



fiducial reference
temperature
measurements



esa

Fiducial Reference Measurements for validation of Surface Temperature from Satellites (FRM4STS)

ESA Contract No. 4000113848_15I-LG

OP-40: Towards Field Inter-Comparison Experiment (FICE) for ice surface temperature

JULY 2017

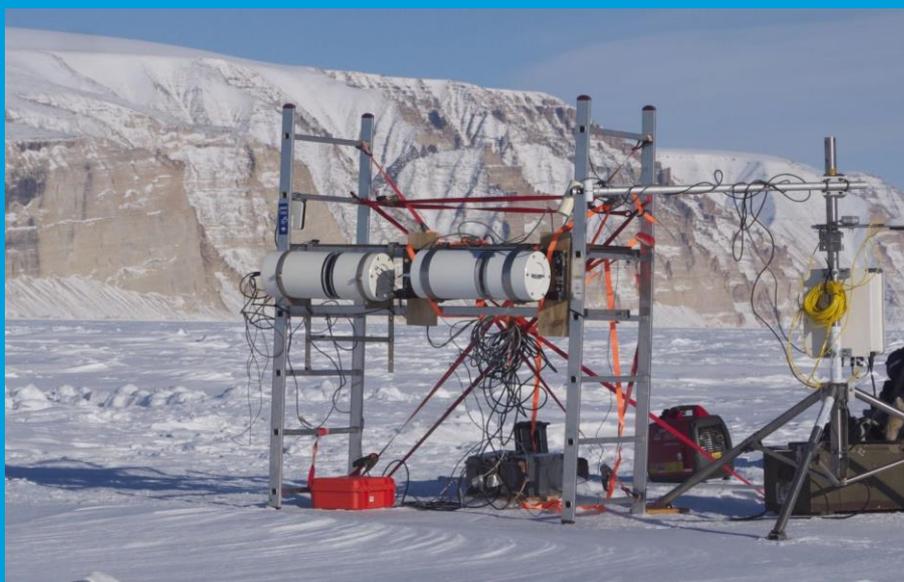
Reference OFE-OP-40-V1-Iss-1-Ver-1-Draft
 Issue 1
 Revision 1
 Date of Issue 7 July 2017
 Document Type TR-5

Approval/Acceptance			
ESA Craig Donlon Technical Officer		NPL Andrew Brown Project Manager	 Andrew Brown, NPL
	<i>Signature</i>		<i>Signature</i>

Report from Field Inter-Comparison Experiment (FICE) for ice surface temperature

Jacob L. Høyer¹ and Andreas M. Lang¹, Rasmus Tonboe¹, Steinar Eastwood², Werenfrid Wimmer³ and Gorm Dybkjær¹

¹Danish Meteorological Institute, ²Meteorological Institute of Norway, ³National Oceanographic Centre, Southampton



DMI
Vejr, klima og hav



fiducial reference
temperature
measurements

Abstract

A field inter-comparison experiment over sea ice was conducted in March-April, 2016 on the sea ice off Qaanaaq, in Northwest Greenland. Six different thermal infrared radiometers participated in the inter-comparison, the first of its kind over sea ice, including two Fiducial Reference Measurements Thermal Infrared Radiometers. The weather conditions were typical for a high Arctic environment, with surface temperatures between -30°C and -10°C and low wind speeds. Pairwise comparisons of the 10 minute averaged brightness temperatures from the radiometers showed mean differences between 0.13 and 1.0 K and standard deviation from 0.05 to 1.20 K. Inter-comparison of the sky brightness temperatures showed that a lower limit of ~ 173 K was reached with both FRM-TIR instruments during cold sky conditions. Several other experiments have been conducted during the field campaign, such as spatial variability study, a freeze-up experiment and an angular dependence experiment. All these experiments are described here and contribute to the understanding of the radiometers inter-comparison experiments. The radiometer installed on the DMI weather station was used for validating the four most used satellite Ice surface Temperature products for a period of four months. Root mean square differences ranged from 1.4 K to 3.5 K, where the Metop_A AVHRR and MODIS Terra products showed the smallest differences to the radiometer observations. The experiment was the first time with several radiometers and a focus upon SI traceability. The extreme conditions which caused difficulties to several of the instruments and a concluding section provides recommendations for the next field campaigns.

© DMI, 2017

Danish Meteorological Institute
Lyngbyvej 100, 2100 Copenhagen Ø

Extracts from this report may be reproduced provided the source is acknowledged
and the extract is not taken out of context.

Table of contents

1. MOTIVATION	1
2. EXPERIMENTS	2
2.1 PROGRAM	2
2.2 EXPERIMENT SITE	3
2.3 WEATHER CONDITIONS	4
3. INSTRUMENTS	6
3.1 ISAR	7
3.2 KT 15.85 II.....	7
3.3 CS IR120	7
4 RESULTS	8
4.1 INTER-COMPARISON EXPERIMENT.....	8
4.1.1 Introduction	8
4.1.2 Data processing	9
4.1.3 Ice brightness temperatures.....	10
4.1.4 Pairwise inter-comparisons	10
4.1.5 Mean inter-comparison	14
4.1.6 ISAR inter-comparison	14
4.1.7 Relation to NPL laboratory calibration	15
4.1.8 Relation to calculated uncertainties.....	15
4.1.9 Effect from different spectral response functions	16
4.1.10 Sky Brightness temperatures.....	16
4.2 SPATIAL VARIABILITY EXPERIMENT	17
4.3 ANGULAR DEPENDENCE.....	19
4.3.1 Ice surface brightness temperatures	19
4.3.2 Sky brightness temperatures	20
4.4 FREEZE-UP EXPERIMENT	21
4.5 SNOW AND ICE THICKNESS SURVEY	22
4.6 SATELLITE VALIDATION	23
4.6.1 Satellite products.....	23
4.6.1 Matchups.....	24
4.6.2 Validation Results.....	25
5 SUMMARY AND CONCLUSION	30
6 LESSONS LEARNED AND FUTURE WORK	31
7 ACKNOWLEDGEMENTS	33
8 REFERENCES	33
APPENDIX A: DETAILS ON DATA PROCESSING	35
DATA PROCESSING FOR ANGULAR DEPENDENCE EXPERIMENT	35
QUALITY CONTROL FOR SATELLITE MATCH-UP.....	35

1. MOTIVATION

Ice surface temperature validation work and SI traceability is a new and only little explored research area. Hence, the scientific potential for even a relatively small research campaign is large. Although challenging, it becomes increasingly important that efforts are made to establish Fiducial Reference Measurements (FRM) of sea Ice Surface Temperature (IST)¹. Few field campaigns have been conducted in the high latitudes with the focus on IST and SI traceability due to the extreme conditions. The establishment of IST FRM is therefore much less mature than for SST and LST. As IST is often significantly colder than most of the other surface temperature measurements from sea or land, it is important to ensure that Infrared (IR) radiometers are validated appropriately for these conditions.

In a recent user requirement survey carried out within the ESA CCI sea ice project, ice surface temperature was ranked 4th out of 22 parameters in terms of how important the users considered the variables. High latitude In situ observations play an important role in the development and validation of satellite surface temperature retrieval algorithms which typically have problems in Arctic conditions (Høyer et al., 2012). In addition, in situ observations are used to quantify the uncertainty of the satellite observations in the context of SI traceability. However, the IST observations that are available in the operational data stream come typically from buoys placed on the sea ice. As it was shown in Dybkjær et al. (2012) these data are very sparse and need manual quality control in order to ensure they represent the ice surface temperature. In addition, none of the buoys are recovered and post calibrated and the SI traceability is therefore lost or at best unreliable. Finally, the vertical temperature gradients in the snow are very high, as snow is an effective insulator (Tonboe et al., 2011), and during the experiment there were differences of over 10 K between the snow and sea ice interface and the snow surface temperature (see Figure 3). Therefore, it is recommended to compare satellite ice surface temperature measurements with in situ surface temperature measurements and not proxies such as 2 m air temperature or the snow temperature measured by buoys buried in snow. The observations from normal contact thermometers used on the buoys rarely observe the actual skin temperature of the surface, which is what the satellites measure after correction of atmospheric effects.

Previous field campaigns carried out by DMI have demonstrated that ice surface temperature observations from infrared radiometers are much more accurate and suitable than conventional temperature observations. Yet, no campaigns have been performed with the aim to inter-compare the FRM TIRs and to assess the uncertainties between different radiometers.

This report describes the work, which has been carried out as one of the tasks: “Field Inter-Comparison Experiment (FICE) for Ice surface temperature” carried out in the FRM4STS project (www.FRM4STS.org) which included the tasks of arranging an IST FICE in Qaanaaq, process the data and present the results in a report.

¹ Throughout this report, IST refers to sea ice surface temperature or snow skin surface temperature when the sea ice is covered with snow

2. EXPERIMENTS

The field program was carried out from March 30th to April 6th, 2016. Three research groups participated with instruments: Danish Meteorological Institute (DMI, Lead), National Center for Oceanography Southampton (NOCS) and Norwegian Meteorological Institute (METNO).

The program for the one week experiment is outlined below.

- **Temporal inter-comparison.** All thermal infrared radiometers observed the same sea ice area continuously for at least 2 days (48 hours)
- **Spatial inter-comparison.** A thermal infrared radiometer was moved within a satellite footprint to observe snow covered sea ice surface temperature variability,
- **Freeze up measurement.** The radiometer observed the surface temperature for 1 day (24 hours) during the freeze up of an open pond carved into the ice to cover a broad temperature range. See the figure 3 below.
- **Angular dependence measurements.** The radiometers observed the sea ice at different incidence angles, to assess the effect of angular emissivity dependence.
- **Snow in situ measurements.** Snow parameters were measured throughout the experiment. Measured parameters were: snow grain size, density, depth, salt content of the forming sea ice and thermodynamic temperature

2.1 Program

Table 1 below shows the measurement program for the field experiment

Table 1: Program for the field experiments

	Morning, AM	Afternoon, PM
Wed. 30.03		Arrival at 16:55 local time. Unpack and prepare, boxes from the airport and storage.
Thu. 31.03	Meeting with local hunters for planning of experiment set-up, transport and security.	18:30-21:00 UTC on ice installation of instruments for the inter comparison experiment. Preparation for freeze-up experiment. Download of weather of station data.
Fri. 01.04	13:15-17:30 UTC snow and ice thickness survey	Inter-comparison test phase extended due to problem with some instruments.
Sat. 02.04	13:00-18:30 UTC spatial variability experiment. Snow and ice thickness survey.	Angles on all instruments adjusted. Continuation of inter-comparison with ISARs and IR120.
Sun. 03.04	13:00-17:52 UTC spatial variability experiment. Snow and ice thickness	Angles on all instruments adjusted continuation of inter-comparison with ISARs and IR120.

	survey.	
Mon. 04.04	13:25-18:56 UTC freeze-up experiment started.	Snow survey. Angles on all instruments adjusted continuation of inter-comparison.
Tue. 05.04	13:10-18:30 UTC End of inter-comparison and start of angular experiment with ISARs. Continuation of freeze-up experiment. Snow survey and spatial variability experiment.	End of experiment at 17:10 UTC. Deployment and download.
Wed. 06.04	Post – processing and packing	Departure 13.35 local time.

More information can be found in the log books, which are available on request.

2.2 Experiment site

The experiment site was in Inglefield Bredning off Qaanaaq in the northwestern Greenland. The site is well suited for conducting a field campaign on the sea ice, measuring the ice surface temperature with radiometers. It is well within the high Arctic at 77N with a dry Arctic atmosphere and cold temperatures in April. The position of the experiment is shown in the figure below (Figure 1).

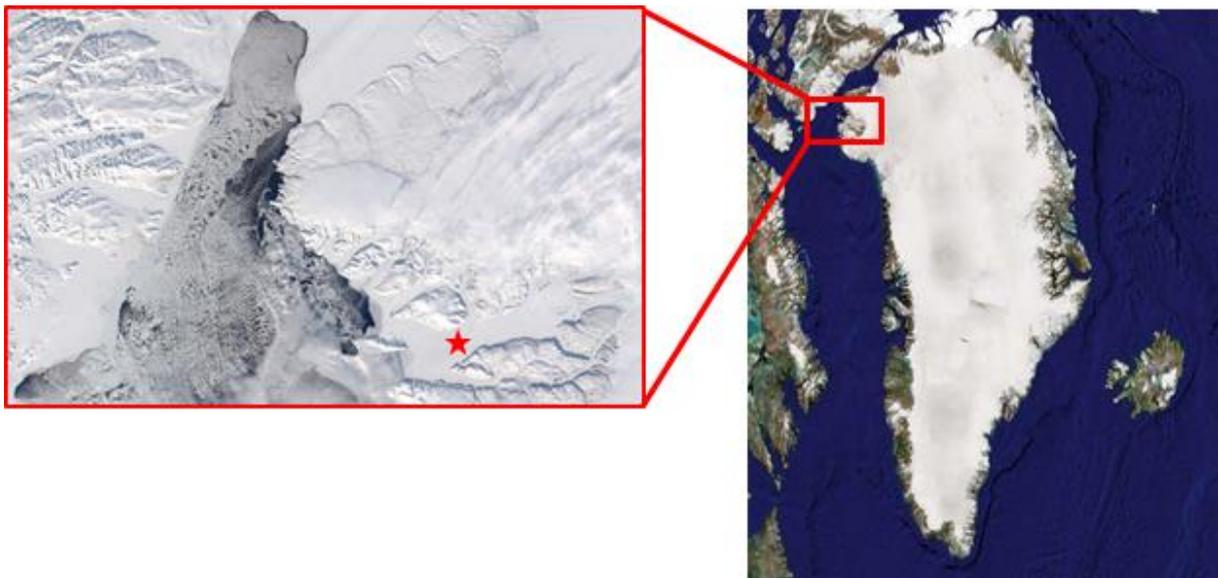


Figure 1: Location of the FRM4STS IST FICE. Qaanaaq is situated in Northwest Greenland at 77°N. The experiments were carried out on the sea ice off Qaanaaq in Inglefield Bredning (left) marked with a red star.

