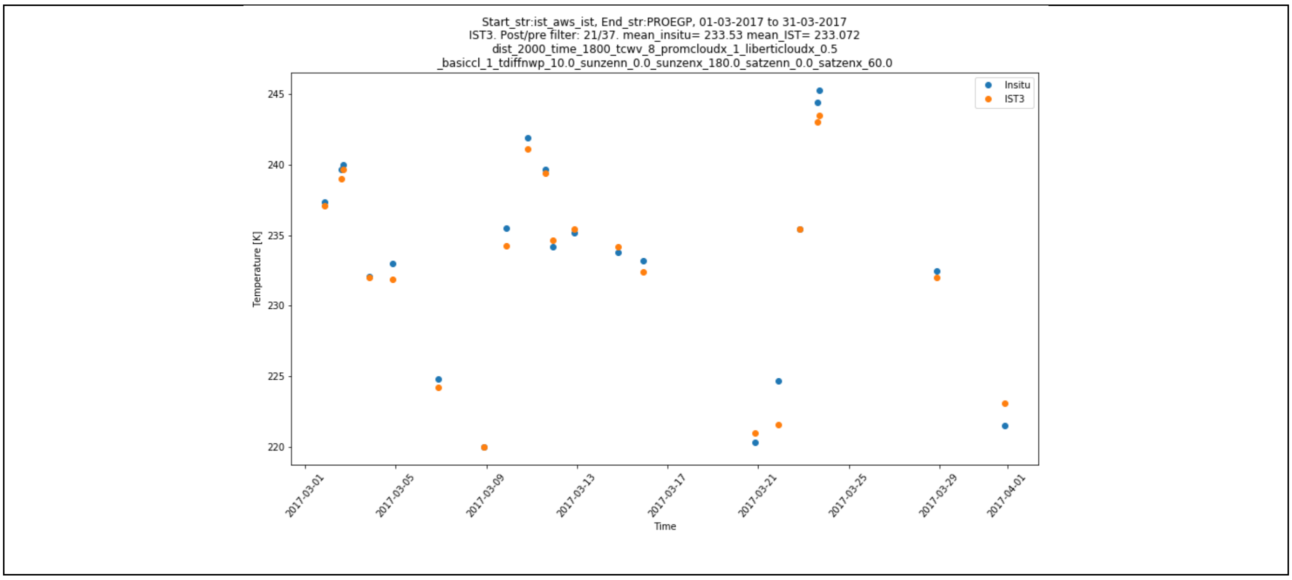
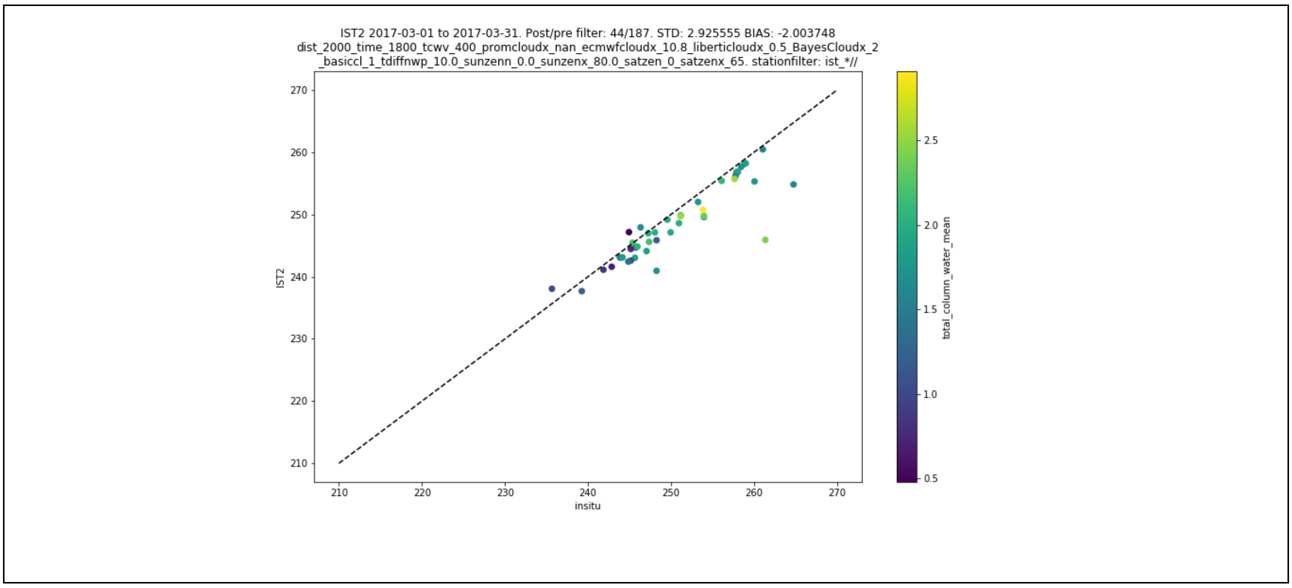


Figure 8 *Time series of IST3 and in situ surface temperatures from EASTGRIP for September 2016 and March 2017. Liberti cloud screening and standard filter is applied. Plots include mainly day time data, due to the application the Liberti cloud mask.*

8.2.3 SLSTR IST versus Southern Hemisphere AWS observations – by example

A comparison of IST2 with air temperatures from four Antarctic weather stations is included as a rough estimate on SLSTR IST performance in the SH (Figure 9). Performance is poorer than performance against PROMICE AWS data on the Greenland ice sheet, but we note the Liberti cloud mask is not tuned for this environment. Furthermore, it is not clear whether the Basic cloud mask is tuned for the Southern Hemisphere. Moreover, and maybe most importantly, the in situ temperature reference is air temperatures in 2 m, which can be several degrees different to the skin temperatures, as documented elsewhere [AD-1]. Antarctic std values range from 3 to 4 K and Bias from -2 to -3 K for day time, twilight and night time.



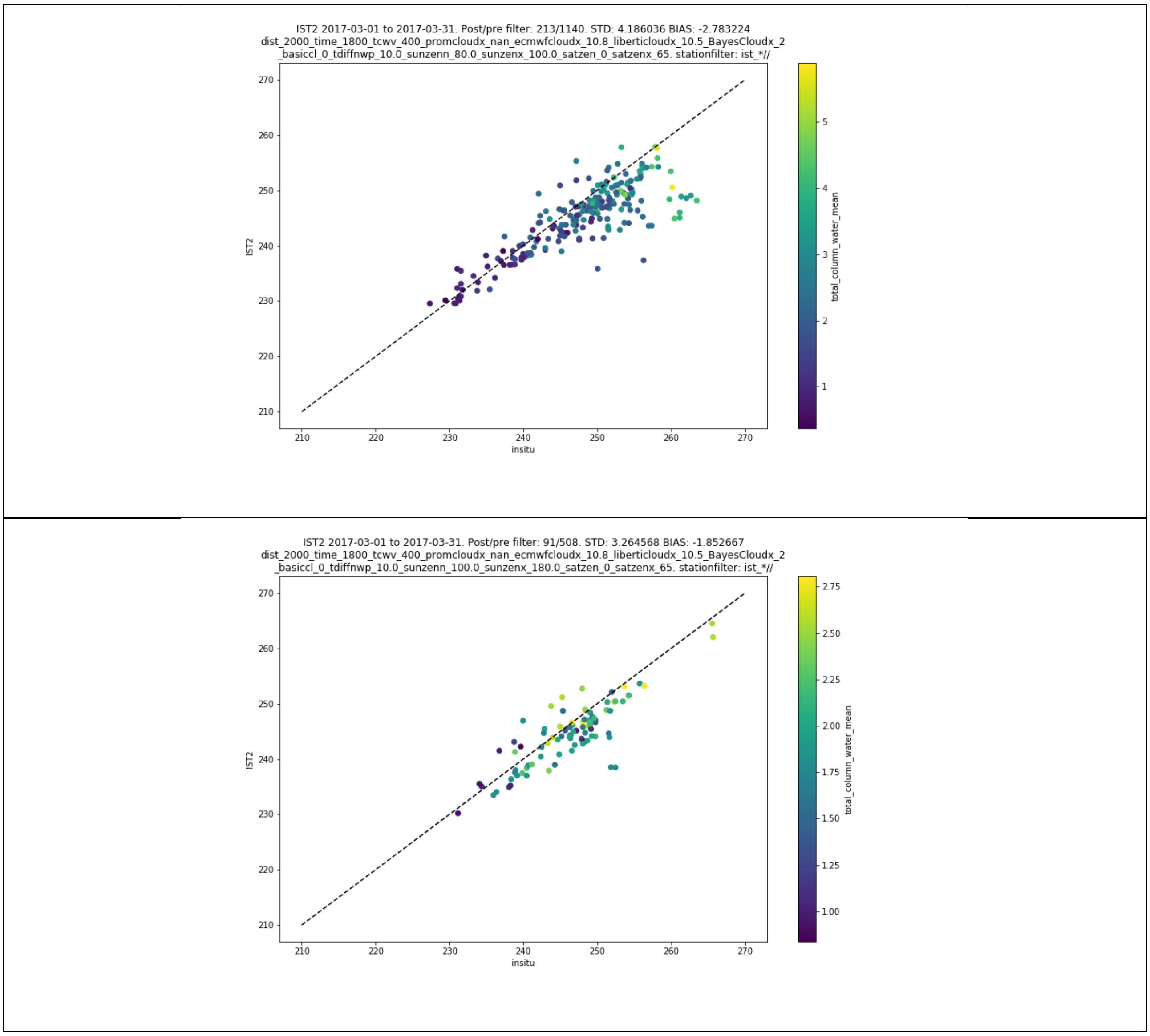


Figure 9 *Scatter plots of IST2 performance against air temperatures from four Antarctic AWS on the ice sheet. Day time (top panel), Twilight (middle panel) and night time (bottom panel)*

It is striking that the data in Figure 9 are scattered on both side of the 1:1 line, whereas in situ data usually are colder than the satellite surface temperature estimate. This can partly be explained by the skin and air temperature differences.

Improvements are needed and cloud masking is most likely the largest limitation of SH applications with this product, but a dedicated Antarctic ice cap retrieval coefficients may contribute to better SH validation statistics.

