Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements

Pia Nielsen-Englyst[1,2], Jacob L. Høyer[2], Kristine S. Madsen[2], Rasmus T. Tonboe[2], Gorm Dybkjær[2] and Sotirios Skarpalezos[2]

5   [1] Technical University of Denmark (DTU), DK-2800 Kongens Lyngby, Denmark
    [2] Danish Meteorological Institute (DMI), DK-2100 Copenhagen Ø, Denmark

Correspondence to: Pia Nielsen-Englyst (pne@dmi.dk)

Abstract.

The Arctic region is responding heavily to climate change, and yet, the air temperature of ice covered areas in the Arctic is
10 heavily under-sampled when it comes to in situ measurements, resulting in large uncertainties in existing weather and reanalysis products. This paper presents a method for estimating daily mean clear sky 2 metre air temperatures (T2m) in the Arctic from satellite observations of skin temperature, using the Arctic and Antarctic ice Surface Temperatures from thermal Infrared (AASTI) satellite dataset, providing spatially detailed observations of the Arctic. The method is based on a linear regression model, which has been tuned against in situ observations to estimate daily mean T2m based on clear sky satellite
15 ice surface skin temperatures. The daily satellite derived T2m product includes estimated uncertainties and covers the Arctic sea ice and the Greenland Ice Sheet during clear skies for the period 2000-2009, provided on a 0.25 degree regular latitude-longitude grid. Comparisons with independent in situ measured T2m show average biases of 0.30°C and 0.35°C and average root mean square errors of 3.47°C and 3.20°C for land ice and sea ice, respectively. The associated uncertainties are verified to be very realistic for both land ice and sea ice, using in situ observations. The reconstruction provides a much better spatial
20 coverage than the sparse in situ observations of T2m in the Arctic and is independent of numerical weather prediction model input. Therefore, it provides an important supplement to simulated air temperatures to be used for assimilation or global surface temperature reconstructions. A comparison of T2m derived from satellite and ERA-Interim/ERA5 estimates shows that the satellite derived T2m validates similar or better than ERA-Interim/ERA5 against in situ measurements in the Arctic.

1 Introduction

25 The Arctic climate is changing rapidly with surface temperatures rising faster than other regions of the world due to Arctic amplification (Graversen et al., 2008; IPCC, 2013; Pithan and Mauritsen, 2014; Richter-Menge et al., 2017), with the maximum warming occurring during late fall and early winter (Box et al., 2019; Screen and Simmonds, 2010). Meteorological measurements in Greenland show a general warming since the 1780s (Cappelen (ed), 2021; Masson-Delmotte et al., 2012; Hanna et al., 2021; Abermann et al., 2017), with the 2000s being the warmest decade in western and southern Greenland, while
30 the 2010s in parts of eastern Greenland were slightly warmer than the 2000s (Cappelen (ed), 2021).

The Arctic surface air temperature is one of the key climate indicators used to assess regional and global climate changes (Hansen et al., 2010; Pielke et al., 2007) and both model simulations and observations indicate that warming in the global climate is amplified at the northern high latitudes (e.g. Collins et al., 2013; Holland and Bitz, 2003; Overland et al., 2018). Traditionally, near surface air temperatures have been measured at the height of 1-2 m using automatic weather stations (AWSs) or buoys (Hansen et al., 2010; Jones et al., 2012; Rayner, 2003; World Meteorological Organization, 2014). Extreme temperatures, winds and the remoteness of the Arctic make in situ observations in the Arctic temporally and spatially sparse (Reeves Eyre and Zeng, 2017). Therefore, it is challenging to achieve climate-quality temperature records for this region.