The experiment site at 77.34N; 69.14W is 3-4 km from the coast at a deep water site (about 800 m depth). The sea ice was about 1 m thick on average with snow on top. The average snow depth on the site was 9 cm distributed in small barchan dunes and around smaller ice deformation features. The ice was level only with small deformation features and no major overshoot features or tilted blocks.

There was no electricity at the site and all instruments needed to use batteries. A generator was used for battery charging during shorter time intervals while working on site. The weather station has, in addition to the battery, a solar panel for charging the weather station battery. Modest power

consumption is in general an advantage for radiometers operating on sea ice. Sea ice is located in remote areas with extended periods of complete darkness during winter.

2.3 Weather conditions

The DMI Automatic Weather Station (DMI AWS) on the site was installed on January 10, 2016 (see Figure 2 on the right) and collected a complete record of meteorological parameters including the snow surface temperature (see instrument description). Air temperatures during the experiment were about -18°C with a few nighttime temperatures colder than -25°C (see Figure 3). Ice temperatures measured by an Ice Mass Balance buoy (IMB) with a 2 cm resolution are shown in Figure 4. The thermistor string is placed in the sea ice and extends about 0.8 meters out of the ice into the air. The IMB shows the differences in variability within the air and snow/ice, where the typical snow and ice temperatures were in the range between -15°C in the top snow and -1.8°C at the ice water interface. The

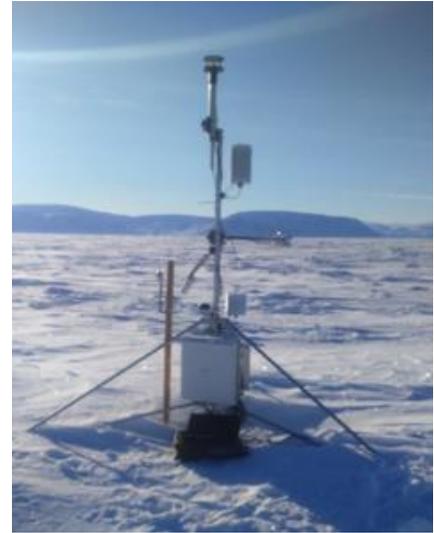


Figure 2: Automatic Weather Station

weather conditions are typical for this time of the year, which is a transition season with both day and night and a pronounced daily cycle in the incoming radiation and the surface temperature. During the experiment the diurnal variability was up to 10°C . The Arctic conditions are extreme for most instruments but the weather was in general favorable for conducting the experiments. Clear sky conditions dominated on the 02.04.2016-03.04.2016, while the remaining days were partly cloudy (see the incoming short wave radiation measured at the DMI AWS shown with the blue curve in Figure 5). Wind speeds were weak and mostly below 5 m/s, with a maximum at around 8 m/s at night and in the early morning of the 02.04.2016 (see the wind speed measured with the DMI AWS shown with the red curve in Figure 5). A small snowfall event occurred on 04.04.2016 at 18.26 UTC resulting in 0.5 cm of new snow on the surface. There was no significant wind during or after the snow event and the snow was not redistributed by wind.

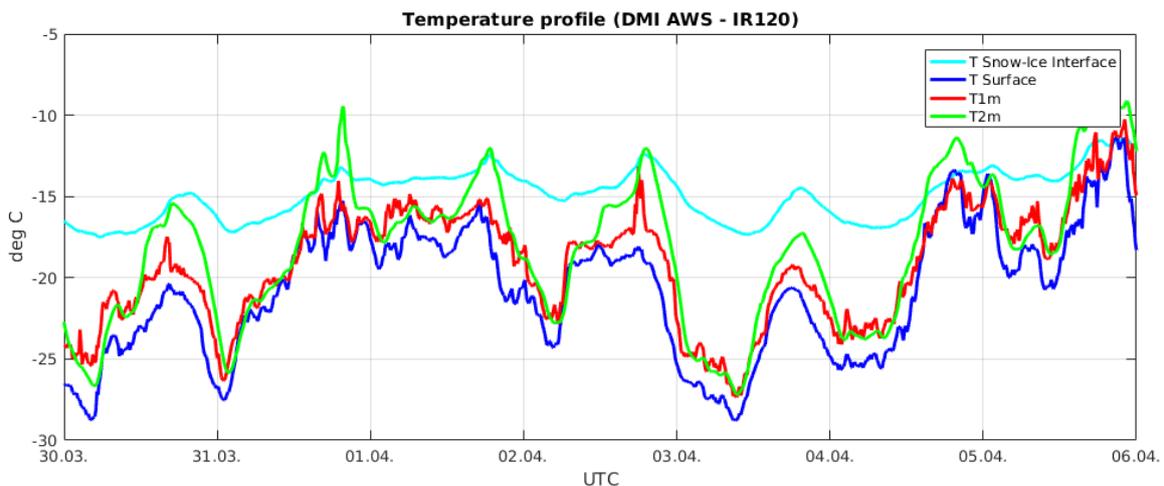


Figure 3: Temperature conditions at different vertical levels during the experiment, as observed by the DMI AWS. The snow-ice interface was under 9 cm of snow.

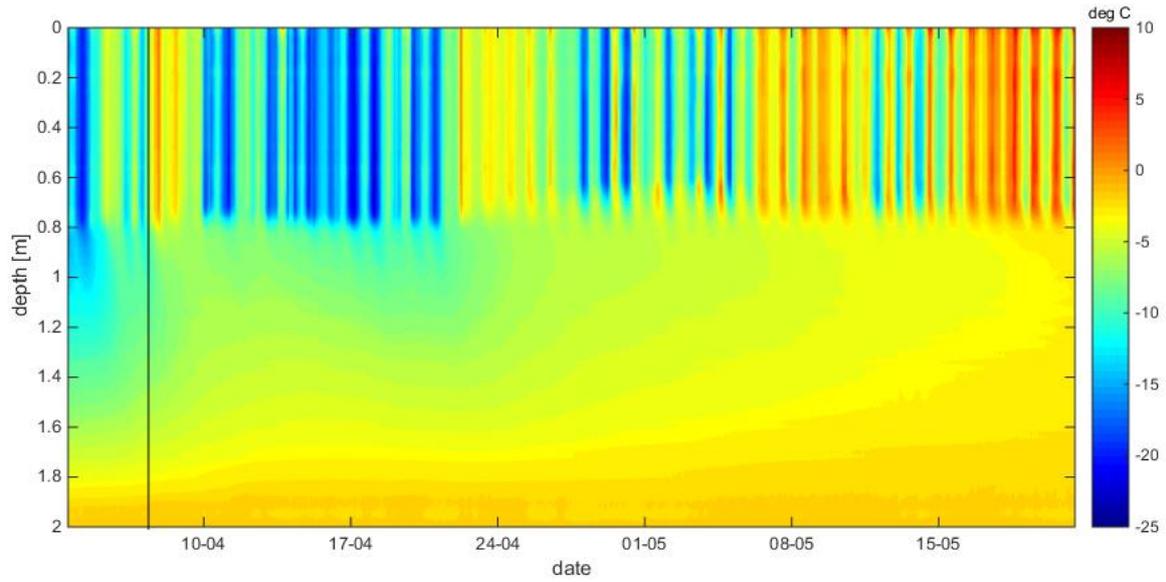


Figure 4: Temporal evolution of vertical temperature profile from the Air through the ice in the water, measured by an Ice Mass Balance Buoy (IMB) at the site from April 3rd to June 7th, 2016. The black vertical line indicates the end of the field campaign.

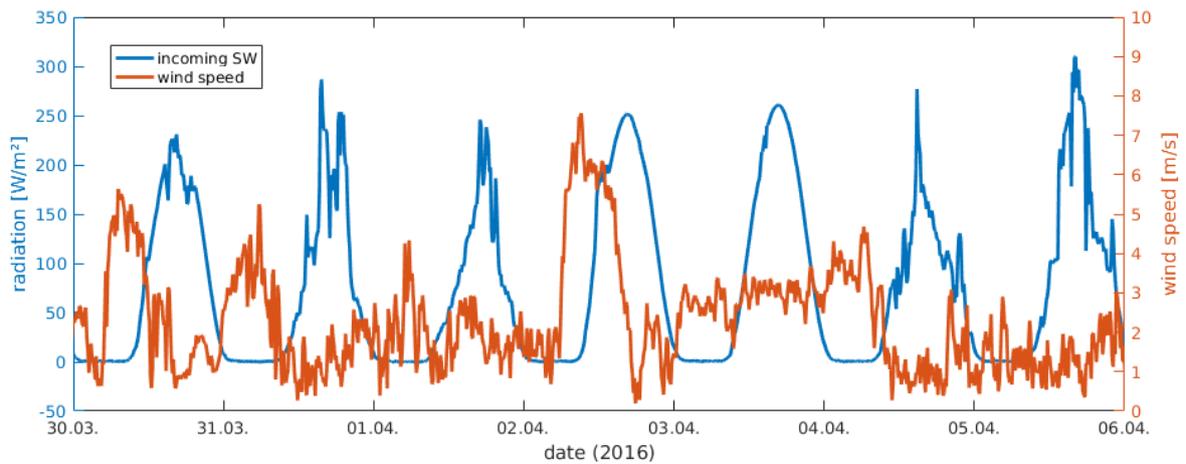


Figure 5: Weather situation during the inter-comparison experiment, including the incoming short-wave radiation (blue) and wind speed (orange).

3. INSTRUMENTS

The three research groups participating in the field campaign brought several instruments to the experiment, measuring the thermal infrared radiation. An onsite weather station measured the surface energy balance and the weather conditions and an Ice Mass Balance buoy (IMB) measured the internal snow and ice temperature profile.

The main instruments used for the fields campaign were two ISARs (from DMI and NOCS), one Heitronics KT 15.85II (DMI), one mobile Campbell Scientific IR120 and one weather station with a Campbell Scientific IR120 and up- and downward looking global radiative fluxes (METNO).

In addition to these instruments, DMI deployed an AWS in the ICE-ARC project, measuring: temperature at several levels (snow-ice interface, 1 and 2 m height over snow surface), radiometric surface temperature (IR 120 radiometer), up and downward looking shortwave and longwave radiation flux, air pressure, relative humidity and wind speed at 2 m height. Ice Mass Balance (IMB) buoys were deployed in the ice, measuring the temperature every 2 cm from the air through the ice and into the water. In addition, ice tethered ocean buoys were deployed in the water, measuring the temperature, salinity and ocean currents in the water under ice.

The AWS, IMBs and ice tethered ocean buoys were installed in January and retrieved in June, thus measuring the conditions for about 5 months.

The different instruments participating in the experiments are listed below with sampling intervals.

Table 2: List of instruments

Radiometer instruments	Institution	Ice sampling rate	Spectral range (μm)	Measured parameters
DMI ISAR08	DMI	2-3 minutes	9.8-11.5	Radiometric IST/Sky temp
NOCS ISAR03	NOCS	2-3 minutes	9.8-11.5	Radiometric IST/Sky temp
KT15.85 II	DMI	1 sec	9.6-11.5	Radiometric IST
MET-WS IR120	METNO	1 min	8-14	Radiometric IST
MET-CS-IR120	METNO	1 min	8-14	Radiometric IST
DMI-AWS IR120	DMI	10 min	8-14	Radiometric IST
Other instruments				
DMI AWS	DMI	10 minute		-Wind -Radiation from NR01 (short and long, in/out) -Humidity - T_{2m} , T_{1m} , $T_{\text{snow/ice}}$ - Radiometric IST (IR120, see top of Table 2)
WS	METNO	1 min		-Radiation (long,in)

				- Radiometric IST (IR120, see top of Table2)
IMB	SAMS/DMI	2 hourly		Vertical Snow and Sea Ice temperature (every 2 cm)

A detailed description of the different thermal radiometers is given below

3.1 ISAR

The infrared SST autonomous radiometer (ISAR) is a self-calibrating instrument measuring infrared emission from the ice surface and atmosphere. It employs two reference blackbody cavities to maintain the radiance calibration of a Heitronics KT15.85D radiometer with an accuracy of $\pm 0.1\text{K}$ and measures IR emission in the spectral waveband 9.8-11.5 μm (Wimmer & Donlon, 2012). One measurement cycle views the sea ice for ~ 70 seconds, the sky temperature for about 20 seconds and the two black bodies for about 60 seconds each. One referenced Ice brightness temperature and Ice surface temperature value are calculated for each cycle using post processing software afterwards. The ice and sky incidence angles are 25° from nadir and vertical, respectively. The post processing software also provides uncertainty values for each referenced measurement. These values have been derived for Sea Surface Temperature (SST) according to Wimmer and Robinson (2016).

3.2 KT 15.85 II

The precision radiometer KT15.85 IIP is similar to the one used in ISAR. It measures the thermal IR radiance between 9.6 μm and 11.5 μm and has a temperature resolution of 0.03 K with an accuracy of ± 0.3 K (See e.g. Theocharous et al., 2010).

3.3 CS IR120

The CS IR120 is an infrared radiometer from Campbell Scientific, with a thermopile that detects the thermal radiation in the 8-14 μm spectral range with an aperture of 20° . The calibrated range for the CS IR120 is from 25°C below to 25°C above body (ambient) temperature with an operating range from -25 to $+60^\circ\text{C}$. The accuracy for of the target temperatures against a blackbody is estimated to 0.2 $^\circ\text{C}$ (See: IR100/IR120 Infra-red Remote Temperature Sensor, user manual (2015)).

4 RESULTS

4.1 Inter-comparison experiment

4.1.1 Introduction

The main experiment of the field campaign was the inter-comparison experiment, where all six radiometers observed the same sea ice area simultaneously for both day and night time conditions. All radiometers thus participated in this experiment. The experiment was set up 40 m to the North of the DMI AWS. The snow surface was visually homogeneous, however with some structure, as can be seen around the site (see Figure 6 for the setup).