8.2.4 SLSTR IST versus NH Sea Ice observations – by example

A test of SLSTR IST against radiometric brightness temperature measurements from the DMI sea ice based AWS in Qaanaaq is included as an example of performance against fiducial sea ice measurements. Scatterplots of IST2, IST12, and in situ brightness temperatures are plotted in Figure 10. Both algorithms show clear linear relation with in situ data with a shift towards warmer SLSTR IST values. This well-defined warm shift is caused by the black body radiometric in situ observation that is not corrected for snow emissivity as the satellite IST observation is.

This result is, despite its limited representativeness, extremely promising with respect to general SLSTR IST performance on sea ice, because the sea ice environment is fundamentally different from the environment of the PROMICE AWS stations that are used in the full year performance evaluation.

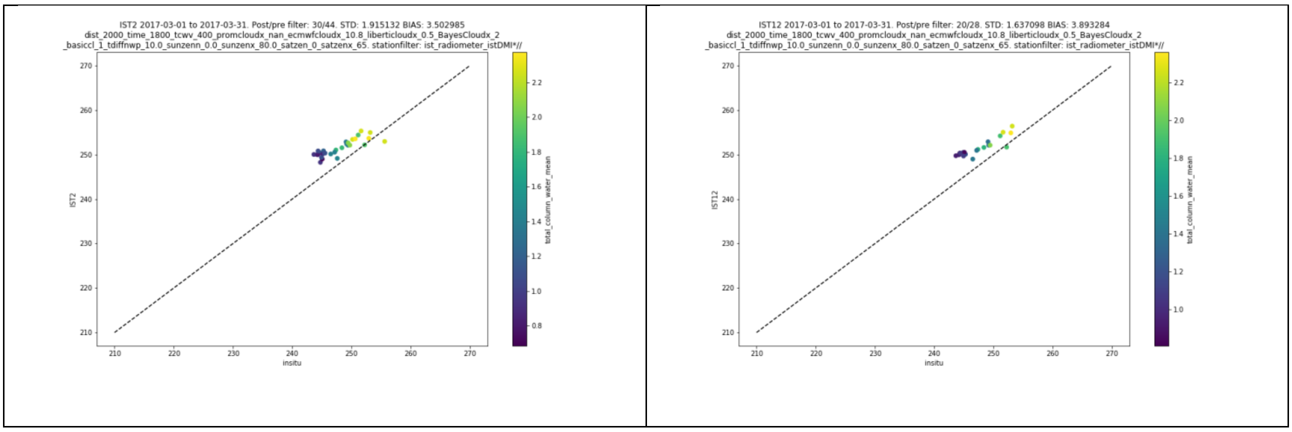
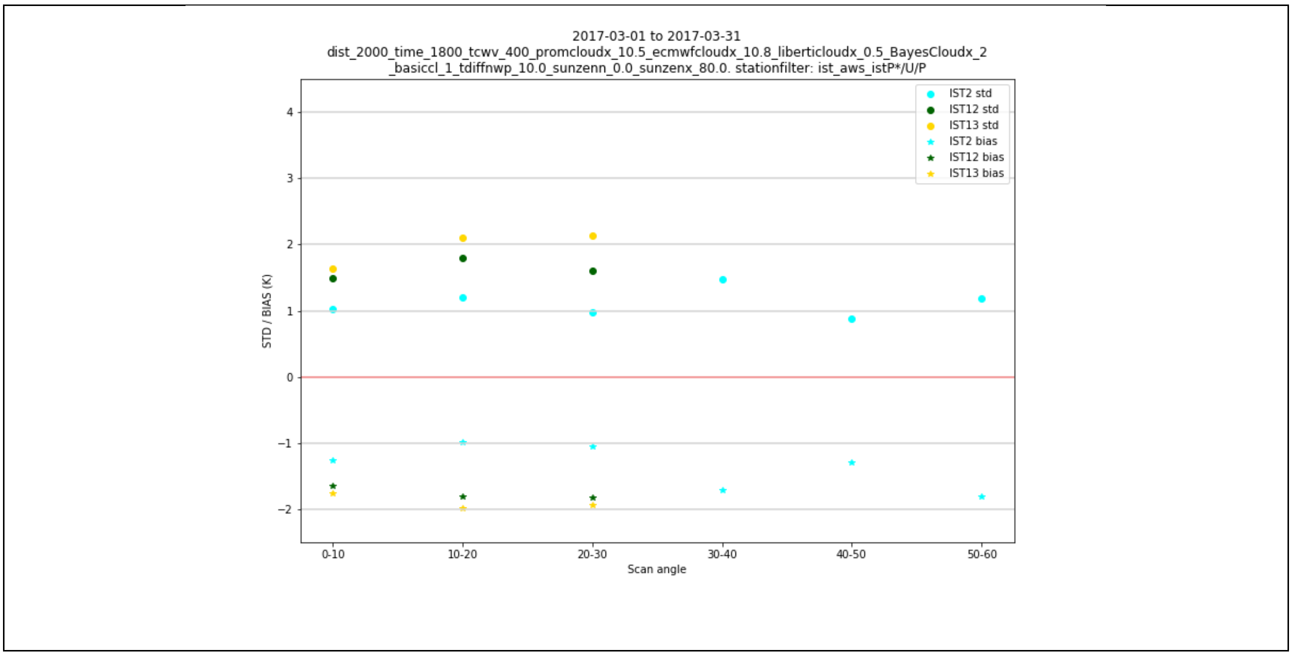


Figure 10 *Scatter plots of IST2 (left) and IST12 (right) against DMI sea ice radiometric thermal infrared brightness temperatures from day time in March 2017.*

8.2.5 SLSTR IST angular dependent performance – by example

The algorithm sensitivity to satellite view/scan angle was evaluated after indications in the first validation report [AD-8] that view angle correction does not necessarily improve the performance of IST12. Here IST2, IST12 and IST13 are tested for view angle sensitivity, against PROMICE Upper and EASTGRIP AWS's. IST13 is essential identical to IST12, but with a view angle correction term [AD-8]. IST13 is included to evaluate the effect of an angular correction term as opposed to IST12 using general scan angle calibration coefficients. The test is carried out on March data for day, twilight and night time in bins of 10 degrees (Figure 11).



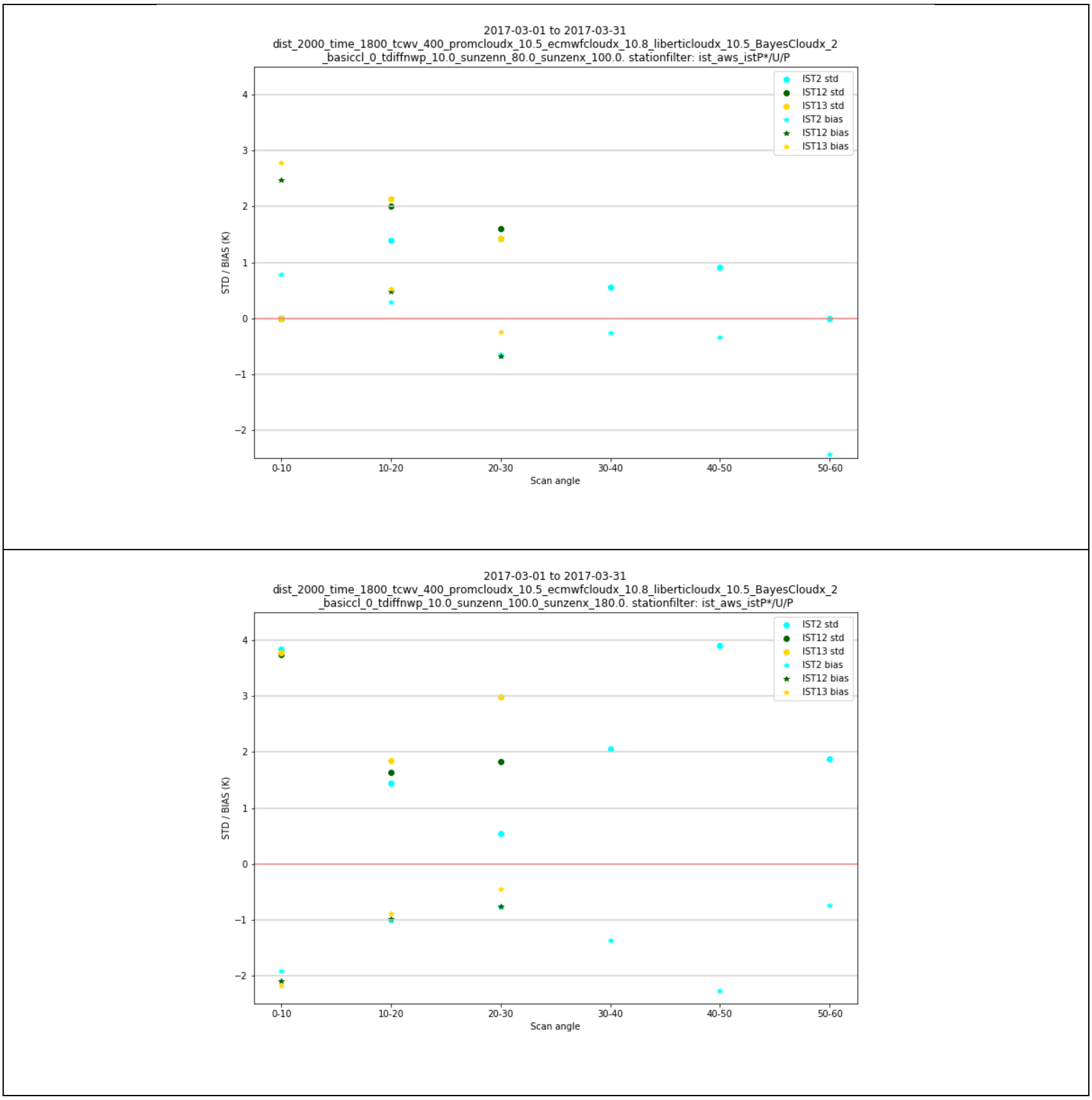


Figure 11 *Validation of IST2, IST12 and IST13 as a function of view angle in bins of 10 degrees, for day time (top), twilight (middle) and night time (bottom). Values scan angle bins 0-10, 40-50 and bin 50-60 for twilight not valid, due to only 1, 2 and 1 data point, respectively. In addition, night time values are not statistically robust either with only 5 to 13 data points.*

Within a tolerance for small variations caused by the limited number of match-ups, the std and bias seem to be independent of view angle during day time for all three algorithms (Figure 11). IST2 is performing best for all scan angles, both with respect to error and bias, and IST13 is performing worst.