The key data sets used to assess the Arctic temperature changes are global gridded near surface air temperature datasets that are derived using in situ observations (Hansen et al., 2010; IPCC, 2013; Morice et al., 2012; Smith et al., 2008; Vose et al., 2012). These datasets typically have higher uncertainties in the Arctic region due to the limited availability of in situ observations (Cowtan and Way, 2014; Lenssen et al., 2019; Rapaić et al., 2015). In addition, global reanalysis products such as ERA-Interim (ERA-I) and ERA5 (Dee et al., 2011; Hersbach et al., 2020) are frequently used to study the changes in the Arctic and to force ocean and sea ice models. Despite the assimilation of in situ data in the global reanalysis models, significant model differences have been reported for the Arctic (Davy and Outten, 2020; Delhasse et al., 2020; Lindsay et al., 2014; Wesslén et al., 2014) as well as large deviations from observations of T2m over Arctic sea ice (Wang et al., 2019).

Observations from polar orbiting satellites offer a very good supplement to the in situ observations through high spatial and temporal coverage of the high latitudes and may improve the surface temperature products and the assessment of the Arctic climate changes. Therefore, daily near surface air temperatures derived from satellite temperature observations have the potential to increase the amount of information in the data sets and improve the quality of the climate records, as recognized in Merchant et al. (2013) and (Rayner et al., 2020).

Two fundamental challenges exist when deriving a T2m product from infrared satellite observations. The first challenge is that infrared sensors (in the atmospheric window region of 10-12 micrometre wavelength) measure the ice surface skin temperature (IST_{skin}), whereas the current global temperature products include the near surface air temperature as measured continuously by AWSs and buoys. The surface skin temperature may differ considerably from the near surface air temperature measured by AWSs or buoys. Previous studies have compared satellite retrieved IST_{skin} and T2m from AWSs located on the Greenland Ice Sheet (GrIS; Dybkjær et al., 2012a; Hall et al., 2008, 2012; Koenig and Hall, 2010; Shuman et al., 2014) and over the Arctic sea ice (Dybkjær et al., 2012) and found temperature differences of which a significant part could be attributed to the temperature difference between T2m and IST_{skin}. Other studies have investigated the relationship between T2m and IST_{skin} over ice using in situ observations (Adolph et al., 2018; Hall et al., 2008, 2004; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma et al., 2008). Nielsen-Englyst et al. (2019) found that on average T2m is 0.65-2.65°C higher than IST_{skin} with variations depending on the location of the measurement i.e. over sea ice, seasonal snow cover, and the following zones of the GrIS: lower ablation zone, upper-middle ablation zone, and accumulation zone. The T2m-IST_{skin} difference was found to vary seasonally with largest differences during the winter (when inversions are most common), and during melting conditions in

the summer (where the surface temperature is fixed at the melting point). In Nielsen-Englyst et al. (2019), wind speed and cloud cover were identified as key parameters determining the T2m- IST_{skin} difference.

The second challenge, related to the use of satellite-derived infrared IST_{skin} to derive T2m, is that the availability of IST_{skin} observations is limited to clear sky conditions while T2m is measured continuously by AWSs and buoys. Previous studies

5   have shown that a satellite-derived, clear-sky, surface temperature record can be significantly colder than an all-sky surface temperature record (Koenig and Hall, 2010; Nielsen-Englyst et al., 2019). To benefit from the good coverage of satellite surface temperature data, above mentioned challenges should be considered with caution. This work, starting with Nielsen-Englyst et al. (2019), has been initiated to estimate clear sky T2m from satellite observations (whenever these are available) for the Arctic sea ice and the GrIS in order to provide spatially-detailed observations for the areas unobserved by in situ stations and to

10   supplement the in situ observations already available. Here, special attention have been given to above mentioned challenges and the relationships between the near surface air temperature and the satellite skin measurements have been explored in detail. A regression-based approach has been used to estimate daily T2m using satellite IST_{skin} and a seasonal cycle function as predictors based on the work presented in Høyer et al. (2018). The derived product covers only days with none or limited clouds, when satellite skin temperature observations are available. However, for those days when the satellite derived T2m

15   product is available, it provides an estimate of the daily averaged all-sky T2m since it has been regressed towards in situ measurements from both clear and cloudy conditions. In order to further facilitate the usage of the derived product in modelling and for monitoring purposes, each satellite retrieved T2m estimate comes with uncertainties.