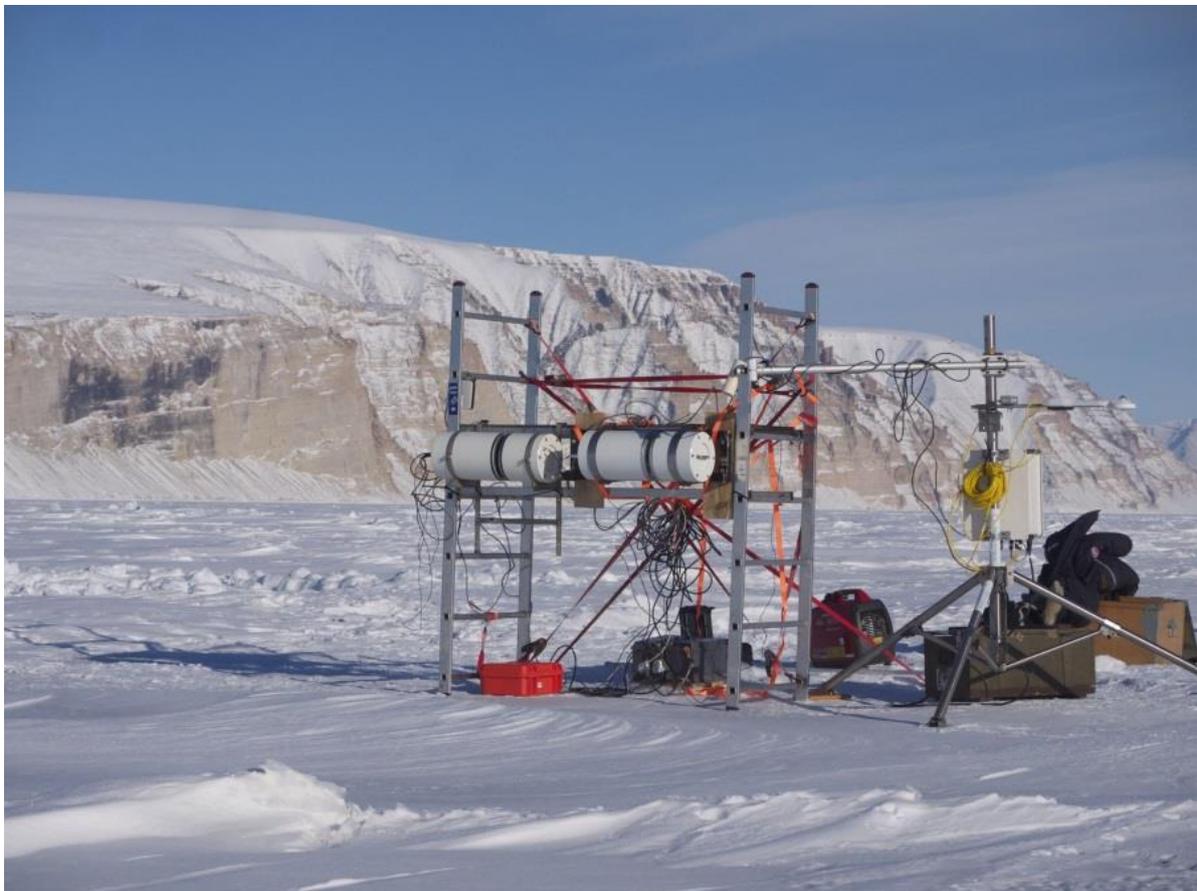


Figure 6: Photo of the inter-comparison experiment, provided by W. Wimmer.

For the comparison all radiometers were mounted to look at the surface with an incidence angle of 25 degrees off nadir. Similarly, sky temperatures were measured by the ISARs with a 25 degree angle off zenith. The rack with the two ISAR instruments was moved two times by about 40 cm during the inter-comparison experiment to look at three different surfaces (see the log-book for more details). The move ensured that the two ISAR radiometers looked at new surfaces three times for extended

time periods during the comparison and to test whether potential differences were due to the radiometers looking at different surface conditions. The measurements were done on a small optically uniform barchan snow dune composed of wind packed snow with a high density and small and mechanically broken down snow grains with some hoar in patches at the surface (snow depth: 6 cm, snow density in the upper 6 cm: 332 kg/m³, surface snow grain size: 0.5-1.2 mm). This was visually a very typical surface for the site. The overall surface (measured in the upper 6 cm of the snowpack with a 5 cm snow density shovel) snow density for the area on average was 340 kg/m³ (min: 200 kg/m³, max: 468 kg/m³) and the average snow depth was 9.1 cm (min: 0.5 cm, max: 61 cm). Following a wind event on March 15, 2016 with about 8 m/s, some hoar crystals had formed on the surface and a weak/low density layer of depth hoar had formed above the snow - ice interface. The measuring time for each radiometer in the inter-comparison experiment is shown in Figure 7.

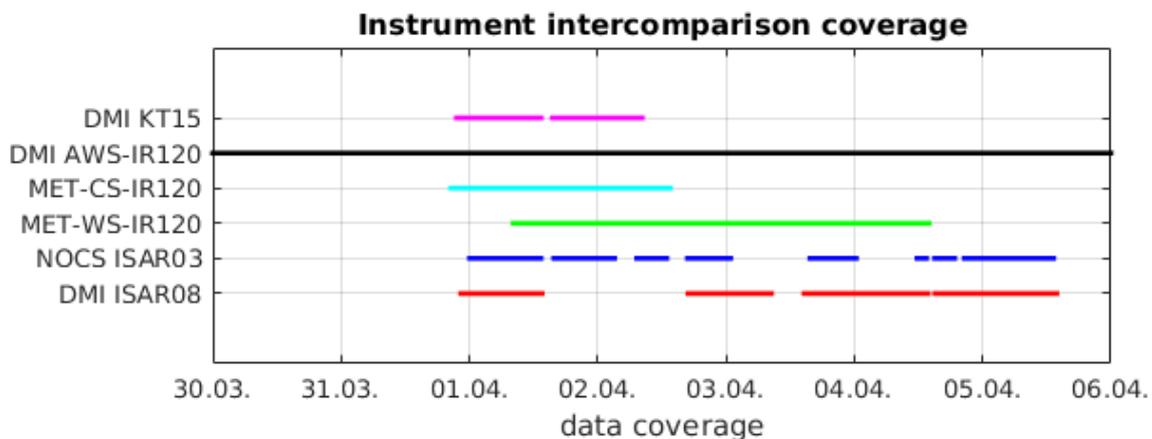


Figure 7: Temporal coverage of the radiometers participating in the inter-comparison experiment

The cold and remote conditions are challenging for the electronics in radiometers and the capacity of the batteries and this put a limitation on several of the instruments. The ISAR03 from NOCS was shutting down when the temperature dropped below -22°C, which happened during two nights on the 3rd and 4th of April. The ISAR08 from DMI has gaps in the record due to power loss and the DMI KT15 broke after 1.5 days into the inter-comparison exercise. This was not a surprise as data gaps are expected when conducting field work in extreme conditions. Except for the DMI KT15, which failed and the MET-CS IR120, which was used for other experiments, all the radiometers measured the same ice surface for at least 48 hours and at least two radiometers were functioning at all times during the inter-comparison experiment.

4.1.2 Data processing

The radiometers used in the inter-comparison experiment have different sampling rates dt , from 1 second to 10 minutes (see Table 2). To perform an inter-comparison, it is therefore necessary to adjust them to a common time sampling to allow for a statistical comparison of the different datasets. Here, all ice surface temperature (IST) data records have been averaged to the sampling rate $dt = 10$ min, which is the sampling rate of DMI-AWS IR120 and at the same time the lowest sampling rate of the six radiometers. The datasets with a higher temporal sampling were averaged to this sampling rate calculating the 10 min average centered on the DMI-AWS IR120 intervals. The two ISAR datasets with sampling rates of several minutes were first linearly interpolated to a temporal sampling of 1 minute and subsequently averaged to the 10 minutes sampling intervals.

Tests were carried out with an averaging interval of 5 minutes for all the radiometer records, but the results were very similar to the 10 minute results. We therefore chose a sampling rate of $dt = 10$ min to have the same sampling times as the DMI-AWS IR120.

4.1.3 Ice brightness temperatures

As Figure 7 shows, the DMI-AWS IR120 (black line) is the only radiometer that recorded for the full period of the inter-comparison without gaps. The only time span where all six instruments were operating simultaneously, comprises around 3 hours on April 1st. The varying number of available records complicates the inter-comparison, but at least two radiometers were operating at any time. Figure 8 shows the brightness temperature measurements from the six participating instruments during the inter-comparison experiment. It is evident that the instruments agree well when temperatures do not exhibit much variation, while they differ most for relative temperature maxima (esp. on the 4th of April) and minima (2nd and 3rd of April). As mentioned before, the NOCS ISAR03 has difficulties resolving very low temperatures (< -22 °C) and cuts off the data record. Although of the same instrument type, DMI ISAR08 and NOCS ISAR03 measurements have an offset of approximately 0.9 K, which is far more than the calculated uncertainties of around 0.13 K could explain. Further, DMI ISAR08 exhibits a higher variability than NOCS ISAR (with standard deviations of 2.72 K and 1.76 K, respectively).

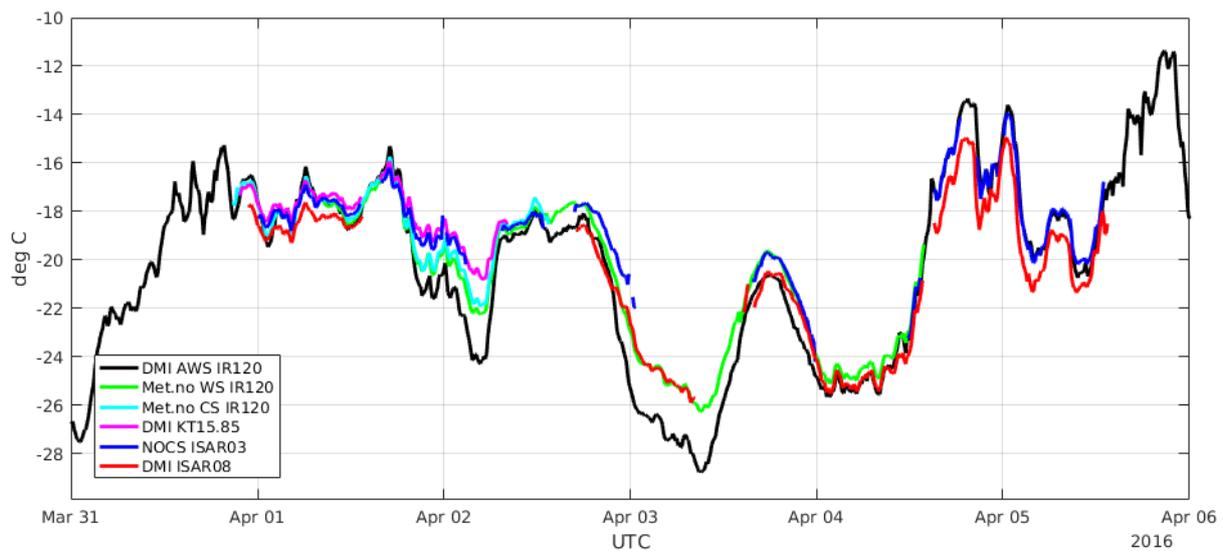
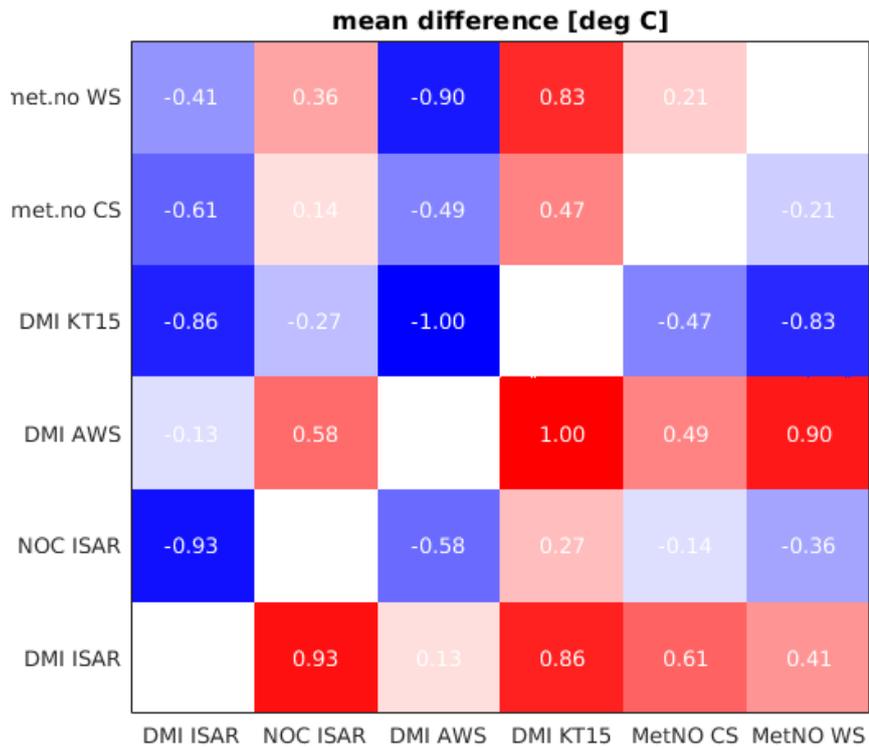


Figure 8: Time-series of brightness temperatures observed by the 6 radiometers participating in the inter-comparison experiment.