In addition, during twilight, there is no clear angular performance dependency for any of the three algorithms shown in Figure 11 (bin 0-10 and 50-60 excluded due to too few data). However, fluctuations in performance across the scan angles seem larger than during day time, but that is assumed a consequence of a limited amount of data.

During night time IST2 performs best at view angles between 20 and 30 degrees with a relative large sensitivity to view angle. The angular corrected IST13 seems more sensitive to view angle than IST12 with no correction, where IST12 still performs best of the two with respect to std and similar to IST13 with respect to bias.

In the Requirement Baseline document [AD-1] it was shown that for incidence angles smaller than 30 degrees (approximately the view angle range for dual view algorithms) the emissivity reduction due to view angle is small (<0.005). Further, the in situ data used in this analysis are below very dry atmospheres (<~3 kg/m2) which makes split window algorithms efficient for atmospheric corrections. This implies that simple algorithms may work as good as more complex algorithms in these circumstances.

As the dual view algorithms are calibrated using data from 0 and 30 degree view angles they are tuned for average view angle, which may explain slightly improving performance around 20 degrees. IST2 is tuned for view angle between 0 and 60 degrees which may explain the best performance around 30 degrees for night time data.

8.2.6 SLSTR IST comparison with NWP and Metop IST data

A comparison of SLSTR IST2 with NWP surface temperature is shown in Figure 12 for unfiltered data (left panel) and quality level 5 (right panel). The unfiltered plot reveals large differences between the two temperature estimates, with no particular systematic difference, except that cloud screened data are cold biased, as expected. This is different from comparison with observations, as exemplified in Figure 2. After filtering for quality level 5 only, an expected correlation is clear, but with a warm satellite IST BIAS. The warm IST3 Bias is surprising, because the NWP surface temperatures in the Arctic Ocean usually are warmer than observations.

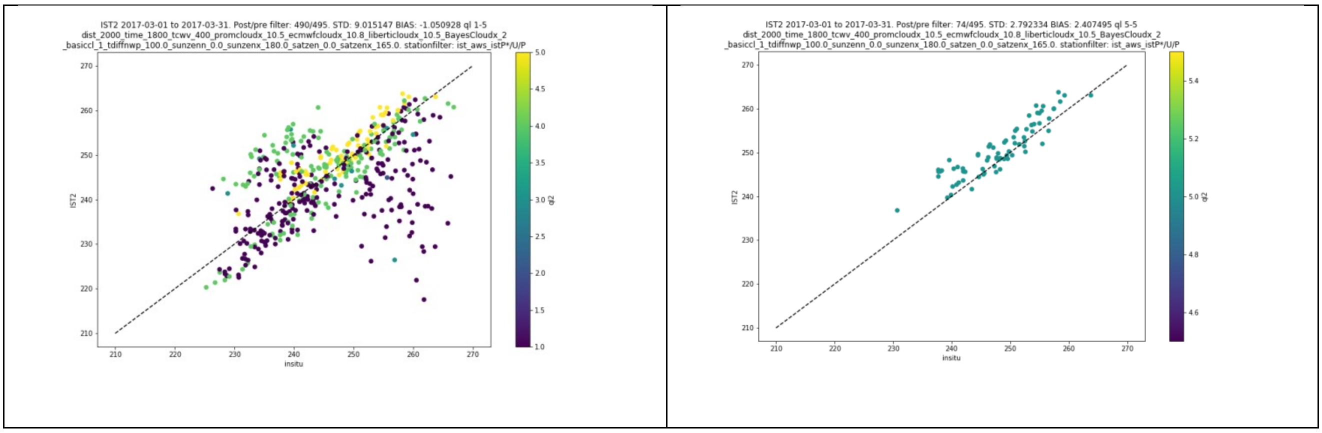


Figure 12 *IST2 from all PROMICE upper stations and EASTGRIP from March 2017 plotted against corresponding ECMWF NWP surface temperatures, with quality level 1-5 in the left panel and quality 5 only in the right panel.*

The identical algorithms SLSTR IST3 (~IST2, see [AD-8]) and OSI SAF Metop AVHRR IST (OSI-205) are evaluated individually and by match-up in Figure 13. The statistical comparison (top left vs top right panel) suggests that the SLSTR IST product has less fluctuation around the 1 to 1 line than OSI-205, but with larger Bias; std values are 2.5 K and 2.1 K and Bias -1.9 K and -2.9, respectively.

A match-up between QL 4+5 data from IST3 and OSI-205 show a high level of agreement, with a std of 0.9 K and a bias of -1.0 K (Figure 13, lower panel).

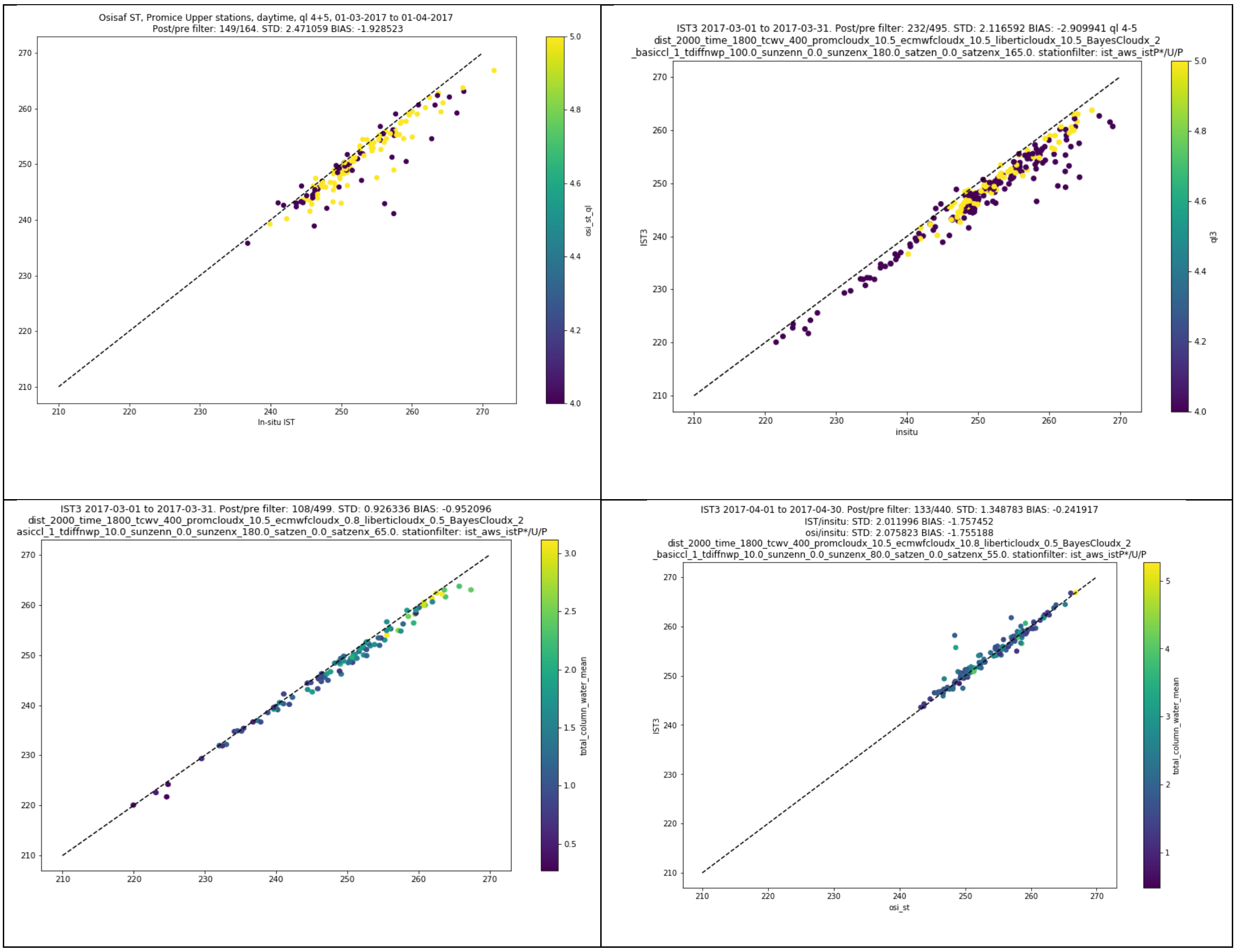


Figure 13 *Statistical comparison between OSI-205 (Metop AVHRR IST, quality levels 4+5) and SLSTR IST3 (quality levels 4+5), for March 2017, left and right top panels, respectively, against* PROMICE EASTGRIP *and Upper stations. Match-up between OSI-205 and IST3 data are shown as a scatterplot in the lower panels, for March and April 2017, left and right panels, respectively.*

8.2.7 Sentinel-3 A SLSTR IST vs. Sentinel-3 B SLSTR IST

An inter-comparison between SLSTR-A IST and SLSTR-B IST was carried out to ensure consistency. The inter-comparison was done by matching SLSTR-A IST data to SLSTR-B IST using temporal and spatial match up criteria of 60 minutes and 5 km, respectively. SLSTR-A and SLSTR-B algorithms are calibrated individually, following the procedure described in the ATBD, where the specific A and B are listed (AD-4.2). For the S3 A versus B inter-comparison a number of areas were identified as dummy match-up areas. Dummy match-up data were generated for March 2017 and only ice covered areas were sampled in this analysis, i.e. the dummy areas with red crosses in Figure 14. The pairing of A and B data is done by calculating the average IST of the central 25x25 pixels in SLSTR A in each dummy area and then matched with the mean value of the corresponding 25x25 SLSTR B pixels, where the centre SLSTR B pixel comply with the maximum distance threshold of 5 km (usually less than 1 km, i.e. pixel size). Only night time data are used (Sun-Zenith angle > 100 degrees) in order to minimize effects from surface warming and cooling during sun-lit hours in the match-up. The cloud mask is not applied, because of issues handling a pixel wise cloud mask in a 25x25 pixel window in a sensor comparison study, thus the comparison comprise both clear sky and cloudy pixels. Scatterplots of the paired SLSTR-A and

SLSTR-B IST data are shown in Figure 15 for IST2 and IST12, where bias is calculated as instrument B minus A.