Similar efforts have been done to estimate clear sky near surface air temperatures (and corresponding uncertainties) over land, ocean and lakes using satellite observations to cover all surfaces of the Earth (Good, 2015; Good et al., 2017; Høyer et al.,

20   2018). The previous work has mostly been done as a part of the European Union's Horizon2020 project EUSTACE (EU Surface Temperatures for All Corners of Earth, 2015-2019, https://www.eustaceproject.org) with the overall aim to produce a globally complete gap-free daily near surface temperature analysis since 1850. It is outside the scope of this paper to produce a continuous gap-free daily near surface temperature analysis. However, within EUSTACE this has been done using a statistical model to combine satellite derived clear sky near surface air temperatures and in situ observations, and their respective

25   uncertainty estimates (Morice et al., 2019; Rayner et al., 2020). The clear sky T2m product derived in this paper has been used to generate this daily gap-free EUSTACE T2m product for the GrIS and the Arctic sea ice, while similar clear sky temperature products have been used over land, ocean, and lakes, respectively.

This paper is structured such that Sect. 2 describes the in situ data and the satellite data. Section 3 presents the method used to estimate clear sky daily T2m and uncertainties. The resulting T2m dataset and its validation are presented in Sect. 4 and

30   discussed in Sect. 5. Conclusions are given in Sect. 6.

2 Data

2.1 In Situ data

In situ observations of near surface air temperatures have been collected from weather stations, expeditions and campaigns covering ice and snow surfaces to assemble the DMI-EUSTACE database. The database includes quality controlled and uniformly formatted temperature observations covering ice and snow surfaces during the period 2000-2009 (Høyer et al., 2018). For the GrIS we use the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) data provided by the Geological Survey of Denmark and Greenland (GEUS; Ahlstrøm et al., 2008; van As et al., 2011) and the Greenland Climate Network data (GC-Net; Kindig, 2010; Shuman et al., 2001; Steffen and Box, 2001). Only PROMICE data from the middle-upper ablation zone and accumulation zone have been used to ensure that data are only acquired over permanently snow or ice covered surfaces. Observations covering seasonal snow have also been used from the Atmospheric Radiation Measurement (ARM) Program from the two sites: Atqasuk (ATQ) and Barrow (BAR), at the North Slope of Alaska (Ackerman and Stokes, 2003; Stamnes et al., 1999). Data from Arctic sea ice are primarily retrieved from the meteorological observation archive at the European Centre for Medium-Range Weather Forecasts (ECMWF) MARS data storage facility, providing 196 unique data series from drifting buoys. These sea ice data are supplemented with data from 10 U.S. Army Cold Regions Research Engineering Laboratory (CRREL) mass balance buoys (Perovich et al., 2016; Richter-Menge et al., 2006) and observations from a weather station located 29 m above the sea surface on the research vessel, *Polarstern*, operated by the Alfred-Wegener-Institute in the sea ice covered parts of the Arctic Ocean (Knust, 2017; König-Langlo et al., 2006). We also use air temperature measurements obtained from ice buoys deployed in the Fram Strait region within the framework of the Fram Strait Cyclones (FRAMZY) campaigns during the years 2002, 2007, and 2008 as well as air temperatures from the Arctic Climate System Study (ACSYS) campaign in 2003 (Brümmer et al., 2011b, c, 2012b, a). Finally, we use data from two ice buoy campaigns operated by the Meteorological Institute of the University of Hamburg within the framework of the integrated EU research project DAMOCLES (Developing Arctic Modelling and Observing Capabilities for Long-term Environmental Studies; Brümmer et al., 2011a).

The different in situ types measure the air temperature at different heights that furthermore differ over time depending on the amount of snow fall, snow drift and snow melt. Here, we will refer to T2m for all observation types regardless of these variations. Nielsen-Englyst et al. (2019) showed small changes (<0.22°C) in T2m-IST$_{skin}$ differences when using only observations within the measurement range of 1.90-2.10 m in height compared to using all measurements (ranging in measurement height from 0.3 m to 3 m). The observations from *Polarstern* at 29 metres height are not included in the derivation of the near surface air temperature data set, but only used for the validation. The accuracy of the air temperature sensors for all observation sites is approximated to 0.1°C (Hall et al., 2008; Høyer et al., 2017b). Few data sources provide both skin and air temperatures e.g. the PROMICE and ARM stations. The PROMICE skin temperatures have been calculated from up-welling longwave radiation, measured by Kipp & Zonen CNR1 or CNR4 radiometer, assuming a surface longwave emissivity of 0.97 (van As, 2011). All in situ data have been screened for spikes and other unrealistic data artefacts by visual inspection.