4.1.4 Pairwise inter-comparisons

It is evident from Figure 8 that it is difficult to make an average IST curve for all radiometers that the individual radiometers can be compared against. The many data gaps in the different observational records make it impossible to create an average IST time series with uniform error statistics throughout the experiment. Instead, pairwise inter-comparisons of all the combinations of radiometers were performed for the 10 minute observations. Figure 9 shows the number of pairs, the mean difference and the standard deviation of the differences for all the combinations of instrument pairs. The mean differences between the datasets range between 0.14-1.0 K, depending on the respective instrument pairs; best agreement is found between the METNO instruments as well as between DMI's KT15 and the NOCS ISAR03. In contrast, the mean difference between DMI-

AWS IR120 with DMI's KT15 is, with 1 K, highest between all pair-constellations. The two ISARs correlate quite well, but have an offset of 0.93 K on average, with DMI ISAR08 reading colder than NOCS ISAR03. The standard deviation of the differences is highest between any instrument and the DMI-AWS IR120. This is probably due to the low temporal sampling that makes this data record a snapshot every 10 minutes, whereas the other instrument records are all averages over a 10 minute interval. The DMI KT15 and DMI ISAR08 differences exhibit the lowest standard deviation (0.05 K). These two instruments were pointing at nearly the same spot on the surface.



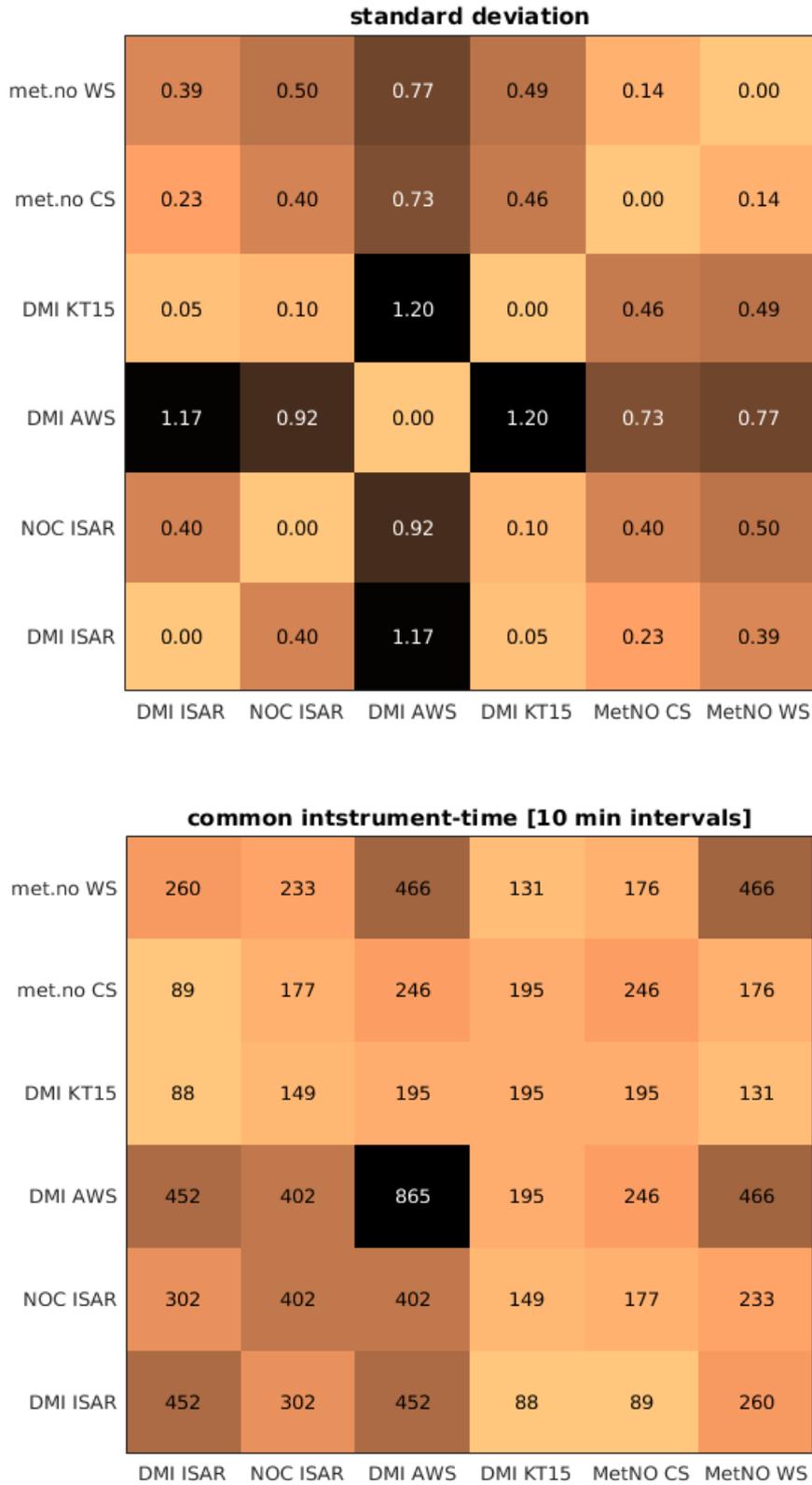


Figure 9: Pairwise statistics of the 6 instruments used for the inter-comparison project: (a) mean difference, (b) standard deviation of differences and (c) common time steps. Read signs of mean difference from x to y, e.g. NOCS ISAR is on average 0.93 °C warmer than DMI ISAR.

4.1.5 Mean inter-comparison

As shown above, the data-sets are partly fragmented and contain gaps; of the total 865 time steps during the inter-comparison period, common record time for certain instrument pairs (DMI KT15 or METNO CS with DMI ISAR08) comprises only 89 time steps (corresponding to less than 15 h). Hence, the respective mean to compare with critically depends on single instruments, suggesting that such a comparison may not be optimal. Instead we compute the statistics for each instrument, where each pairwise inter-comparison has equal weight. The difference of each participating instrument's IST measurements from the mean of all other instruments' measurements is shown in Figure 10 summarizing that the DMI KT15 display the largest differences compared with the other instruments, with an average of 0.69 K warmer, whereas the METNO WS is on average 0.09 K colder than the other instruments.

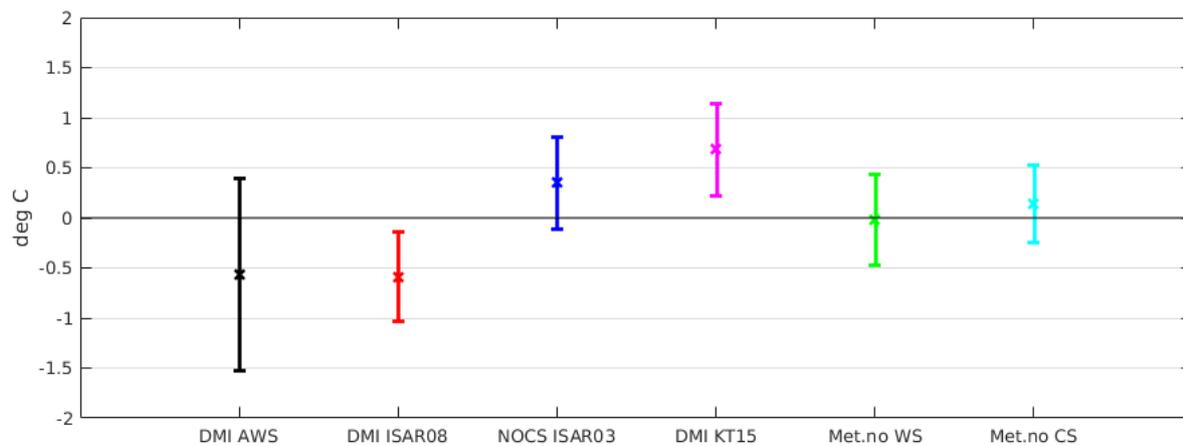


Figure 10: Mean difference (cross) and mean standard deviation (errorbars) of each radiometer to the remaining instruments, with each radiometer equally weighted (irrespective of common instrument time).

4.1.6 ISAR inter-comparison

The two ISARs are the only instruments in field campaigns that are self-calibrating and correcting for the reflected sky radiation. The accuracies are therefore expected to be higher than the other instruments. The scatter plots in Figure 11 shows the ice surface temperatures (corrected for sky effects) from DMI and NOCS. It is evident that the significant bias between the DMI and the NOCS ISAR is present for the full temperature range observed. In addition, the color coding that represents the incoming solar radiation (left) and the observed wind speed (right) show no systematic dependency of these quantities in the inter-comparison.

A post deployment calibration experiment revealed a misalignment of the DMI ISAR08 scan drum encoder, which may have resulted in a mispointing of the radiometer. In addition, the window had to be replaced due to contamination effects. These effects have probably resulted in the DMI ISAR08 not performing to its specifications, and may explain the ISAR differences.

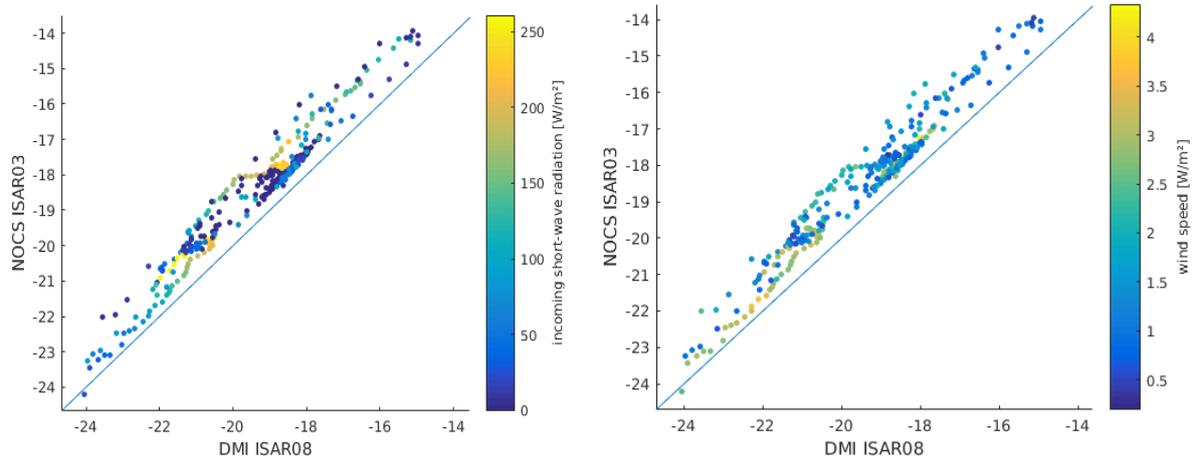


Figure 11: Scatter-plots of ice surface brightness temperatures from the DMI ISAR08 against NOCS ISAR03. The colourbar denotes the incoming SW radiation (left) and the wind speed (right).

4.1.7 Relation to NPL laboratory calibration

Both ISARs participated in the NPL 2016 laboratory radiometer comparison in June 2016, after the IST FICE (Barker-Snook et al., 2016). The experiments were conducted at room temperature (about 25°C) and viewed blackbody targets with temperatures between -30 and +45°C. Unfortunately, no comparison was available for the DMI ISAR08 at -30°C, but the NOCS ISAR03 showed a warm bias of 1.3K when viewing the -30°C reference target. At -15°C, the DMI ISAR08 overestimated with 0.05 K and the NOCS ISAR03 with 0.4 K. As discussed in the Barker-Snook et al., 2016 report, elevated errors are expected when the internal reference blackbodies are far away from the target temperatures. The performance of the ISARs is likely to improve when the target and ambient temperatures have similar temperatures. It is, nevertheless, interesting to note that the NPL lab experiments confirm what is seen from this inter-comparison experiment, that the NOCS ISAR03 is warmer than the DMI ISAR08 for cold temperatures and that biases can be more than one degree Kelvin. A calibration experiment could be conducted in cold air temperatures to assess the true performance of the radiometers in high latitude conditions. This is discussed in section 6.

4.1.8 Relation to calculated uncertainties

The standard ISAR processing software provides an uncertainty estimate, which has been derived for SST radiometer observations (Wimmer and Robinson, 2016). The minimum, maximum and average uncertainties for the two ISARs during the IST-FICE measurement campaign are shown in Table 3.

Table 3: Uncertainties for the ISARs during the inter-comparison campaign.

	Minimum uncertainty (K)	Maximum uncertainty (K)	Mean uncertainty (K)
NOCS ISAR03	0.12	0.31	0.13
DMI ISAR08	0.12	0.20	0.13

With an observed average difference of 0.93K between the two instruments, it is clear that these uncertainties do not apply to the actual IST observations in this experiment. The reason for the inconsistency is probably due to the misalignment error in the DMI ISAR08 explained above and the fact that these uncertainties have been derived for SST conditions. More work is therefore needed to assess the true uncertainty budget for the radiometers when measuring IST in cold conditions and to assess the validity of these values when the ISAR instruments are performing to specifications.

4.1.9 Effect from different spectral response functions

The emissivity of the ice surface differs between the Heitronics KT15s and the Cambell Scientific IR120 radiometer observations due to differences in the spectral windows (9.6-11.5 μm versus 8-14 μm , respectively). The difference in emissivity will lead to a difference in actual observed brightness temperature as the reflected sky temperature tends to be much colder than the ice surfaces. These differences are thus included in the pairwise ISAR vs. CS IR120 inter-comparisons. To assess the impact of this effect, an infrared snow and ice emissivity model was run with the spectral response functions for the two types of instruments. This resulted in averaged emissivities of 0.998 for the KT15 and 0.996 for the IR120 spectral windows for a typical snow surface and an incidence angle of 25 degrees. Using realistic sky and snow surface temperatures, the maximum difference in the observed brightness temperatures was calculated to be 0.06 K, with CS IR120 being colder than KT15.