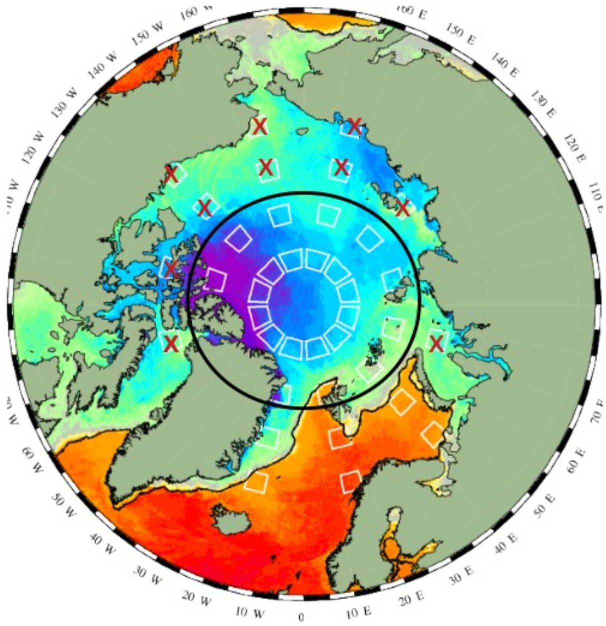
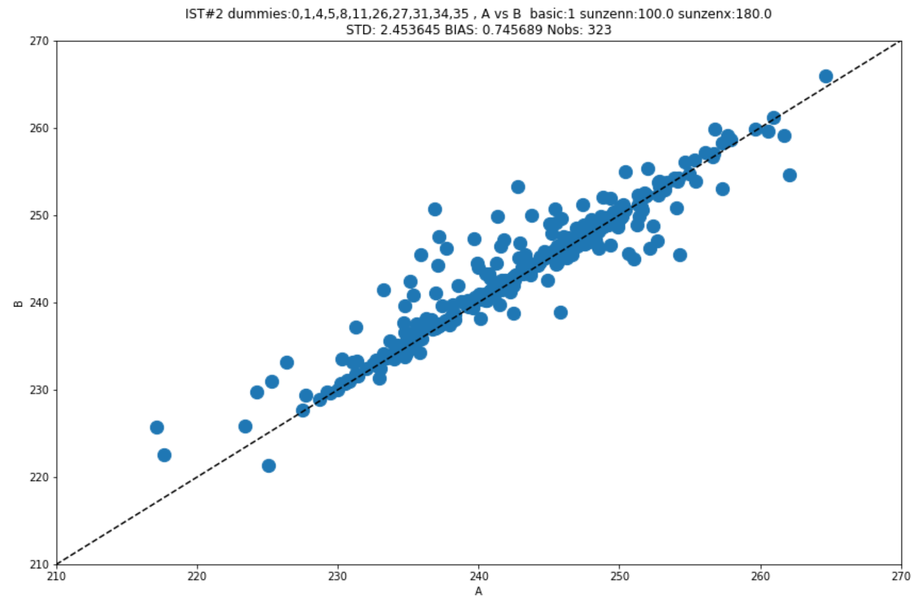


Figure 14 *Dummy Match-Up areas (white squares) for Sentinel-3 SLSTR-A vs B IST for March 2017. Areas with night time data (red crosses) are included in the A vs B comparison. All locations containing only twilight and day time data are excluded (North of black circle).*

Some scattering around the 1:1 line is present and expected due to effects from different views and changing atmospheres between A and B overpasses. This is the case for both IST2 and IST12 data, where the scattering for IST12 is larger than for IST2. The larger scattering around the 1:1 line in IST12 data is most likely originating from the fact the IST12 is a dual view algorithm and has larger mismatch errors than for the single view IST2 algorithm. This indicates that further work on instrument grid homogenisation in post processing is needed.

The plots also reveal that most data are on the 1:1 line indicating that most of the paired data are un-biased. However, both IST2 and IST12 algorithms have positive biases indicating warmer SLSTR-B IST. It is not clear what drives the asymmetry.



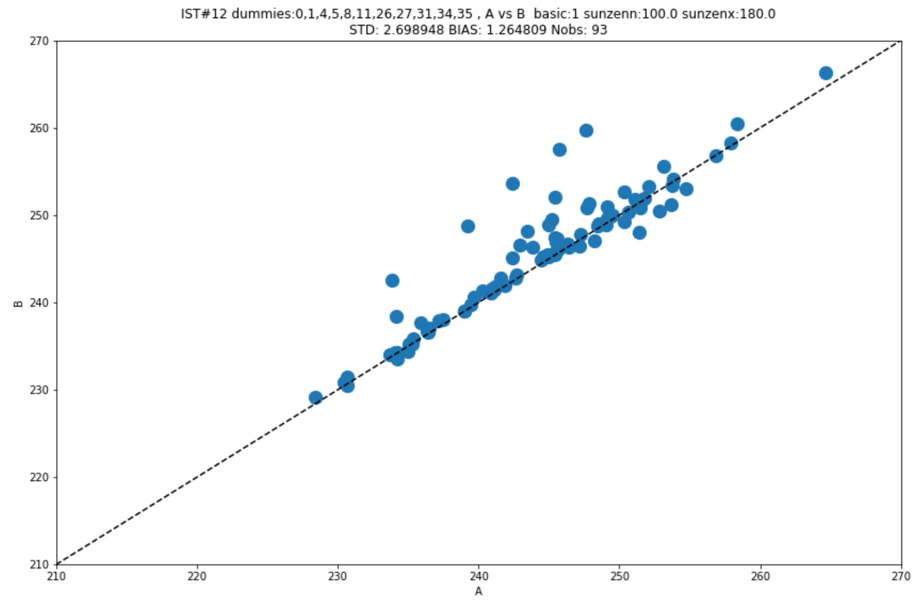


Figure 15 *Inter-comparison of SLSTR-A IST with SLSTR-B IST, night data with no cloud-mask applied. SLSTR-A IST along the primary axis and SLSTR-B IST along the secondary axis. IST2 in the top panel and IST12 in the bottom panel. Bias is B-A.*

These results show good agreement between SLSTR-A and SLSTR-B but some further work is needed on harmonisation. It is recommended to run the IST processer with SLSTR-A and SLSTR-B in parallel for a longer period to obtain improved intercomparison statistics.

8.2.8 Evaluation of Quality Levels

The SLSTR IST output is accompanied by an estimate of the IST quality level (QL) as determined by the decision tree algorithm described in the ATBD [AD-4.2]. The algorithm uses a penalty system based on a range of tests to be passed. Each failed test is known to deteriorate the IST quality statistically. That is, a specific IST value with QL 4 is not necessarily more accurate than a specific QL 5 IST value, but it will be from a statistical point of view.

The QL is determined, validated and verified using the MUDB, in 6 levels from 0 to 5.

The QL description:

QL 0: No Data. Missing or corrupt data

QL 1: Bad Data. Not cloud free according to cloud mask. More than 5 penalty points

QL 2: Worst Quality. 4 or 5 penalty points

QL 3: Low Quality. 3 penalty points

QL 4: Acceptable Quality. 1 or 2 penalty points

QL 5: Best Quality. Zero penalty points.

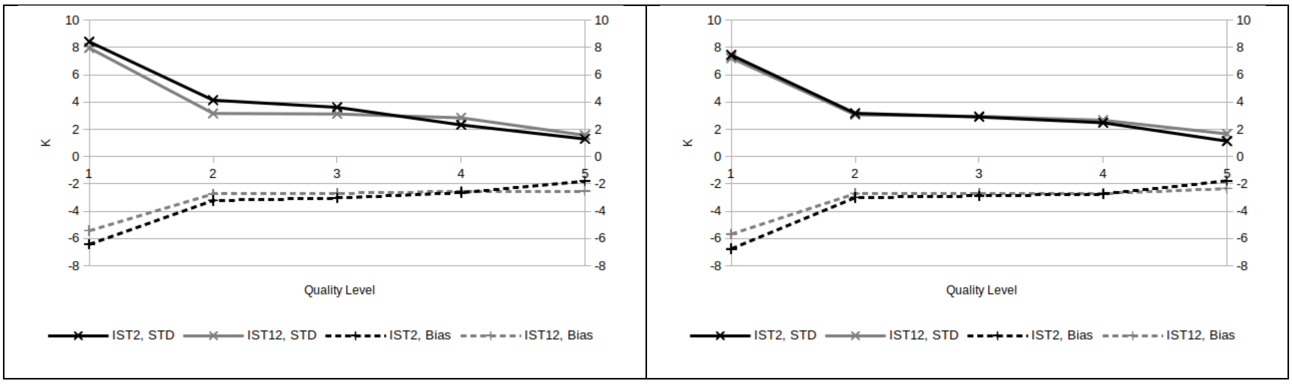


Figure 16 *Evaluation of quality levels, February-March-April (FMA) (left) and August-September-October (ASO) (right), based on in situ data from PROMICE Upper stations and EASTGRIP. Performance values are aggregated values of all data with QL X and higher. The percentage of samples in a given QL category or higher is ql1=100 %, ql2=47 %, ql3=46 %, ql4=44 % and ql5=16 % during FMA and ql5=100 %, ql5=45 %, ql5=44 %, ql5=40 % and ql5=6 % during ASO for IST2 (IST12 sample distribution is similar).*

The QL's were evaluated for two periods of 3 month each, i.e. for February, March and April and for August, September and October, and results are shown in Figure 16. The IST performance improves with increasing quality level in both periods as expected. The figure also indicate that IST12 performs slightly better for low QL's and IST 2 performs equal or best at QL 4 and 5. std and bias are below 2 K for both algorithm IST2 and IST12 for both evaluation periods, except for IST12 bias during ASO that is slightly above 2 K.

Analysis data from QL 2 to QL 4 is limited by the low number of match-ups and so a re-evaluation of the quality level algorithm tests may be needed in order provide a clearer stratification between QL 2, 3 and 4 once more match-ups are available.