Afterwards, the in situ observations have been averaged to daily temperatures using all available observations. Figure 1 shows the number of daily averaged in situ observations each year (2000-2009) of IST_{skin} and T2m over Arctic land ice and sea ice. The two ARM stations are included as land ice stations in this analysis, and only data from snow covered periods are used. In total 65,810 observations with daily T2m and 7,057 observations with daily IST_{skin} are available over land ice. See Table 1 for more information on the in situ observations used in this study.

2.2 Satellite data

The satellite data used in this study is from the Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors (AASTI; Dybkjaer et al., 2018; Dybkjær et al., 2014; Høyer et al., 2019) data set, covering high latitude seas, sea ice, and ice sheet with clear sky surface temperatures based on satellite infrared measurements from the CLARA-A1 data set compiled by EUMETSAT's Climate Monitoring, Satellite Application Facility (CM-SAF; Karlsson et al., 2013). The data set is based on one of the longest existing satellite records from the Advanced Very High Resolution Radiometer (AVHRR) instruments on board a long series of NOAA satellites. AASTI contains swath based (i.e. Level 2; L2) ice surface skin temperature (IST_{skin_L2}) data processed and error corrected on the original Global Area Coverage (GAC) grid. The first version of the AASTI product, which is used in this study, is available from 2000 to 2009 in the original projection and resolution (L2), i.e. ~0.05 arc degree resolution and multiple daily coverage. Since 2000, seven different AVHRR instruments have been orbiting the globe, each 14 times per day, and thus providing approximately bi-hourly coverage of the Polar Regions (Figure 2). The number of operational satellites has increased from 2 to 6 from 2000 to 2009. The IST algorithm used to generate the AASTI data set is based on thermal infrared brightness temperatures of AVHRR channel 4 (centre wavelength at ~11 microns) and 5 (centre wavelength at ~12 microns), and the satellite zenith angle. The algorithm is a split window algorithm, working within three temperature domains for each individual satellite (Key et al., 1997). The retrieval calibration of each domain has been done by relating modelled surface temperatures with modelled top-of-atmosphere brightness temperatures, determined by a radiative transfer model (Dybkjær et al., 2014). Cloud masking has been performed using the Polar Platform System (PPS) processing cloud processing software (Dybbroe et al., 2005a, b).

As discussed in Merchant et al. (2017), satellite-based climate data records should include uncertainty estimates. The AASTI IST_{skin_L2} data come with uncertainties divided into three independent uncertainty components, each with different characteristics: the random uncertainty (μ_{rnd_L2}), a locally systematic uncertainty (μ_{local_L2}) and a large-scale systematic ("global") uncertainty (μ_{glob_L2}). These three components have been chosen since they behave differently when aggregating the observations in time or space (see Sect. 3.2). This uncertainty methodology has been developed within the sea surface temperature (SST) community (Bulgin et al., 2016; Rayner et al., 2015) and will be followed here. The total uncertainty on the IST_{skin_L2}, μ_{total_L2}, is calculated by summing each component in quadrature (i.e., square root of sum of squares). Excluding the cloud mask uncertainty, grid-cell systematic uncertainties (μ_{glob_L2}) are set to a fixed value of 0.1°C to represent systematic uncertainties in the forward models (see e.g. Merchant et al., 1999; Merchant and Le Borgne, 2004). The AASTI IST_{skin_L2}

data also come with a quality level (QL) from 1 (bad data) to 5 (best quality), with the addition of level 0 (no data) (GHRSST Science Team, 2010).