4.1.10 Sky Brightness temperatures

Figure 12 shows surface brightness temperatures (dashed lines) together with sky brightness temperatures (solid lines). METNO and DMI AWS NR01 sky brightness temperatures T have been derived using the Stefan-Boltzmann-law

$$j = \sigma T^4$$

with the Stefan–Boltzmann constant $\sigma = 5.670373\text{E-}8$ and the incoming radiation j in W/m^2 . Note that ISAR instruments only measure wavelengths of 9.6-11.5 μm , whereas the sky temperatures derived from longwave incoming radiation from DMI and METNO WS have a much wider spectral band-width (from 4-40 μm for the DMI NR01 Hukseflux radiometer). The difference in spectral windows explains the difference in calculated sky brightness temperatures as seen in Figure 12. The ISAR instruments effectively measure within an atmospheric window, which is not the case for the longwave incoming measurements, hence, the higher temperatures of these instruments. Note that it appears that the ISAR instruments reach their lower limit and fail to properly operate at low temperatures in clear sky conditions (between 2nd and 4th of April), where the data shows sky temperatures of -100 °C.

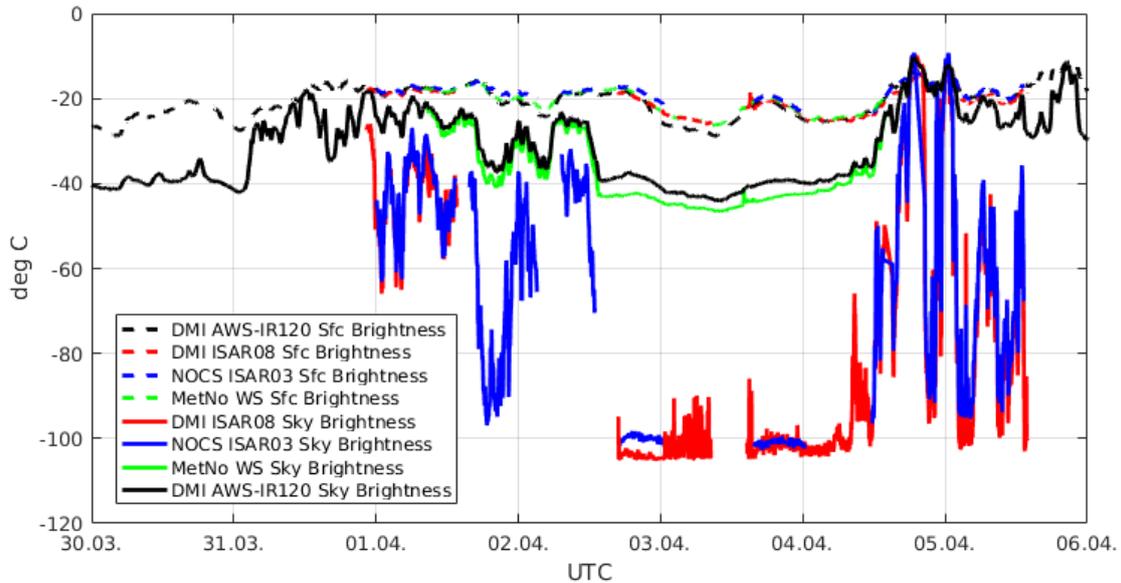


Figure 12: Sky (bold lines) and surface (hatched lines) brightness temperatures for the ISARs and derived from longwave incoming radiation observations.

4.2 Spatial variability experiment

The goal of this experiment was to quantify the spatial temperature variability within one satellite footprint centered on the experiment site with the weather station. In the first part of the spatial variability experiment the MET-CS IR120 radiometer was mounted on a sledge and a 1 km² area centered on the weather station was surveyed twice in a star-shaped pattern with cross-overs near the center. Due to the relatively high variability due to uneven terrain, it was realized that much of the observed variability could have been caused by tilting of the radiometer and local incidence angle variability and the experiment was repeated, with a focus upon minimizing the variability in the sledge tilt and the radiometer incidence angle.



Figure 13: Performing the spatial variability experiment, provided by W. Wimmer.

The two experiments were conducted within a 1 km² area around the main experiment site and comprise measurements within one hour in the early afternoon of April 2nd (part 1) and the morning of April 3rd (part 2), respectively. The track-paths including time-series of the measured IST using the MET-CS IR120 are shown in Figure 14.

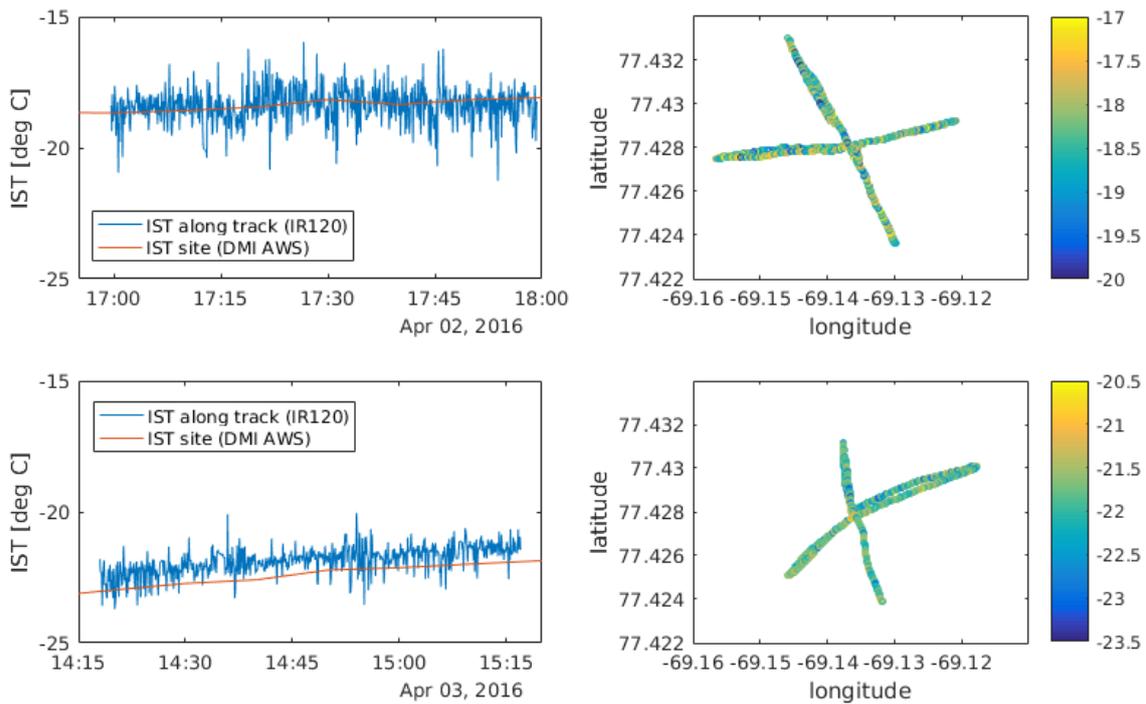


Figure 14: Spatial variability experiment, part 1 (top) and part 2 (bottom). Tracks are indicated on the right with the color-scale showing the de-trended temperatures in degree Celsius

Since – especially during the 2nd part of the experiment – the ice was warming substantially, the time series have been detrended before calculating statistics. The mean difference between DMI-AWS IR120 and MET-CS IR120 on April 3rd is around 0.5 K, while both instruments agree well on April 2nd. Note however that the DMI-AWS IR120 has a sampling rate of $dt = 10$ min, and therefore contains only 6 data-points during the experiment. Variability in terms of the standard deviation ranges between 0.69 K in the first and 0.42 K in the second part of the experiment. The spatial variability was calculated as the standard deviation over 100 m-track (detrended) IST averages (see Figure 15).

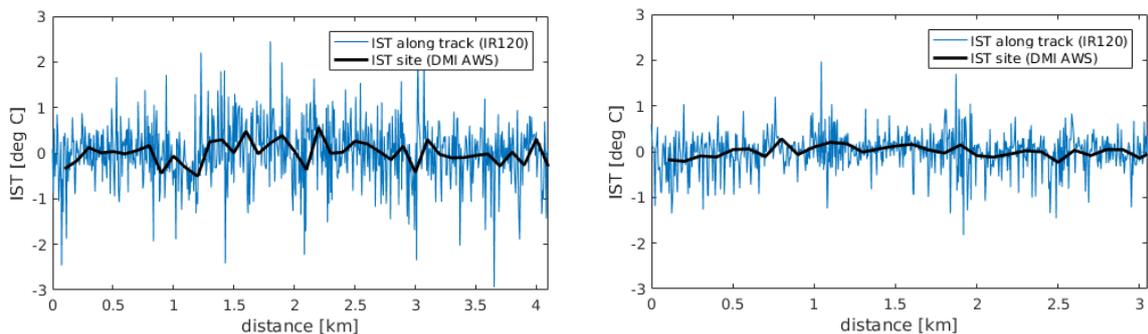


Figure 15: Original (blue) and low-pass filtered (black) IST using a 100 m averaging window for part 1 (left) and part 2 (right) of the spatial variability experiment.

The statistical numbers for the two sledge dragging experiments are shown in Table 4. The reduced variability in the second experiment due to the minimization of the sledge tilt is seen as a reduction of the IST variability from 0.69 K to 0.42 K. The significant contribution from the radiometer incidence

angles to the IST variability is in agreement with the results from the angular dependence experiment outlined in Section 4.3, where the sensitivity of the IST observations to the viewing angle can be up to 0.5 K.

Table 4: Statistics for the spatial variability experiment

	Nobs	Distance (km)	Duration	Stdv (K)	Bias to DMI-AWS IR120 (K)	Spatial stdv (K)
Part 1 (Apr-02)	718	4.08	00:59:45	0.69	-0.01	0.25
Part 2 (Apr-03)	709	3.04	00:59:00	0.42	0.50	0.12

4.3 Angular dependence

This experiment was designed to assess the effect of emissivity angular dependence using the two ISAR instruments. The experiment was started Tuesday April 5th and both instruments were programmed for an angular measuring cycle surveying a small barchan dune with optically uniform conditions and corresponding temperatures for sky.

The experiment was conducted using the DMI ISAR08 and NOCS ISAR03 for the angles 25, 45 and 55 degrees towards sky, as well as the angles 155, 135 and 125 towards the surface. 0 degrees correspond to zenith. DMI ISAR08 measurements for angle 45 were missing (see Appendix A.1).

4.3.1 Ice surface brightness temperatures

As Figure 16 shows, the observed ice surface brightness temperatures are sensitive to the instrument angle. The flattest angle of 125° from zenith (55° ice incidence angle) yields lower temperatures than the steepest angles of 155° (25° ice incidence angle). However, the temperatures from a 135° scan angle in the DMI ISAR08 observations exhibit higher values than the flatter 125° angle until 16:00 UTC, whereas after this time, those temperatures are found within the same range as 125° scan angle temperatures. In contrast, temperatures measured with NOCS ISAR03 show a clear dependence on the incidence angle: The flatter the angle towards the surface, the lower the KT15 brightness temperature. The observations with incidence angles of 55° from nadir were in general 0.25-0.5K colder than an incidence angle of 25°, but the differences can be more than 1K. The reason for the differences between the two ISARs might be due to the misalignment of the DMI ISAR08.

The instrument rack was moved around 16:10 UTC, which might explain the change in ice relative surface temperature patterns after that time.

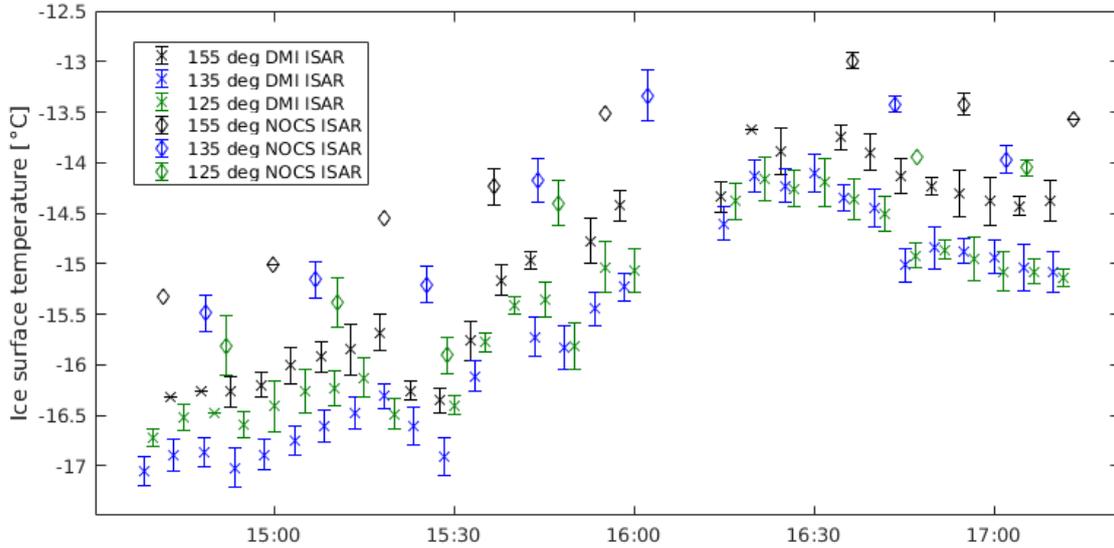


Figure 16: Brightness temperature measurements of the ice surface from the DMI ISAR08 (x) and NOCS ISAR03 (diamonds) for the measurement angles 125, 135 and 155 degrees from zenith. Error bars indicate the variability during the measurement cycle.

4.3.2 Sky brightness temperatures

The corresponding ISAR observations from different sky view angles are shown in Figure 17. It is seen that the temperatures are critically dependent on the angle with a difference of 6 K for an angle difference of 30 degree, with lower temperatures for angles closer to zenith. This is due to the smaller atmospheric path distance and thus colder sky temperature measurements. Note that the standard deviations of the NOCS ISAR03 sky temperatures are significantly higher than those from the DMI ISAR08.