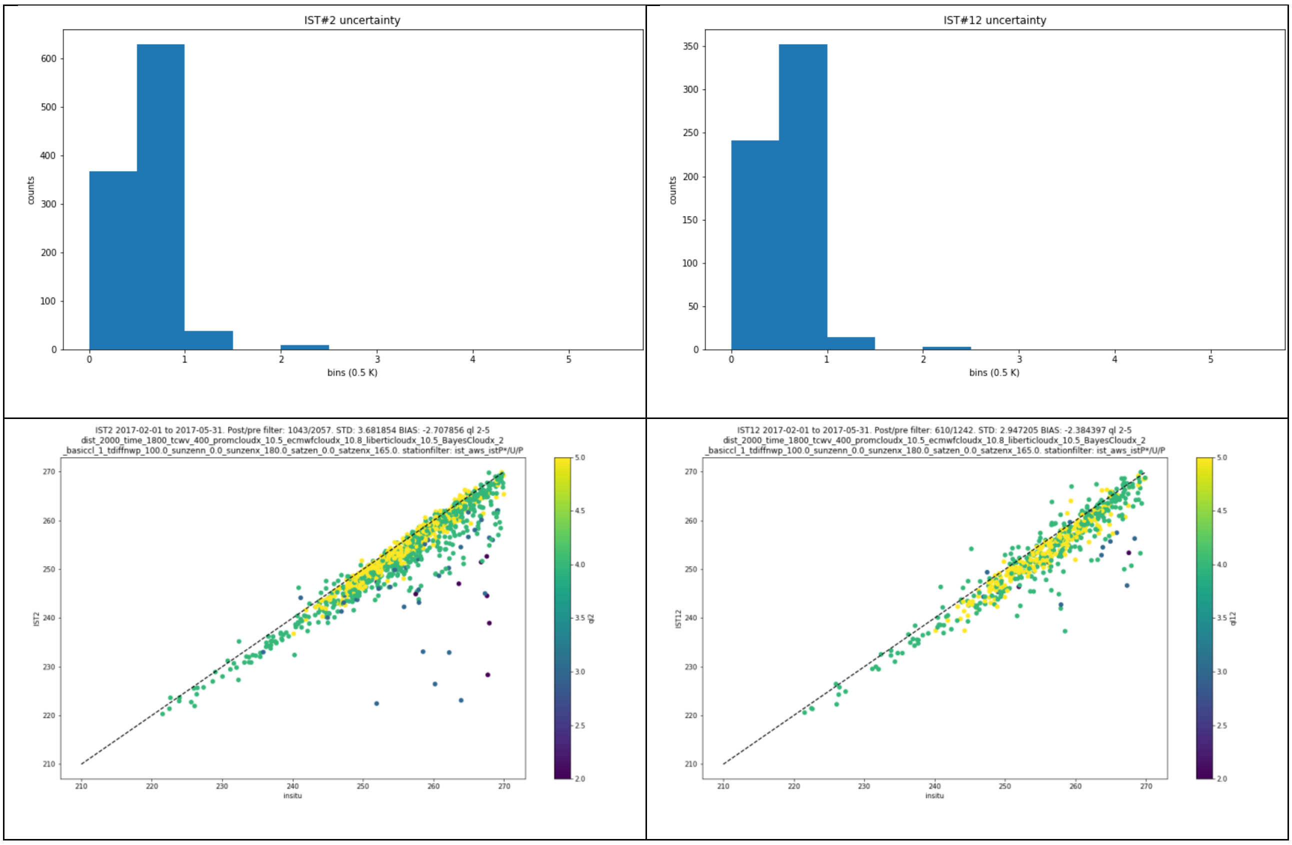
8.2.9 Evaluation of uncertainties

Each SLSTR IST retrieval is accompanied by an estimate of the IST retrieval uncertainty (σ_{sat}) as determined by the uncertainty algorithm [AD-4.2]. σ_{sat} is part of the total theoretical uncertainty (σ_{total}) that describes the expected uncertainty when comparing a given in situ measurement with the corresponding satellite IST measurement. The total uncertainty is validated against PROMICE observation data from the MUDB during the period February 1 to May 31, 2017. Here, the uncertainty evaluation consists of a comparison of the total theoretical uncertainty and the corresponding measured uncertainty of the system (σ_{sys}), i.e. the standard deviation of the differences between satellite IST and in situ observation. The measured system uncertainty shall compare with the total theoretical uncertainty. This comparison is performed within bins of 0.5 K of total theoretical uncertainty.

The total theoretical uncertainty (σ_{total}) is composed by the satellite retrieval uncertainty (σ_{sat}), the uncertainty of the ground instrument (σ_{ground}), the spatial uncertainty (σ_{space}) from comparing a point with an area measurement and the time uncertainty (σ_{time}) that originate from a time lag between satellite and in situ measurement. This approach is taken from Ghent et al. (2016).

$$\sigma_{total} = \sqrt{\sigma_{sat}^2 + \sigma_{ground}^2 + \sigma_{space}^2 + \sigma_{time}^2}$$

In this evaluation the *ground*, *space* and *time* uncertainties are considered constant and will behave as an offset uncertainty. This offset is estimated to be 0.75 K, which is the root of the squared summed of: 1) instrument uncertainty (0.2 K estimated for PROMICE surface measurements), spatial uncertainty (0.12-0.25 K within 1 km) and uncertainty from temporal sampling (0.7 K for 30 minutes). The static uncertainty values are shown in table 2 of the Requirement Baseline document report [AD-1].



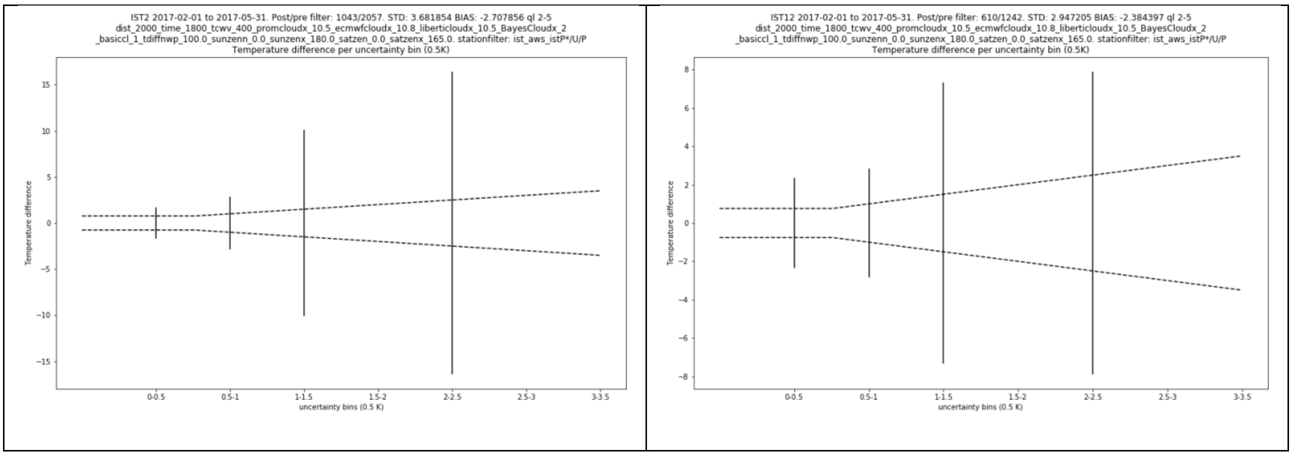


Figure 17 *Uncertainty evaluation is based on IST data with QL > 1 (i.e. cloud free) for the period February 1 to May 30, 2017, for all Upper PROMICE stations and EAST GRIP station. IST2 related plots are shown in the left column and IST12 related plots in right column. From top row: Histograms of σ_{sat} in bins of 0.5 K (top row). Scatterplot of IST as a function of corresponding in situ observation – with QL as colour-labels (middle row). Actual std of errors are plotted as bars (σ_{sys}) as a function of 0.5 K bins of total theoretical uncertainty in the bottom row. The punctured lines in the bottom plots delimit the theoretical uncertainty that σ_{sys} ideally shall resemble.*

Before looking at the uncertainty evaluation, it is worth looking at the histogram of satellite uncertainties in Figure 17. This histogram is mainly populated in the first two bins (0.0-0.5 K and 0.5-1.0 K) and only poorly populated in the subsequent bins. This is an effect from a non-ideal QL algorithm, where the most important component of the IST retrieval uncertainty originates from the global uncertainty. In the QL evaluation above, it is pointed out that not many IST values have QL 2 and 3, which will limit the distribution of the IST retrieval uncertainty. This is reflected in the uncertainty evaluation in Figure 17. The uncertainty plots show the system uncertainty (measured uncertainty) as function of bins of the total theoretical uncertainty, where the punctured line indicate the system uncertainty. If the bars within a given bin is larger than the theoretical uncertainty (punctured lines) then the uncertainty algorithm underestimates the uncertainty and vice versa.

The results show a minor underestimation of uncertainties for IST2 for small uncertainties and a moderate underestimation of uncertainties for IST12. For large theoretical uncertainties, the underestimation is large for both algorithms. Part of the underestimated uncertainties comes from erroneously classified cloud free pixels. They can be seen as even extreme outliers in scatterplot (middle row) of Figure 17. Falsely classified cloud-free pixels will most likely divert from the ground observation and at the same time often have a low global uncertainty. Another explanation of the underestimated uncertainties lies most likely in the limited dynamics of the QL algorithm that is discussed above. This results in nearly bimodal QL distribution in QL4-5 and QL2.

From these results, it is advised to revisit both the QL algorithm as well as the uncertainty algorithm, not the least the global component to the satellite retrieval uncertainty. The QL algorithm must provide a more even distribution of quality levels in order to use the full range of uncertainty.

9 Conclusions

The SLSTR IST processor provides IST from two different algorithms, a traditional split window algorithm (IST2) and a single channel – dual view algorithm (IST12). The performance of these algorithms are evaluated against automatic weather stations on Greenland ice sheet from the PROMICE project. .

Observation data from Greenland Ice sheet is not ideal for evaluating the SLSTR sea-ice surface temperature algorithms. However, it was necessary to use ice cap data for the validation, as no other in-situ data source provides the necessary data volume and quality. This is a limitation of the current evaluation because issues like atmospheric water are less relevant in the dry environment of the Greenland ice cap and the Marginal Ice Zone performance cannot be evaluated.