Here, we have aggregated the AASTI IST_{skin_L2} observations into 3 hourly and daily, gridded Level 3 (L3) averages of IST_{skin_L2} on a fixed 0.25 by 0.25 degrees regular geographical grid. This grid was chosen within the EUSTACE project to ensure a common grid to be used globally. The daily gridded averages (IST_{skin_L3}) are calculated by averaging all available IST_{skin_L2} observations with a quality flag of 4 (good) or 5 (best) for a given date and within the 0.25 degree bin. This has been done to facilitate the development of the relationship model and to ease the user uptake. The data in the daily aggregated files contain mean surface temperature observations from 00:00 to 24:00 hours local solar time, 3-hourly bin averages of surface temperatures and also the number of observations in the eight time bins during each day. The 3-hourly numbers of observations are used to estimate the satellite sampling throughout the day, and the 3-hourly temperature data to gain confidence in the daily cycle estimates (see quality checks below). Figure 3 shows the mean number of observations per day in each of the eight time intervals given in local time for the Arctic region. The variation in coverage throughout the day is a combined effect of the satellite overpassing, performance of the cloud screening algorithm, and the cloud free conditions during the day. In addition, the fixed 0.25 degrees regular geographical grid results in a decreasing L3 bin area when approaching the North Pole. The maximum satellite coverage is generally seen around 80°N with a minimum at the North Pole. Cloud free conditions over the GrIS are primarily observed around noon and the early afternoon.

In order to best resolve the diurnal cycle with satellite information, we require data both during the night (between 18:00 and 6:00 local solar time) and the day (between 6:00 and 18:00 local solar time) in order to calculate IST_{skin_L3}. To identify sea ice, we use an ice mask for which sea ice is characterised by sea ice concentrations above 30 lar according to the EUMETSAT OSISAF Global Sea Ice Concentration Climate Data Record (Tonboe et al., 2016). A few more checks have been set up in order to minimize the temporal sampling errors, the effects of undetected clouds and outliers as well as inconsistencies between the ice mask and the surface temperatures. Following Høyer et al. (2018), the IST_{skin_L3} is discarded if one of the following criteria is met:

- IST_{skin_L3} exceeds +5°C, indicating inconsistency between the ice mask and the surface temperatures.
- The standard deviation of satellite IST_{skin_L2} during one day exceeds 7.07°C, corresponding to a sinusoidal daily cycle with a difference between day and night of 20°C.
- The difference between IST_{skin_L3} and the average of all available 3 h bin averages exceeds 10°C.
- IST_{skin_L3} is more than 10°C colder than the corresponding average of up to 24 neighbouring cloud free observations (in a 5 by 5 grid cell square) with the same surface type.

The criteria above have been derived from analysis and inspection of the satellite data and with considerations to the results presented in Nielsen-Englyst et al. (2019). Inconsistencies between the ice mask and surface temperature typically occur along the coasts and sea ice edge, where the OSISAF product is subject to land-spillover effects causing spurious ice in ice-free areas (Lavergne et al., 2019). Using a surface temperature threshold of 5°C reduces the land-spillover effects and results in increased consistency between the ice mask and the surface temperatures.

The satellite-derived surface temperature has seasonal differences in daily variability, with largest standard deviations during the summer in Greenland and during the winter for sea ice, when the freeze-up of sea ice causes higher variability along the sea ice margin (Fig. 4). The main uncertainty components of the IST_{skin_L3} estimates are erroneous cloud screening and the spatial variance of snow and ice surface emissivity, which are not accounted for in the retrieval algorithm. The presence of non-detected clouds will contribute to increased standard deviations and usually a cold IST_{skin_L3} bias, since the cloud tops and other atmospheric constituents generally are colder than the surface (Dybkjær et al., 2012).