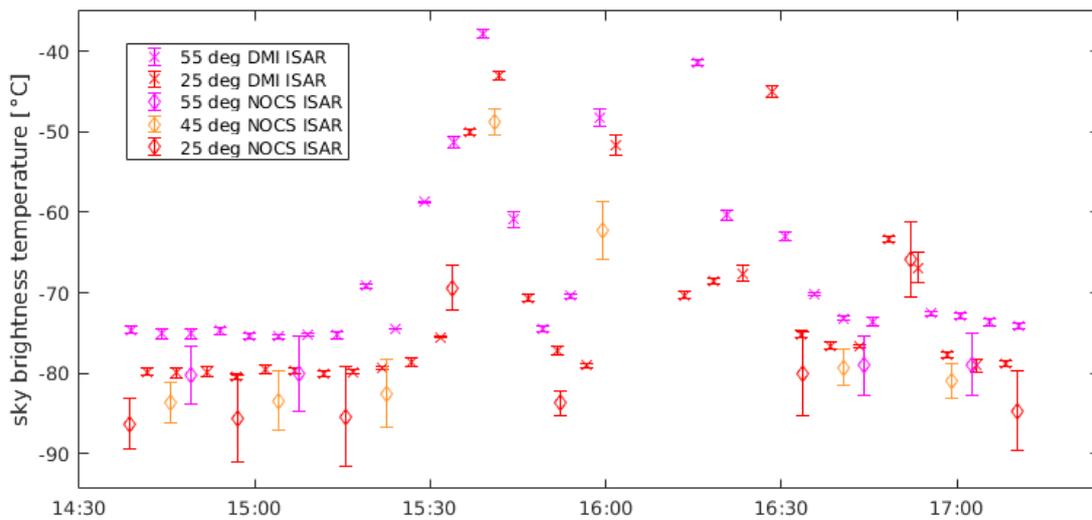


Figure 17: Same as Figure 16, but for upward looking angles, relative to zenith (25, 45, 55 degrees).

4.4 Freeze-up experiment

The experiment started on Monday morning April 4th using the MET-WS IR120 + flux measuring the surface temperature of the rapidly freezing pond. In the first attempt, a large basin was drilled out in the sea ice with an ice drill Thursday 31st April. The bottom of the basin consisted of about 20 cm sea ice. However, the first basin unexpectedly filled with water overnight without any radiometric measurements. A second smaller basin was therefore dug out Monday April 4th and after removal of ice chips, water was let in while observing the surface with the MET-WS IR120. Ice thickness and salinity measurements were made during the freeze-up experiment.

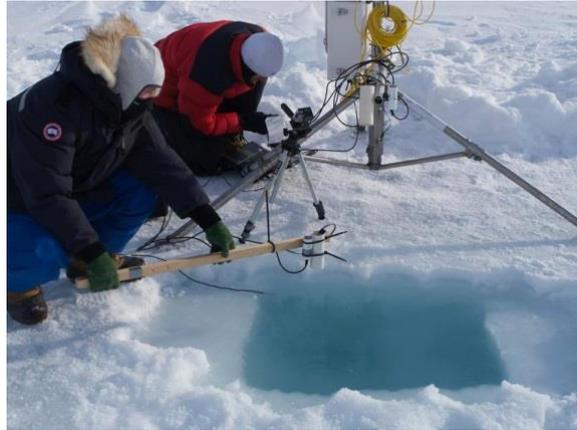


Figure 18: Set-up of the Freeze-up experiment, photo, provided by W. Wimmer.

The freezing pond temperature was measured by the MET-WS IR120 during roughly 24 hours. However, the raw pond temperature below the freezing point at the beginning of the experiment indicates that the IR120 has measured a wider area than the pond only. A simple correction can be calculated, assuming the relative contribution from the ice temperature to be constant and assuming that the initial water temperature in the pond is known. The pond temperature T_{pond} has thus been corrected using the adjacent ice surface temperature T_{AWS} and assuming the corrected pond temperature T_{corr} to lie at the freezing point T_{freeze} at the beginning of the experiment:

$$T_{corr} = T_{pond} - a \cdot T_{AWS}$$

where the correction factor a is $\frac{(T_{pond(1)} - T_{freeze})}{T_{AWS(1)}}$, and $T_{freeze} = -1.7$ °C.

Figure 19 shows the evolution of the corrected pond temperature ($a = 0.13$) together with the pond's ice thickness and the ambient ice surface temperature as measured by the DMI-AWS IR120. Ice thickness in the pond was around 1 cm after two hours and 6 cm at the end of the freeze-up experiment. However, the pond was found to have different ice thicknesses, depending on the exact measurement position. As evident from Figure 19, the pond temperatures are related to the adjacent ice surface temperatures as measured by DMI-AWS IR120 and show a weaker cooling at the beginning due to increasing ambient temperatures, while the pond temperatures drop faster when the ambient surface temperatures decrease too. Within 24 hours, the pond temperature fell by approx. 3-4 K, while the pond's ice thickness grew by 4-5 cm.

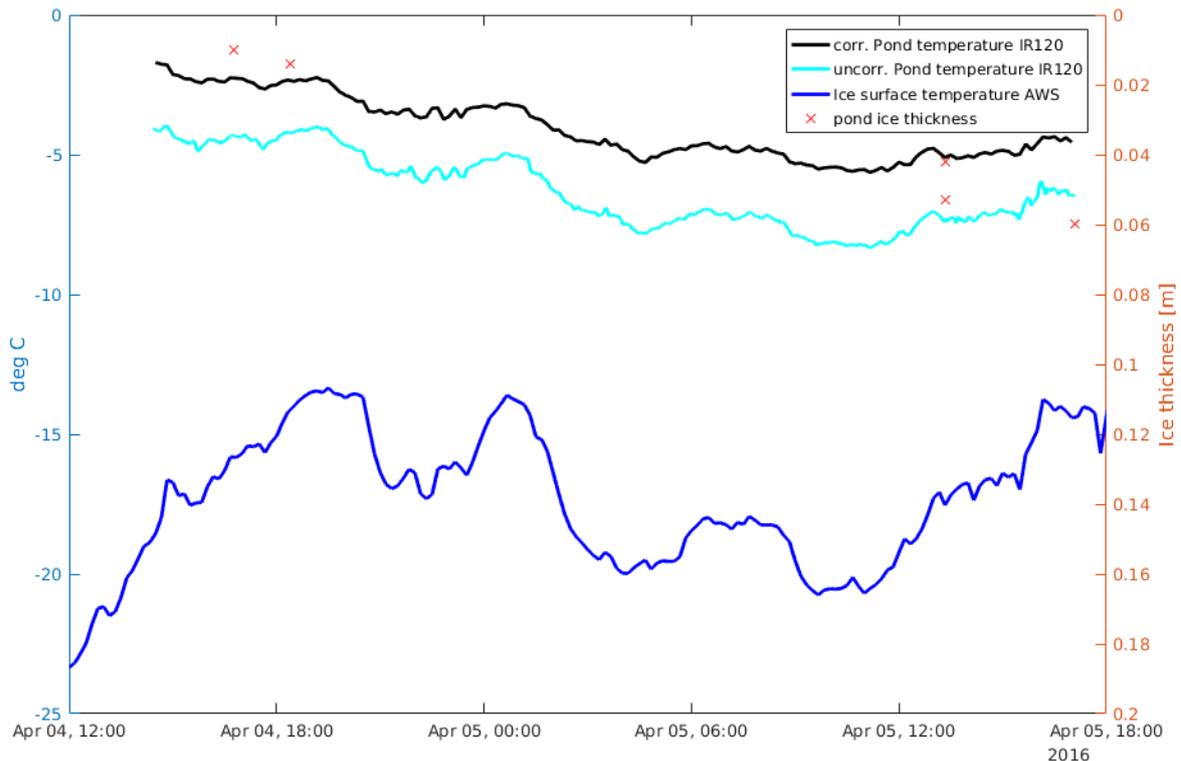


Figure 19: Freeze-up experiment. Uncorrected (light blue) and corrected (black) temperatures of the freezing pond together with site ice surface temperature (dark blue). Ice thickness in the pond is indicated in red asterisks on the right y-axis (reversed).

4.5 Snow and ice thickness survey

The aim of this experiment was to produce a sea ice snow and ice thickness dataset, including snow density and microphysics relevant for sea ice thermodynamics and IR measurements. The snow conditions were measured along 200 m lines for every 10 m starting near the experiment site and pointing in all directions. All the individual measurements are given in the snow survey data sheet, which is available upon requests. At every site the snow grain size was estimated by scraping the surface with a thin plastic sheet with printed millimeter mesh and a magnifying glass. The maximum and minimum grain size of the distribution was estimated. The snow density was measured using a 250 cm³ snow shovel (height 5 cm) whenever possible. The snow density was not possible to measure when the snow depth was less than 6 cm. The way the shovel is constructed means that the center of gravity of the measurement is located at near 4-6 cm from the surface. Occasionally, the density measurement reached into the depth hoar layer located above the ice surface. This resulted in samples with lower density since the depth hoar layer had lower density than the surface layer. The snow depth was measured with a ruler in the same spot as the density was measured and the snow ice interface temperature was measured with two different handheld thermometers. The snow density ranged from 200 to 448 g/cm³ with an average of 340 g/cm³ for the more than 100 samples. The minimum grain sizes were 0.11 mm and maximum 2.44 mm. The average snow thickness was 9 cm lying on top of a roughly 1 m thick sea ice cover.

4.6 Satellite validation

Observations from FRM TIRs are crucial to perform an accurate calibration and validation of the satellite observed IST. The site off Qaanaaq is located about 3-4 km from the coast and is thus well suited for a comparison with thermal Infrared satellite observations with footprints about 1 km. In addition, the high latitude ensures many overpasses during one day and therefore a large sample size available for the validation statistics. Despite the many overpasses, the week-long experiment was too short to perform a proper satellite validation. However, the 4.5 months observations from the DMI-AWS IR120 radiometer were able to give much more satellite versus in situ matches. The inter-comparison experiments have shown that the DMI-AWS IR120 performs similar to the other radiometers but was probably ~ 0.5 K colder than the other radiometers (see Figure 9 and 10). In any case, the DMI-AWS IR120 performs reliable measurements of the thermal radiation of the skin snow and ice surface similar to the Infrared satellites and these observations are thus well suited for a validation of Infrared satellite IST products.

4.6.1 Satellite products

Four infrared satellite IST products were obtained for the inter-comparisons with the radiometer observations (see Table 4). The Metop_A AVHRR product is OSI-205A generated by the EUMETSAT OSI-SAF team (Dybkjær et al., 2016, Dybkjær & Eastwood, 2016) and has a spatial resolution of 1 km. The MODIS Aqua and Terra products are the MYD29 and MOD29, respectively (See e.g. Hall et al., 2004). The products are version 6 and have been obtained from obtained from NASA (Hall, D. K. and G. A. Riggs. 2015). The NPP SUOMI VIIRS IST product is the operational product from NOAA, known as the Environmental Data Record (EDRs), which were developed under the National Polar-Orbiting Environmental Satellite System (NPOESS) (see e.g. Tschudi et al., 2016; Key et al., 2013; Justice et al., 2013, Román, et al., 2014) obtained from the NOAA Comprehensive Large Array-Data Stewardship system (CLASS). The different satellite products are shown in Table 4.

Table 4: List of polar orbiting thermal infrared satellite datasets used for validation

Satellite product	Spatial resolution	File granule	Data Provider
Metop_A AVHRR OSI 205	1.1 km	3 min	EUMETSAT OSI-SAF
NPP SUOMI VIIRS	750 m	5 min	NOAA
MODIS Terra (MOD29.006)	1 km	5 min	NASA-GSFC
MODIS Aqua (MYD29.006)	1 km	5 min	NASA-GSFC

Examples of the satellite products are shown in Figure 20 where four cloud free passages have been recorded within 3 hours and 14 minutes on March 29, 2016. The similarities between the different products are obvious, but notice the MODIS Aqua satellite, that might display elevated geolocation uncertainties.

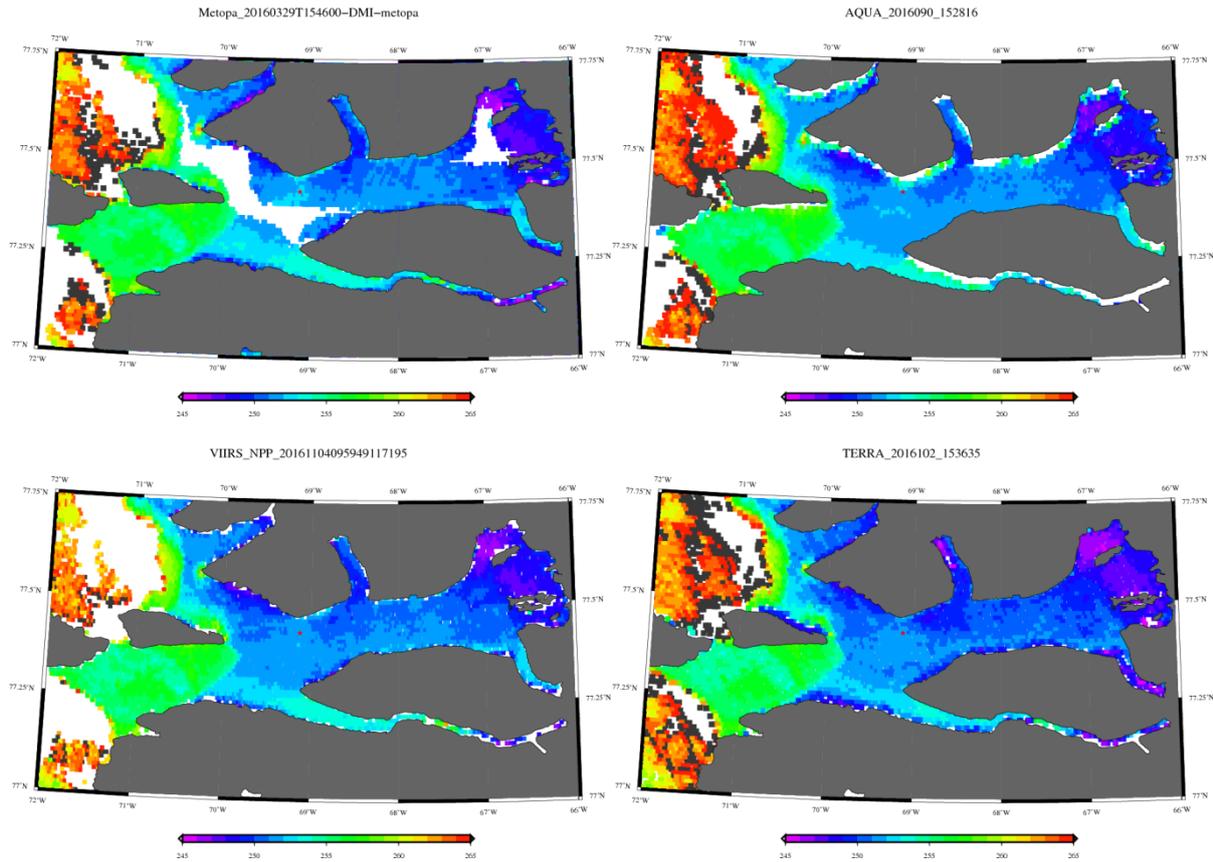


Figure 20: Examples of satellite IST products for the study area for Metop AVHRR (top left), MODIS Aqua (top right), Suomi NPP VIIRS (Lower left) and MODIS Terra (lower right) for March 29, 2016. The satellite observations are within 3 hours and 14 min, from 15.46 UTC (Metop AVHRR) to 19.00 UTC (MODIS Terra). The position of the IST FICE is shown with a red star.