Validation against a small, and high quality in situ data set from sea ice, revealed that both algorithm perform excellently for sea-ice surface. Here the dual view IST12 algorithm performs best. It shall be noted that the algorithm coefficients are tuned for sea ice and that performance over sea ice may generally be superior to validation statistics against the PROMICE ice cap data. That remains to be proven.

The SLSTR IST algorithms produce IST that perform equally or superior to existing state-of-the-art operational products, with std and bias less than 2 K, for QL 5 data. The performance during twilight is close to threshold requirement of 2 K. The night time precision is up 3 K for QL 4 that is the highest possible QL during non-sunlit hours.

The performance of the SLSTR IST products is robust through time and across seasons over 1 year. Both algorithms are relatively insensitive to satellite view angles.

Two cloud methodologies have been evaluated. A combination of the native Basic cloud mask (night time) with the Liberti cloud mask (day time) and the day and night time algorithm from UoL. Both methodologies perform well during sunlit hours, and moderately successful during dark hours. The Liberti cloud mask has a tendency to partly mask out very cold IST values that are positively cloud free. In some case, obviously cloudy pixels are classified as cloud free introducing large errors into the SLSTR IST validations. The combined Liberti/Basic cloud screening methodology seems to perform slightly better than the UoL cloud screening, but more work is needed to make a strong conclusion and both approaches can be improved. However, all current cloud masking methods fall short of what is needed to minimise cloud screening errors in the SLSTR IST product.

Our evaluation of the IST theoretical uncertainties underestimates the IST uncertainties. This is believed to be caused primarily by poor cloud masking, and troublesome quality level assignments. The quality level algorithm implemented for IST2 and IST12 works reasonable well. However, the QL 2 and 3 are poorly represented and it is recommended to develop the "penalty" algorithm when issues of cloud screening are improved. The unevenly distributed quality levels distorts the performance of the SLSTR IST uncertainty estimation and the QL and uncertainty estimation algorithms should be looked at as a coupled system in order to work well. This is mainly due to a global uncertainty term that depends on the quality level.

10 Future work

The conclusions presented here strongly support further work in this area. In particular, we recommend additional activities 1) To develop means and facilities for fiducial sea ice temperature observations, and 2) to focus on improving SLSTR cloud masking over sea ice.

Regarding cloud masking, further testing of the UOL Leicester cloud mask is recommended, as more tuning may improve the results. It is also recommended to implement and test the PPS software for SLSTR data, which is an alternative to the cloud screening methods tested here applied to other sensors (e.g. AVHRR) by the EUMETSAT NWC SAF.

Regarding sea ice measurements, this work underlines the urgent need for fiducial in-situ surface temperatures for sea ice. High quality sea ice observations are only available from field campaigns or irregular and short lasting buoys deployed. Sea ice data are needed for algorithm development, because important details in algorithm development are not reflected in the available observation data. Furthermore, there are no observation data sets currently existing of required volume and quality to evaluate IST and SST algorithms in marginal ice zones.

Finally, we recommend further work on harmonisation of the SLSTR-A and SLSTR-B IST algorithms. It is needed to use data from both sensors at the same time. In addition, the current quality level and uncertainty algorithms need to be revisited

10.1 Considerations on future in situ sea ice observations platforms

The main objective for fiducial surface temperature measurements is to provide traceable, high precision data, at representative temporal and spatial resolutions for the area of interest.

For temperature monitoring of the Arctic ocean that means a large number of in situ platforms distributed over characteristic sea ice domains like 1) closed ice in the central Arctic, 2) the marginal ice zones with complex atmospheric compositions, 3) in deformation zones along northern Greenland and Canadian archipelago, and 4) in seasonal ice formation zones along the Siberian sector of the Arctic. Furthermore, with surface warming and cooling rates up to 10 K/h in spring and autumn, the recording frequency shall preferable be around 10 minutes. Finally, the precision of the observations must be well below the target precision requirement (1 K) of the IST products, no worse than half the target accuracy.

The technology to measure fiducial surface temperatures exists today in two forms, 1) as radiometric measurements, similar to satellite measurements, or 2) as thermal measurements from Ice Massbalance Buoys, where thermistors on a string measure the temperature profile from the air through the snow pack to the ice. The drawbacks of these two solutions are the price and a relatively complicated deployment procedure that requires dedicated personnel. Both setups consist of an instrument (radiometer or thermistor string), a data logger, a communication package and a battery pack. The price for one unit of either of these solutions is 8000-9000 EUR.

With our experience from Arctic field work and associated logistics, it will be challenging to establish a sustainable Arctic network at a required density using either of these platforms.

At DMI we are conceptualizing a new idea based on the proven concepts of traditional drifting buoys, like SVP/ISVP platforms. The advantage of SVP platforms is the price and the plug-and-play deployment concept. The SVP buoys cost around 1300 EUR and they can be deployed by everyone by simply putting the buoy on the ice. The challenge is to combine the SVP technology and wrapping with a device to obtain fiducial surface temperatures.

The idea is to add a thermistor stick to the SVP housing. A stick of approximately 0.5 m in length will at practically any time be able to reach from the air to the snow/ice interface. 10-20 evenly distributed thermistors on the stick will provide a temperature profile resolution of 2.5-5 cm resolution. This is sufficient to resolve the actual surface skin temperature, from the closest thermistor over the snow surface. Most often the coldest temperature from the temperature profile resembles the skin temperature very precisely. Occasionally other means must be applied to determine the skin temperature, e.g. by analyzing the diurnal temperature dynamics.

An alternative solution could be to develop a package with well characterized infrared radiometric instruments that can measure the skin surface temperature of the snow and sea ice during the harsh Arctic Environment. Such a package could be tied onto already existing ice tethered platforms. The advantage of this solution is, that the instrument measures the same physical parameter as the satellite, namely the very top skin of the snow and sea ice and no additional processing has to take place to identify the interface. This method relies on the other hand, on emissivity assumptions, which can introduce uncertainties whether it is bare ice or fresh snow. Comparisons from Qaanaaq show however that satellite versus in situ differences can be as low as 1°C with this type of instruments.

It is outside the scope of the SLSTR IST/SST project to develop a fiducial observation system, but data records from the integrated observation system in the DMI winter observatory in Qaanaaq, NW Greenland, enables us to perform initial analysis on the precision of the such alternative in situ devises. With little effort we can provide thorough estimate of the expected performance of e.g. a modified SVP setup or a radiometer based in situ platform. If analysis shows that the concepts can provide adequate precision, we believe that new concepts can be distributed as current traditional drifters - but with fiducial ice and snow surface temperature.

11 Acknowledgement

The work is funded by the European Union under the EUMETSAT Copernicus contract EUM/CO/18/4600002129/AOC

12 References

Bamber, J.L. and A. R. Harris. The atmospheric correction for satellite infrared radiometer data in polar regions. GEOPHYSICAL RESEARCH LETTERS, Vol. 21, No. 19, pp.2111-2114, 1994.

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. and S. Eastwood. Validation Report for the OSI-SAF High Latitude L2 Sea and Sea Ice Surface Temperature. Presently in final revision. osisaf_cdop2_ss2_valrep_hl-l2-sst-ist_v1p0.pdf, http://osisaf.met.no/docs/, 2016.

Dybkjaer, G., S. Eastwood, R.H. PFeiffer and E. Howe. OSI SAF High Latitudes L2 Sea and Sea Ice Surface Temperature Product User Manual, OSI205 Version 1.1, 2017.

Fox, N. FRM4STS Workshop Proceedings & FRM4STS Scientific Roadmap (http://www.frm4sts.org/wp-content/uploads/sites/3/2018/10/OFE-D-160_170-V1-Iss-1-Ver-2-signed.pdf) 2018.

Ghent, D., *Ortiz, P.,Tonboe, R.,and Corlett, G.* Validation Report for Surface Temperature Uncertainties. EU SURFACE TEMPERATURE FOR ALL CORNERS OF EARTH (EUSTACE) DELIVERABLE D3.1. H2020 GRANT AGREEMENT 640171, 2016.

Ghent, D. and Sembhi, University of Leicester. Thermal Infrared Probabilistic Cloud Detection for Land Algorithm Theoretical Basis Document for Sentinel-3, S3-ATBD_UOL_CLOUDS_V3, 2017.

Høyer, J.L., E. Alerskans, P. Nielsen-Englyst, P. Thejll, G. Dybkjær and R. Tonboe. Fiducial Reference Measurements for validation of Surface Temperature from Satellites (FRM4STS): Ice Surface Temperature Comparison of Participants Radiometers., Technical Report ESA Contract No. 4000113848_15I-LG, 2017b.

Høyer, J. L., Lang, A., Tonboe, R. T., Eastwood, S., Wimmer, W., Dybkjær, G. (2017) Report from Field InterComparison Experiment (FICE) for ice surface temperature, FRM4STS report, available at: www.frm4sts.org.

Høyer, J. L., E. Alerskans, P. Nielsen-Englyst P., P. Thejll, G. Dybkjær and R. Tonboe, Detailed investigation of the uncertainty budget for Non-recoverable IST observations and their SI traceability, FRM4STS Technical Report, ESA Contract No. 4000113848_15I-LG, 2018.

Liberti, G.L. Cloud screening over sea-ice and marginal ice zones: Final Report. Official SLSTR cloud mask for over ice, EUMETSAT/Copernicus, ITT No. 15/211424, 2017.

APPENDIX A

Day (sunzen 0-80)

IST_algo #	parameter	08-2016	09-2016	10-2016	11-2016	12-2016	01-2017	02-2017	03-2017	04-2017	05-2017	06-2017	07-2017
0	std	1.55	1.96	2.82	4.33	nan	nan	2.85	1.4	1.78	1.94	2.42	1.66
	bias	-1.72	-1.8	-1.01	-5.96	nan	nan	-1.3	-1	-1.66	-1.92	-2.23	-1.62
	Datapoints %	59.74	47.89	35.29	57.14	nan	0.00	46.43	36.67	42.81	77.34	60.00	78.08
	data post	46	102	18	4	0	0	13	66	125	215	60	57
	data pre	77	213	51	7	0	1	28	180	292	278	100	73
1	std	1.71	2.06	2.82	4.38	nan	nan	2.84	1.36	1.86	1.96	2.39	1.78
	bias	-2.17	-2.25	-1.61	-6.75	nan	nan	-1.7	-1.56	-2.28	-2.85	-2.95	-2.37
	Datapoints %	59.74	47.89	35.29	57.14	nan	0.00	46.43	36.67	42.81	77.34	60.00	78.08
	data post	46	102	18	4	0	0	13	66	125	215	60	57
	data pre	77	213	51	7	0	1	28	180	292	278	100	73
2	std	1.61	1.91	1.73	0.19	nan	nan	1.51	1.18	2.03	1.51	1.73	1.04
	bias	-1.44	-1.53	-1.48	-1.9	nan	nan	-1.71	-1.28	-1.76	-1.42	-1.55	-1.22
	Datapoints %	57.14	49.44	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	176	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
3	std	1.61	1.91	1.73	0.19	nan	nan	1.51	1.18	2.03	1.51	1.73	1.04
	bias	-1.45	-1.54	-1.5	-1.92	nan	nan	-1.73	-1.31	-1.78	-1.44	-1.57	-1.23
	Datapoints %	57.14	49.44	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	176	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
4	std	2.16	2.09	1.24	0.15	nan	nan	1.18	1.62	1.69	1.45	1.66	1.43
	bias	-0.39	-0.89	-1.99	-2.17	nan	nan	-1.95	-1.13	-1.38	-0.86	-0.75	-0.63
	Datapoints %	51.95	37.74	33.33	28.57	nan	0.00	35.71	33.71	40.07	70.86	56.00	76.71
	data post	40	80	17	2	0	0	10	60	117	197	56	56
	data pre	77	212	51	7	0	1	28	178	292	278	100	73
5	std	1.62	1.96	1.75	0.23	nan	nan	1.49	1.19	2.05	1.51	1.74	1.03
	bias	-1.75	-1.88	-1.79	-2.23	nan	nan	-2.04	-1.63	-2.13	-1.85	-2	-1.61
	Datapoints %	57.63	50.29	36.78	28.57	nan	0.00	34.04	40.68	47.59	74.81	68.12	80.36
	data post	68	172	32	2	0	0	16	120	217	300	94	90
	data pre	118	342	87	7	0	1	47	295	456	401	138	112
6	std	2.06	2.41	1.31	0.08	nan	nan	1.4	1.86	1.83	1.6	1.78	1.64
	bias	-0.69	-1.01	-2.31	-2.43	nan	nan	-2.24	-1.39	-1.6	-1.05	-1.01	-0.74
	Datapoints %	50.65	37.26	33.33	28.57	nan	0.00	35.71	33.71	40.07	69.78	55.00	76.71
	data post	39	79	17	2	0	0	10	60	117	194	55	56
	data pre	77	212	51	7	0	1	28	178	292	278	100	73
7	std	2.43	3.09	2.2	0.51	nan	nan	2.57	2.41	2.7	2.24	2.64	2.29
	bias	-1.77	-2.15	-3.09	-2.93	nan	nan	-3.76	-3.42	-3.75	-3.08	-3.22	-2.51
	Datapoints %	50.65	37.26	33.33	28.57	nan	0.00	35.71	33.15	39.73	70.65	56.57	77.78
	data post	39	79	17	2	0	0	10	59	116	195	56	56
	data pre	77	212	51	7	0	1	28	178	292	276	99	72
8	std	1.65	1.8	1.73	0.26	nan	nan	1.43	1.23	2.1	1.51	1.79	1.05
	bias	-1.85	-1.92	-1.97	-2.42	nan	nan	-2.2	-1.81	-2.37	-2.2	-2.41	-1.89
	Datapoints %	57.14	49.16	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	175	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
9	std	1.65	1.8	1.71	0.24	nan	nan	1.42	1.22	2.1	1.51	1.78	1.05
	bias	-1.98	-2.03	-2.07	-2.58	nan	nan	-2.32	-1.92	-2.48	-2.33	-2.54	-2
	Datapoints %	57.14	49.16	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	175	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
10	std	1.61	1.91	1.73	0.19	nan	nan	1.51	1.18	2.03	1.51	1.73	1.04
	bias	-1.44	-1.54	-1.49	-1.92	nan	nan	-1.72	-1.29	-1.77	-1.44	-1.57	-1.23
	Datapoints %	57.14	49.44	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	176	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
11	std	1.61	1.91	1.74	0.19	nan	nan	1.52	1.18	2.03	1.51	1.73	1.04
	bias	-1.42	-1.52	-1.47	-1.89	nan	nan	-1.7	-1.27	-1.74	-1.4	-1.52	-1.2
	Datapoints %	57.14	49.44	37.50	28.57	nan	0.00	34.04	40.54	47.07	74.81	65.73	80.36
	data post	68	176	33	2	0	0	16	120	217	300	94	90
	data pre	119	356	88	7	0	1	47	296	461	401	143	112
12	std	2.38	2.28	1.29	0.16	nan	nan	1.33	1.65	1.76	1.54	1.78	1.56
	bias	-0.91	-1.43	-2.58	-2.79	nan	nan	-2.6	-1.79	-2.03	-1.5	-1.7	-1.31
	Datapoints %	51.95	37.74	31.37	28.57	nan	0.00	35.71	33.71	40.07	71.22	56.00	76.71
	data post	40	80	16	2	0	0	10	60	117	198	56	56
	data pre	77	212	51	7	0	1	28	178	292	278	100	73
13	std	2.19	2.58	1.42	0.09	nan	nan	1.63	2.03	1.92	1.48	1.97	1.92
	bias	-1.06	-1.49	-2.82	-2.9	nan	nan	-2.81	-1.91	-2.12	-1.41	-1.65	-1.28
	Datapoints %	50.65	36.79	31.37	28.57	nan	0.00	35.71	33.71	39.73	69.42	56.00	75.34
	data post	39	78	16	2	0	0	10	60	116	193	56	55
	data pre	77	212	51	7	0	1	28	178	292	278	100	73
14	std	2.23	2.6	1.44	0.09	nan	nan	1.66	2.04	1.93	1.5	1.99	1.92
	bias	-1.02	-1.45	-2.8	-2.88	nan	nan	-2.79	-1.89	-2.09	-1.38	-1.62	-1.24
	Datapoints %	50.65	36.79	31.37	28.57	nan	0.00	35.71	33.71	39.73	69.42	56.00	75.34
	data post	39	78	16	2	0	0	10	60	116	193	56	55
	data pre	77	212	51	7	0	1	28	178	292	278	100	73

APPENDIX B

Twilight (sunzen 80-100)

IST_algo #	parameter	08-2016	09-2016	10-2016	11-2016	12-2016	01-2017	02-2017	03-2017	04-2017	05-2017	06-2017	07-2017
0	std	0.36	1.62	2.31	3.42	2.53	2.17	2.8	1.48	4.12	1.1	0.28	0
	bias	-0.83	-1.4	-1.06	-2.18	-1.79	-1.65	-2.14	-0.89	-3.67	-0.68	-0.14	0.81
	Datapoints %	25.00	38.10	36.06	47.26	54.92	47.32	17.93	28.79	19.05	44.44	50.00	14.29
	data post	5	48	75	95	106	106	26	19	12	8	6	1
	data pre	20	126	208	201	193	224	145	66	63	18	12	7
1	std	0.33	1.74	2.41	3.64	2.57	2.18	2.63	1.54	4.04	0.89	0.11	0
	bias	-1.18	-1.37	-0.91	-2.14	-1.77	-1.63	-2.05	-0.71	-3.9	-1.25	-1.03	0.28
	Datapoints %	25.00	38.10	36.06	47.76	54.92	47.32	17.93	28.79	19.05	44.44	50.00	14.29
	data post	5	48	75	96	106	106	26	19	12	8	6	1
	data pre	20	126	208	201	193	224	145	66	63	18	12	7
2	std	0.69	1.81	1.69	2.99	1.98	2.02	1.84	1.33	3.58	3.03	0.84	0.79
	bias	-0.51	-0.7	-0.79	-1.91	-1.52	-1.31	-2.33	-0.29	-2.73	-1.36	-0.17	0.73
	Datapoints %	24.14	31.13	29.32	38.37	49.68	44.47	15.20	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	132	157	165	38	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
3	std	0.7	1.81	1.69	3.01	2	2.02	1.85	1.34	3.58	3.03	0.85	0.79
	bias	-0.51	-0.7	-0.79	-1.93	-1.57	-1.37	-2.37	-0.4	-2.75	-1.38	-0.18	0.72
	Datapoints %	24.14	31.13	29.32	38.37	49.68	44.47	15.20	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	132	157	165	38	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
4	std	0.48	2.03	2.04	3.04	2.4	2.54	3.33	1.82	4.5	1.05	0.97	0
	bias	-0.61	-0.6	-0.56	-1.57	-1.14	-0.93	-1.74	0.44	-1.52	-0.86	-0.23	3.24
	Datapoints %	20.00	39.68	40.29	47.26	54.40	49.55	19.44	31.25	23.81	44.44	50.00	14.29
	data post	4	50	83	95	105	111	28	20	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	12	7
5	std	0.61	1.55	1.63	2.73	1.99	1.94	2.06	1.3	3.64	3	0.79	0.8
	bias	-0.84	-1.11	-0.99	-2.29	-1.78	-1.59	-2.45	-0.54	-3.16	-1.67	-0.6	0.39
	Datapoints %	25.00	29.37	29.21	36.36	49.51	43.77	14.57	25.47	28.89	38.00	18.75	25.93
	data post	7	42	92	120	151	158	36	27	26	19	6	7
	data pre	28	143	315	330	305	361	247	106	90	50	32	27
6	std	0.56	1.96	2.19	3.27	2.67	2.83	3.04	1.95	4.68	1.2	1.19	0
	bias	-0.74	-0.43	-0.65	-1.6	-1.22	-0.98	-2.31	0.41	-1.53	-1.14	-0.47	3.54
	Datapoints %	20.00	38.10	40.29	47.26	54.92	49.11	18.75	31.25	23.81	44.44	50.00	14.29
	data post	4	48	83	95	106	110	27	20	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	12	7
7	std	0.68	2.55	2.6	3.39	2.72	2.88	4.89	2.06	4.95	1.37	1.47	0
	bias	-0.68	-0.71	-0.75	-1.64	-1.17	-0.71	-2.05	0.56	-2.38	-1.14	-0.6	3.2
	Datapoints %	20.00	38.89	39.32	46.77	55.44	49.11	19.44	29.69	23.81	44.44	54.55	14.29
	data post	4	49	81	94	107	110	28	19	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	11	7
8	std	0.44	1.91	1.75	3.01	2.01	2	2.02	1.4	3.51	2.93	0.58	0.8
	bias	-0.86	-0.67	-0.78	-1.87	-1.67	-1.44	-2.28	-0.36	-3.06	-1.73	-0.88	0.34
	Datapoints %	24.14	31.13	29.32	38.08	50.00	44.20	15.60	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	131	158	164	39	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
9	std	0.43	1.91	1.76	2.9	2.01	1.99	2.01	1.38	3.52	2.91	0.54	0.77
	bias	-0.86	-0.82	-0.92	-1.9	-1.77	-1.55	-2.37	-0.49	-3.12	-1.79	-0.91	0.28
	Datapoints %	24.14	31.13	29.32	37.79	50.00	44.20	15.60	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	130	158	164	39	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
10	std	0.68	1.81	1.69	2.99	1.98	2.02	1.84	1.33	3.58	3.03	0.84	0.79
	bias	-0.51	-0.69	-0.75	-1.91	-1.52	-1.31	-2.32	-0.29	-2.73	-1.36	-0.18	0.73
	Datapoints %	24.14	31.13	29.32	38.37	49.68	44.47	15.20	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	132	157	165	38	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
11	std	0.7	1.81	1.69	2.99	1.98	2.02	1.84	1.33	3.58	3.04	0.86	0.8
	bias	-0.51	-0.69	-0.75	-1.91	-1.52	-1.31	-2.33	-0.29	-2.73	-1.36	-0.16	0.74
	Datapoints %	24.14	31.13	29.32	38.37	49.68	44.47	15.20	25.93	29.79	38.00	17.65	25.93
	data post	7	47	95	132	157	165	38	28	28	19	6	7
	data pre	29	151	324	344	316	371	250	108	94	50	34	27
12	std	0.52	1.96	2.17	2.97	2.48	2.68	3.6	1.91	4.46	1.14	1.07	0
	bias	-0.99	-0.8	-1.08	-1.93	-1.63	-1.43	-2.27	-0.06	-2.18	-1.49	-0.91	3.12
	Datapoints %	20.00	38.10	40.29	47.26	55.44	50.00	19.44	31.25	23.81	44.44	50.00	14.29
	data post	4	48	83	95	107	112	28	20	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	12	7
13	std	0.69	2.14	2.39	3.4	2.83	3.07	2.61	1.87	4.74	1.33	1.29	0
	bias	-0.96	-0.68	-1.13	-1.89	-1.68	-1.32	-2.5	0.23	-1.99	-1.58	-0.85	4.39
	Datapoints %	20.00	38.10	40.29	46.27	55.44	48.66	18.06	29.69	23.81	44.44	50.00	14.29
	data post	4	48	83	93	107	109	26	19	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	12	7
14	std	0.69	2.15	2.4	3.42	2.85	3.08	2.64	1.89	4.76	1.33	1.31	0
	bias	-0.94	-0.65	-1.1	-1.86	-1.65	-1.29	-2.48	0.26	-1.95	-1.56	-0.83	4.37
	Datapoints %	20.00	38.10	40.29	46.27	55.44	48.66	18.06	29.69	23.81	44.44	50.00	14.29
	data post	4	48	83	93	107	109	26	19	15	8	6	1
	data pre	20	126	206	201	193	224	144	64	63	18	12	7