4.6.1 Matchups

Match-ups of satellite IST observations with DMI-AWS IR120 measurements were performed for the time period 09-01-2016 - 26-05-2016, corresponding to the full observational period of the DMI-AWS IR120. The satellite IST values were selected from those pixels in each IST satellite overpass that corresponded to latitude and longitude coordinates within the range of the site coordinates ($-69.13685 / +77.42798$) ± 1.5 km. Only observations that passed all data quality control procedures provided by the satellite producers (cloud free, no land, etc.) were selected for the validation. A detailed description of the IST quality control screening for each of the satellite products is given in Appendix A.2

For each satellite product, 3 different IST estimates have been matched with the DMI-AWS IR120 data. The three different IST estimates are:

1. IST from the single pixel that covers the actual site
2. IST average over available satellite pixels within the site ± 1.5 km that pass the quality control
3. IST average only if all pixels within ± 1.5 km pass the quality control

A matchup data set was generated, where the different satellite estimates are paired with the DMI-AWS IR120 observations closest in time. The largest possible time difference between satellite and in situ observations is thus 5 minutes.

4.6.2 Validation Results

Figure 21 shows the resulting ice surface temperatures from the different satellite products against DMI-AWS IR120 between January and June 2016. The corresponding statistics of the differences data are shown in Table 4 and are visualized in Figure 22, including a scatterplot and a probability distribution for each of the 3 IST matchup estimates (closest pixel, spatial average and spatial average for cloud-free). In general, the best agreement is found for Metop_A AVHRR, with smallest mean and standard deviation of the differences to DMI-AWS IR120. Suomi NPP VIIRS and MODIS AQUA satellite observations differ the most from the DMI-AWS IR120 data.

The number of very cold satellite outliers in Figure 22 indicates that the cloud screening is challenging for all the products. This is evident in all satellite products but from the PDFs in Figure 22 right, it is clear that the cold tail for Suomi NPP VIIRS is more pronounced than for the other products. The requirement of all pixels within 1.5 km to be cloud free reduces the number of cold outliers for all products, except the MODIS AQUA as seen in the Figures 21 and 22.

When the cold outliers are disregarded, the scatterplots between satellite and DMI-AWS IR120 ice surface temperatures (see Figure 22, left) indicate that the different satellite IST products tend to be too warm for cold temperatures and too cold for warmer temperatures. As this effect is recognized for several of the satellite products it could be an artefact of the DMI-AWS IR 120 being too warm for cold temperatures and too cold for warm temperatures. However, the radiometer inter-comparison in Figure 8 shows an opposite behavior, with the DMI-AWS IR120 being colder for cold ISTs and warm for warmer ISTs.

The difference in the validation statistics is not large when the spatial averages are used instead of the single pixel matchups. This is the case for both bias and standard deviation, but the numbers of observations available for a match-up are significantly larger for the spatial averages (see Table 5). Yet, it does not lower the standard deviation and some outliers are still present.

The requirement of cloud free averages improves not only on the cold outliers. The improved performance of cloud free IST average is evident in both bias and standard deviation for all the products. However, this requirement reduces the number of match-up points substantially and all products except for the Metop_A AVHRR suffer from a low number of observations and hence little robustness, with NPP SUOMI VIIRS having only 26 matches left from an initial 197 for the single pixel matches.

Table 5: Pairwise statistics between DMI-AWS IR120 Ice surface temperatures and the 4 different satellite IST products for closest pixel (top), for averaging within 1.5 km (middle), and cloud-free only average (bottom).

Closest pixel	N(matches)	Mean difference (K)	Std. dev of differences (K)	RMSE (K)
Metop_A AVHRR	227	-0.4	1.9	2.0
NPP SUOMI VIIRS	197	-1.7	3.2	3.6
MODIS TERRA	122	-1.4	3.3	3.5
MODIS AQUA	165	-1.9	4.4	4.8

Spatial average	N(matches)	Mean difference (K)	Std. dev of differences (K)	RMSE (K)
Metop_A AVHRR	352	- 0.4	2.0	2.0
NPP SUOMI VIIRS	349	-2.0	3.8	4.3
MODIS TERRA	147	-1.4	3.1	3.4
MODIS AQUA	207	-1.9	3.9	4.3

Cloud-free average	N(matches)	Mean difference (K)	Std. dev of differences (K)	RMSE (K)
Metop_A AVHRR	173	-0.2	1.7	1.7
NPP SUOMI VIIRS	26	-0.9	2.7	2.8
MODIS TERRA	52	-0.6	1.3	1.4
MODIS AQUA	75	-1.7	3.1	3.5

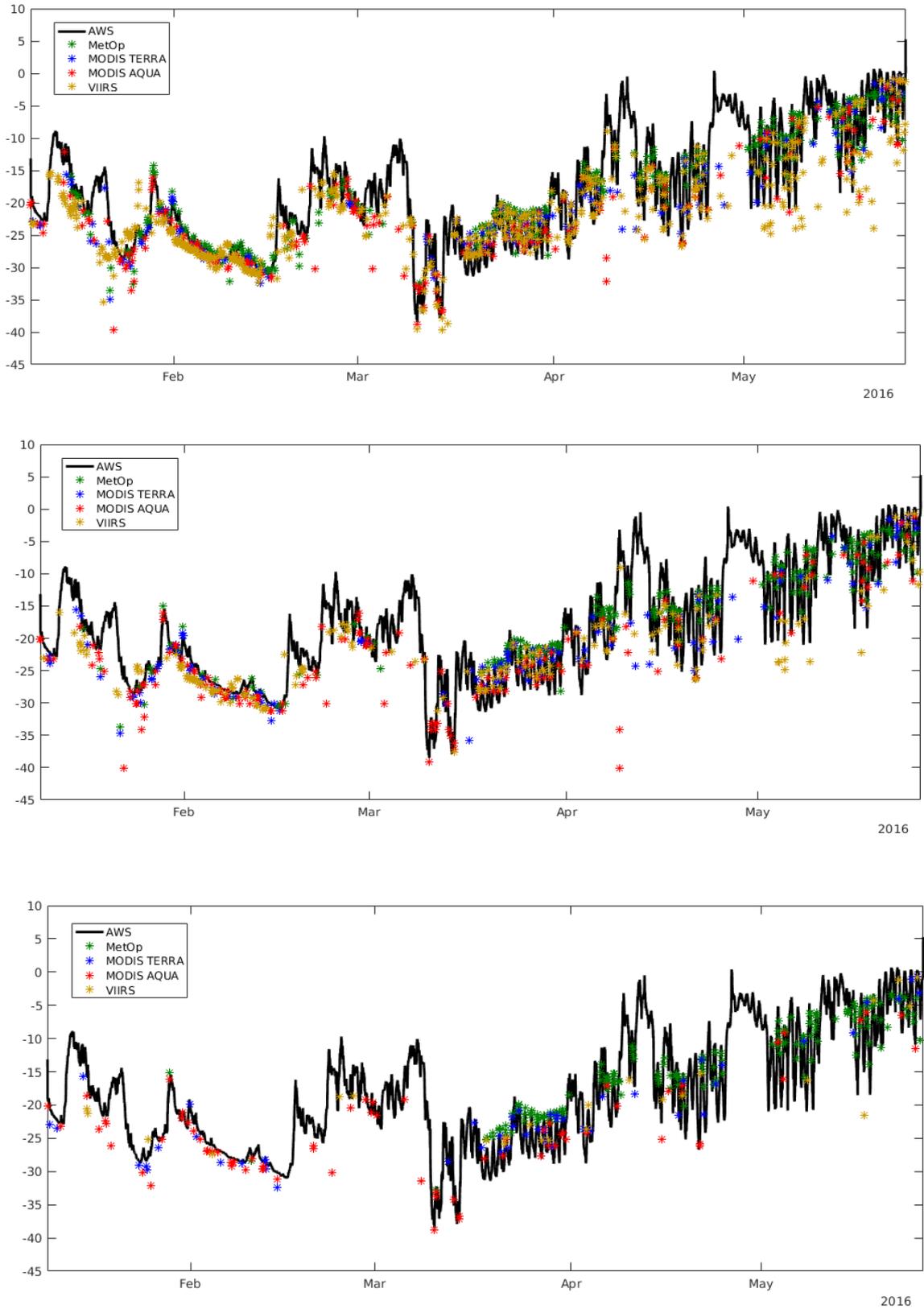


Figure 21: Satellite match-up (asterisk) with DMI-AWS IR120 (black line) for closest pixel (top), for averaging within 1.5 km (middle) and cloud-free only averaging within 1.5 km (bottom).

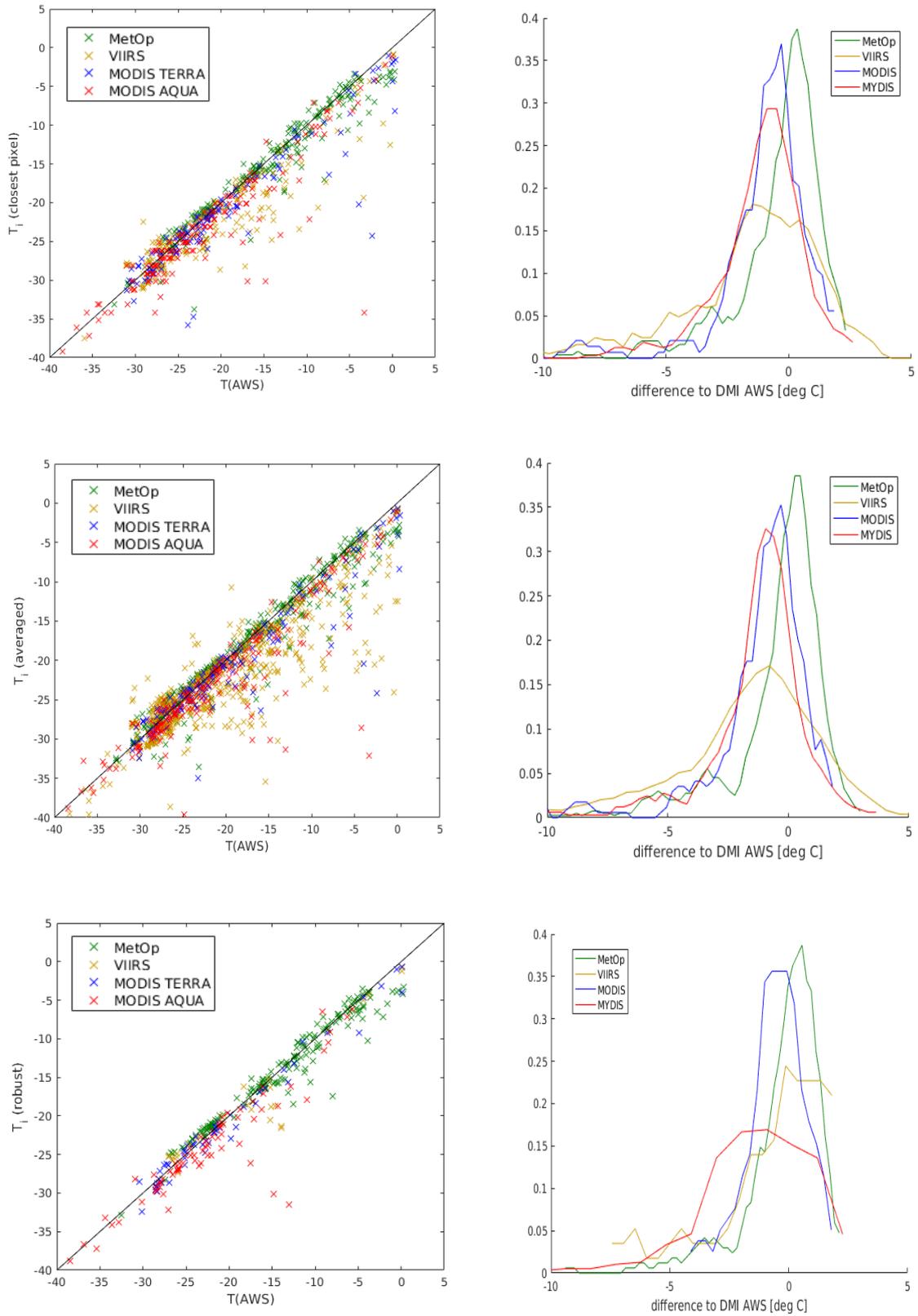


Figure 22: IST (DMI-AWS IR120) vs IST (satellite) (left) and probability distribution of IST difference between satellite and in-situ data (right) for closest pixel (top), for averaged available pixel (middle) and for cloud-free average (bottom).