APPENDIX C

Night (sunzen 100-180)

ST_algo #	parameter	08-2016	09-2016	10-2016	11-2016	12-2016	01-2017	02-2017	03-2017	04-2017	05-2017	06-2017	07-2017
0	std	nan	0.83	2.92	3.13	2.9	3.3	2.93	2.46	2.74	nan	nan	nan
	bias	nan	-0.83	-2.01	-1.43	-2.51	-2.29	-1.3	-1.72	-1.55	nan	nan	nan
	Datapoints %	nan	52.00	55.26	54.55	59.79	54.55	43.14	55.81	66.67	nan	nan	nan
	data post	0	13	42	36	58	42	22	24	8	0	0	0
	data pre	0	25	76	66	97	77	51	43	12	0	0	0
1	std	nan	0.8	2.89	3.14	2.97	3.12	3.01	2.29	2.83	nan	nan	nan
	bias	nan	-1.39	-2.05	-1.34	-2.44	-1.93	-1.57	-2.07	-2	nan	nan	nan
	Datapoints %	nan	52.00	55.26	54.55	60.82	53.25	45.10	55.81	66.67	nan	nan	nan
	data post	0	13	42	36	59	41	23	24	8	0	0	0
	data pre	0	25	76	66	97	77	51	43	12	0	0	0
2	std	nan	1.81	2.51	3.15	2.44	2.69	1.84	2.65	3.81	nan	nan	nan
	bias	nan	-0.42	-1.39	-1.58	-1.9	-2.18	-0.95	-1.34	-1.25	nan	nan	nan
	Datapoints %	nan	54.10	50.00	52.00	56.48	49.66	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	65	109	73	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
3	std	nan	1.82	2.51	3.16	2.46	2.68	1.83	2.65	3.8	nan	nan	nan
	bias	nan	-0.44	-1.41	-1.62	-1.94	-2.23	-1	-1.38	-1.28	nan	nan	nan
	Datapoints %	nan	54.10	50.00	52.00	56.48	49.66	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	65	109	73	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
4	std	nan	3.54	3.25	3.22	3.14	3.23	3.35	2.31	3.19	nan	nan	nan
	bias	nan	-0.14	-1.9	-0.73	-2.01	-1.91	-1.88	-0.61	-1.03	nan	nan	nan
	Datapoints %	nan	60.00	58.67	54.55	62.89	51.95	50.00	57.14	66.67	nan	nan	nan
	data post	0	15	44	36	61	40	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0
5	std	nan	1.84	2.53	3.01	2.35	2.8	1.8	2.54	3.81	nan	nan	nan
	bias	nan	-0.74	-1.73	-2.04	-2.32	-2.51	-1.15	-1.75	-1.58	nan	nan	nan
	Datapoints %	nan	55.93	49.17	50.41	55.43	47.48	45.74	59.09	60.00	nan	nan	nan
	data post	0	33	59	61	102	66	43	52	21	0	0	0
	data pre	0	59	120	121	184	139	94	88	35	0	0	0
6	std	nan	3.72	3.14	3.32	2.89	3.58	3.3	2.47	3.44	nan	nan	nan
	bias	nan	-0.06	-1.87	-0.83	-2.09	-2.11	-2.11	-0.69	-1.23	nan	nan	nan
	Datapoints %	nan	60.00	57.33	54.55	59.79	51.95	50.00	57.14	66.67	nan	nan	nan
	data post	0	15	43	36	58	40	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0
7	std	nan	2.6	2.97	3.26	3.06	3.1	4.06	2.68	3.35	nan	nan	nan
	bias	nan	0.69	-1.52	-0.76	-2.05	-1.56	-2.22	-0.87	-1.22	nan	nan	nan
	Datapoints %	nan	56.00	54.67	54.55	60.82	50.65	50.00	57.14	66.67	nan	nan	nan
	data post	0	14	41	36	59	39	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0
8	std	nan	1.86	2.52	3.26	2.5	2.85	1.87	2.6	3.8	nan	nan	nan
	bias	nan	-0.78	-1.61	-1.76	-2.02	-2.45	-1.22	-1.73	-1.69	nan	nan	nan
	Datapoints %	nan	54.10	50.00	52.00	56.48	50.34	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	65	109	74	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
9	std	nan	1.83	2.53	3.26	2.5	2.72	1.87	2.6	3.81	nan	nan	nan
	bias	nan	-0.82	-1.71	-1.87	-2.1	-2.43	-1.28	-1.78	-1.76	nan	nan	nan
	Datapoints %	nan	54.10	50.00	51.20	56.48	49.66	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	64	109	73	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
10	std	nan	1.81	2.51	3.15	2.44	2.69	1.84	2.65	3.81	nan	nan	nan
	bias	nan	-0.42	-1.39	-1.58	-1.9	-2.18	-0.95	-1.34	-1.26	nan	nan	nan
	Datapoints %	nan	54.10	50.00	52.00	56.48	49.66	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	65	109	73	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
11	std	nan	1.81	2.51	3.15	2.43	2.69	1.84	2.66	3.81	nan	nan	nan
	bias	nan	-0.42	-1.39	-1.58	-1.9	-2.18	-0.95	-1.33	-1.24	nan	nan	nan
	Datapoints %	nan	54.10	50.00	52.00	56.48	49.66	45.83	59.55	60.00	nan	nan	nan
	data post	0	33	62	65	109	73	44	53	21	0	0	0
	data pre	0	61	124	125	193	147	96	89	35	0	0	0
12	std	nan	3.61	2.88	2.87	2.9	3.54	3.45	2.33	3.45	nan	nan	nan
	bias	nan	-0.53	-2.04	-0.95	-2.57	-2.58	-2.52	-1.16	-1.7	nan	nan	nan
	Datapoints %	nan	60.00	54.67	53.03	59.79	51.95	50.00	57.14	66.67	nan	nan	nan
	data post	0	15	41	35	58	40	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0
13	std	nan	3.71	2.5	3.01	2.84	4.11	3.41	2.73	3.93	nan	nan	nan
	bias	nan	-0.23	-1.67	-0.97	-2.38	-2.75	-2.64	-1.05	-1.56	nan	nan	nan
	Datapoints %	nan	60.00	50.67	53.03	57.73	51.95	50.00	57.14	66.67	nan	nan	nan
	data post	0	15	38	35	56	40	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0
14	std	nan	3.75	2.52	3.01	2.85	4.12	3.4	2.73	3.92	nan	nan	nan
	bias	nan	-0.19	-1.64	-0.94	-2.36	-2.72	-2.62	-1.02	-1.56	nan	nan	nan
	Datapoints %	nan	60.00	50.67	53.03	57.73	51.95	50.00	57.14	66.67	nan	nan	nan
	data post	0	15	38	35	56	40	25	24	8	0	0	0
	data pre	0	25	75	66	97	77	50	42	12	0	0	0