5 SUMMARY AND CONCLUSION

The primary objective of the experiment was to make an inter-comparison of the thermal infrared radiometer instruments which can be used as fiducial reference measurements for validating satellite ice surface temperature observations, such as Sentinel 3. All six radiometers that participated in the campaign, were recording during the inter-comparison experiment and the campaign can therefore be regarded a success. In addition, other experiments were conducted to assess different components in the uncertainty budget, such as spatial variability and angular dependency when measuring ice surface temperature.

The six instruments that participated in the inter-comparison experiment were: two Infrared Sea Surface Temperature Autonomous Radiometers (ISARs), one Heitronics KT 15.85IP and three Campbell Scientific IR120. As the different radiometers exhibit different sampling rates an interpolation procedure was applied in order to provide time series with a common time sampling that allows a quantitative inter-comparison.

Due to technical problems in the extreme conditions, the different instruments cover only shorter time tracks, hindering a continuous and simultaneous measurement. While the NOCS ISAR instrument had difficulties operating at temperatures below -22 °C, all instruments agree better for warmer temperatures. The mean differences between instrument-pairs ranged between 0.13 K and 1.0 K with standard deviations from 0.05 K to 1.20 K. The mean difference between the two ISARs was 0.93 K, which was higher than anticipated. This may be attributed to the fact that the DMI ISAR08 did not perform according to specifications, as revealed by the post campaign calibration.

The spatial variability experiment indicated that the variation in the IST within a satellite footprint of 1 km² could be around 0.12 - 0.25°C. Experiments with different surface observing angles showed a significant sensitivity of the measured Brightness temperatures to the observing angle. An incidence angle of 55° from nadir was in general 0.25-0.5 K colder than an incidence angle of 25°. This range of angular variation does not explain all the variability in the sledge pulling spatial variability experiment. A significant component of the spatial variability is from actual surface temperature variability. This is also indicated in the differences between the DMI-AWS IR120 and all other instruments in the inter-comparison experiment.

The Freeze-up experiment was not optimal due to a small basin, which required post-processing to extract the actual temperature of the basin. However, it demonstrated the potential for comparing radiometers over a larger range of temperatures from the freezing point of seawater and colder. Within 24 hours, the pond temperature fell by approx. 4 K and the pond's ice thickness grew by 4-5 cm.

A match-up of the DMI-AWS IR120 with satellite data from four different data-products (NPP SUOMI VIIRS EDR, MODIS Terra, MODIS Aqua and Metop_A AVHRR) has further been applied. While the bias is considerable due to poor data quality, best agreement is found with Metop_A AVHRR.

6 LESSONS LEARNED AND FUTURE WORK

Since this is the first time, such an extensive experiment has been conducted, it is important to discuss what could be improved for the next IST FICE campaign. Suggestions for improvements or additional experiments for the next IST FICE are listed below:

The two ISARs were calibrated at NPL for cold target conditions in the laboratory and with room temperatures of about +25°C for the reference targets. The NPL laboratory experiments showed larger errors than expected for the ISARs for cold targets. For the next field experiment, a blackbody reference, such as the second generation Concerted Action for the Study of the Ocean Thermal Skin (CASOTS-II, Donlon et al., 2014) could be transported to the high latitudes and calibration experiments could be performed in both warm and cold conditions, with a cold blackbody target. This would indicate the performance of the radiometers in a real cold environment and also assess the errors introduced when target and reference body temperatures are very far apart. Pre or post campaign calibration experiments at NPL could then link the performance to SI traceable observations.

The ISAR uncertainty estimates were derived in Wimmer and Robinson (2016) for SST conditions. More work should be done to derive and assess realistic radiometer uncertainty values for cold conditions over sea ice, and we recommend that this forms part of a follow-on project. The Ice uncertainty modelling was not part of this option and the uncertainty estimates therefore have to be regarded as preliminary values until a proper assessment has been carried out.

The spatial variability experiments, where a sledge was dragged with mounted radiometers could be improved. The uneven surface meant that the sledge tilted and that the view incidence angles of the radiometers varied during the experiment. Changing incidence angles resulted in changes in the emissivity of the snow and ice surface. The angular dependence experiment showed that this variability could amount to 0.5 K. It can therefore not be ruled out that some of the observed IST variation in this experiment is due to the changes in incidence angles, rather than changes in snow/ice properties or the temperature. Next time, a spatial variability experiment could be conducted using a thermal infrared camera mounted on a drone. Flying the drone at a few hundred meters height around the measurement site, will give a unique spatial distribution of the relative variations in IST within a satellite footprint. Simultaneous FRM TIR observations on the sea ice could then be used to reference the relative observations.

The freeze-up experiment was not optimal for this experiment. A large basin filled with water overnight without instruments mounted. Due to time and logistic constraints, a smaller basin was dug out at the end of the campaign. The smaller size of the basin resulted in contamination of the IR observations from the surrounding cold ice. Next time, a larger basin should be dug out and filled with water immediately, while at least two radiometers were recording.

A follow up project could also address in situ measurements of the snow and ice emissivity and the variability within the field site. As evident from the report, the snow surface at the site was very inhomogeneous and the cold sky temperatures means that even small emissivity errors can have an effect on the measured IST. An assessment of the emissivity changes and the related effects on the observations would be very valuable.

The estimation of uncertainty on the IST radiometer observations is an area where more work could be done, as described above. The inter-comparison experiment showed that uncertainties derived for SST conditions are too low, compared to the actual IST differences. Improved instrument calibration experiments in cold environments could aid in deriving the uncertainty budgets for the FRM TIR observations in cold environments. Note, that the cold environment is challenging for the electronics and mechanics and result in an elevated risk for radiometers breaking or not performing to its specifications in these conditions. Although experience from previous field campaigns and knowledge about the environmental conditions and the instruments are crucial factors to minimize risks, occasional breakdown and instrument failures have to be accepted as part of conducting high latitude field campaigns.

7 ACKNOWLEDGEMENTS

The study is part of the ESA Fiducial reference measurements for satellite derived surface temperature measurements (FRM4STS) project. DMI provided logistic and technical support during the field campaign and the supporting instrumentation and observations in Qaanaaq were part of the EU FP 7 project: Ice, Climate, Economics – Arctic Research on Change (ICE-ARC). The satellite IST observations were provided by EUMETSAT OSI-SAF, NOAA and NASA.

8 REFERENCES

Barker-Snook, I., Theocharous E. and Fox, N. P. (2016) 2016 comparison of IR brightness temperature measurements in support of satellite validation. Part 2: Laboratory comparison of radiation thermometers. *NPL Report env 14*.

Cambell Scientific (2015), IR100/IR120 Infra-red Remote Temperature Sensor user manual, version 25.2.15. (https://s.campbellsci.com/documents/eu/manuals/ir100_ir120.pdf)

Donlon, C., Robinson, I. S., Wimmer, W., Fisher, G., Reynolds, M., Edwards, R., & Nightingale, T. J. (2008) An infrared sea surface temperature autonomous radiometer (ISAR) for deployment aboard volunteer observing ships (VOS). *Journal of Atmospheric and Oceanic Technology*, 25(1), 93-113.

Donlon, C. J., Wimmer, W., Robinson, I., Fisher, G., Ferlet, M., Nightingale, T., & Bras, B. (2014) A Second-Generation Blackbody System for the Calibration and Verification of Seagoing Infrared Radiometers. *Journal of Atmospheric and Oceanic Technology*, 31(5), 1104-1127.

Dybkjær, G., R. Tonboe, and J. Høyer (2012) Arctic surface temperatures from Metop AVHRR compared to in situ ocean and land data, *Ocean Sci. Discuss.*, 9, 1009–1043, 2012 www.ocean-sci-discuss.net/9/1009/2012/ doi:10.5194/osd-9-1009-2012

Dybkjær, G. Eastwood, S., Borg, A. L., Høyer, J. & Tonboe, R. (2016) Algorithm theoretical basis document for the OSI SAF Sea and Sea Ice Surface Temperature L2 processing chain, OSI205b Version 1.2. EUMETSAT Ocean and Sea Ice SAF High Latitude Processing Centre, Doc-id: SAF/OSI/CDOP2/DMI/SCI/MA/223.

Dybkjær, G. and Eastwood, S (2016) Validation Report for the OSI SAF High Latitude L2 Sea and Sea Ice Surface Temperature, OSI-205 Version 1.1. Doc-id: SAF/OSI/CDOP2/DMI/SCI/RP/247

Hall, D.K., J. Key, K.A. Casey, G.A. Riggs and D. Cavalieri (2004) Sea ice surface temperature product from the Moderate Resolution Imaging Spectroradiometer (MODIS), *IEEE Transactions on Geoscience and Remote Sensing*, 42:1076-1087.

Hall, D. K. and G. A. Riggs. (2015) MODIS/Aqua Sea Ice Extent 5-Min L2 Swath 1km, Version 6, Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: <http://dx.doi.org/10.5067/MODIS/MYD29.006>.

Høyer, J. L., Karagali, I., Dybkjær, G., & Tonboe, R. (2012) Multi sensor validation and error characteristics of Arctic satellite sea surface temperature observations. *Remote Sensing of Environment*, 121, 335-346.

Høyer, J. L., Le Borgne, P., & Eastwood, S. (2014) A bias correction method for Arctic satellite sea surface temperature observations. *Remote Sensing of Environment*, 146, 201-213.

Justice, C. O., Román, M. O., Csiszar, I., Vermote, E. F., Wolfe, R. E., Hook, S. J. & Tschudi, M. (2013) Land and cryosphere products from Suomi NPP VIIRS: Overview and status. *Journal of Geophysical Research: Atmospheres*, 118(17), 9753-9765.

Key, J. R., Mahoney, R., Liu, Y., Romanov, P., Tschudi, M., Appel, I. & Meade, P. (2013) Snow and ice products from Suomi NPP VIIRS. *Journal of Geophysical Research: Atmospheres*, 118(23).

Rigor, Ignatius G., Roger L. Colony, and Seelye Martin (2000) Variations in surface air temperature observations in the Arctic, 1979-97. *Journal of Climate* 13.5: 896-914.

Román, M. O., Justice, C., & Csiszar, I. (2014) Land, Cryosphere, and nighttime environmental products from Suomi NPP VIIRS: Overview and status. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 3530-3533). IEEE.

Sobrino, JoséA, and Vicente Caselles (1993) A field method for measuring the thermal infrared emissivity. *ISPRS Journal of Photogrammetry and Remote Sensing* 48.3, 24-31.

Tonboe, R. T., G. Dybkjær, and J. L. Høyer (2011) Simulations of the snow covered sea ice surface temperature and microwave effective temperature. *Tellus* 63A, 1028-1037.

Tschudi, M., Riggs, G. Hall, D. & Román, M. O. (2016) Suomi-NPP VIIRS Ice Surface Temperature Algorithm Theoretical Basis Document (ATBD), version 1.0

Wimmer, W. and Donlon, C. (2012) ISAR User Manual v2.03. National Oceanography Centre Southampton, UK, 44pp.

Wimmer, W., & Robinson, I. S. (2016) The ISAR Instrument Uncertainty Model. *Journal of Atmospheric and Oceanic Technology*, 33(11), 2415-2433.

APPENDIX A: DETAILS ON DATA PROCESSING

DATA PROCESSING FOR ANGULAR DEPENDENCE EXPERIMENT

Data record of DMI ISAR temperatures for angle 45 are missing; those for angle 135 are partly corrupt, likely owing to a software error as the first two measurements of each measurement-cycle after the angle-adjustment are affected by missing values from the previous angle 45 measurements. Mean and standard deviation for all angle 45 measurements have thus been calculated manually.

QUALITY CONTROL FOR SATELLITE MATCH-UP

The different satellite data-sets have different quality assessments, which are either directly incorporated in the IST data, saved as separate flags (Metop_A, Suomi NPP VIIRS) or both (MODIS). Table 2 shows the quality constraints undertaken for the respective data-set. If the criteria are not fulfilled, respective IST values have been set to NaN.

Data-set	Variable	Criteria
Metop_A AVHRR	quality_level [0:5]	ql > 3
Suomi NPP VIIRS	QF1_VIIRSISTEDR QF2_VIIRSISTEDR QF3_VIIRSISTEDR	IST Quality = high cloud confidence indicator CCI = conf. clear neighboring CCI = confidently clear
MODIS Terra	Ice_Surface_Temperature IST_Pixel_QA	5000 < IST < 65535 IST_Pixel_QA = 0
MODIS Aqua	Ice_Surface_Temperature IST_Pixel_QA	5000 < IST < 65535 IST_Pixel_QA = 0

Table A.1: Quality control of ice surface temperature data in the respective data-sets.

Note:

- The quality flags QF_VIIRSISTEDR use a bit-system (GSFC, 2014). Chosen criteria are the strongest in each category, i.e. both bits are set to 0 (Bit 1=0 and Bit 0=0)
- The neighboring CCI quality flag in Suomi NPP VIIRS are found crucial for eliminating outliers in the IST match-up.
- The IST in MODIS (both AQUA and TERRA) has been scaled with the factor 0.01 (Riggs & Hall, 2015).
- The IST in Suomi NPP VIIRS has been scaled with two factors (from the .hdf-file):

$$\text{IST} = \text{IceSurfaceTemperature} * \text{ISTFactors}(1) + \text{ISTFactors}(2)$$

- The resolution of `dlon` and `dlat` in MODIS (both AQUA and TERRA) is with 406x271 coarser than that of `IST` and `IST_QA` (2030x1354). That is, the first element (1,1) in the geolocation SDS corresponds to element (5,5) in `Ice_Surface_Temperature`. A 2D linear interpolation of `dlon` and `dlat` to the finer grid (2030x1354) was done with the MATLAB command `interp2`.