2.2.1 Validation

Additional satellite versus in situ differences arise when comparing satellite observations with pointwise ground measurements due to different spatial and temporal characteristics. To assess the magnitude of these effects, the IST_{skin_L3} data have been validated against in situ observations from the PROMICE and ARM stations. Table 2 shows the validation results of daily IST_{skin_L3} against in situ skin temperatures (IST_{skin_insitu}) and in situ 2 metre air temperatures ($T2m_{insitu}$), respectively. The maximum matchup distance is 14.6 km and the average distance is 8.1 km, considering the AWSs in Table 2. The topography mask included in the HIRHAM5 regional climate model (see e.g. Langen et al. 2015) has been used to calculate the differences in elevation (Δh) between the in situ stations and corresponding satellite pixels. There is no clear correlation between the large biases and large elevation differences from this table, but the elevation effects are contributing to the spatial sampling error. The spatial and temporal sampling errors contribute to the overall uncertainty, but effects from erroneous cloud screening, algorithm simplifications, and uncertainties in the in situ observations are also included in the results. Previous studies find that erroneous cloud screening (undetected clouds) is one of the main reasons for the cold biases observed when comparing satellite observed IST with in situ measurements (Hall et al., 2004, 2012; Koenig and Hall, 2010; Østby et al., 2014; Westermann et al., 2012). Another important contribution is the effect of comparing clear sky satellite observations with all-sky in situ observations, as discussed in Nielsen-Englyst et al. (2019). In general, IST_{skin_L3} correlates better with $T2m_{insitu}$ than with the IST_{skin_insitu}. Moreover, the IST_{skin_L3}-$T2m_{InSitu}$ difference shows smaller standard deviations than IST_{skin_L3}-IST_{skin_insitu}. However, as expected the biases and root mean squared differences (RMS) are larger for the IST_{skin_L3}-$T2m_{insitu}$ differences than for the IST_{skin_L3}-IST_{skin_insitu} differences. The reason is that the radiometric surface skin temperature can be significant different from the surface air temperature measurements (Adolph et al., 2018; Hall et al., 2008; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma et al., 2008). On average, the skin temperature is colder than the air temperature (Nielsen-Englyst et al., 2019), resulting in even more negative biases, when the IST_{skin_L3} is compared to in situ measured T2m, instead of in situ skin temperatures. The generally high correlations are dominated by the synoptic (2-5 days) and seasonal variations, which are pronounced in both IST and T2m.

3 Methods

3.1 Regression model

Nielsen-Englyst et al. (2019) analysed a large number of in situ stations with simultaneous T2m and IST_{skin} observations and showed that empirical relationships exist between T2m and IST_{skin}. However, it was also shown that the relationships varied for different regions. Based upon these results, it was decided to use a simple regression based method in this paper to derive the daily mean T2m from the satellite IST_{skin_L3} observations. Separate regression models have been derived for land ice and sea ice.

To test different types of regression models, the IST_{skin_L3} data have been matched up with in situ observations for each day (Høyer et al., 2018). This is done by requiring a distance to nearest in situ site of less than 15 km. The average matchup distance is 8.6 km and 7.2 km for land ice and sea ice, respectively, which means that all in situ observations are made within the area of the satellite pixel. The corresponding mean elevation difference is 30 m (while the absolute mean elevation difference is 45 m) and is calculated using the topography mask included in HIRHAM5 (Langen et al., 2015) for the 23 GrIS AWSs. Out of the 23 AWSs, four of them (GC-net JAR1, TAS_U, QAS_U and UPE_U) have corresponding elevation differences above 100 m. In the Section 4.3, the effect of these AWSs has been estimated and discussed. All in situ observations, described in Sect. 2.1., have been matched with IST_{skin_L3} data, resulting in a total number of daily matchups of 65,810 from 275 different observation sites (see Table 1). These have been divided into two subsets: one for training and one for validation of the different regression models for land ice and sea ice, respectively. This has been done while ensuring similar coverage of training and validation data over the two domains, which is shown in Fig. 5. The result is that 40 % (13,792 matchups) are used for testing the regression models (and generating the regression coefficients) and the remaining 60 % (20,872 matchups) are left for validation of the regression models over land ice. Over sea ice 48 % (15,035 matchups) are used for testing and 52 % (16,111 matchups) are left for validation.

The regression model is based on multiple linear regression analysis using least squares (Menke, 1989). The multiple linear regression analysis equations can be written in matrix form,

$$\boldsymbol{d^{obs}} = \mathbf{G}\boldsymbol{m} + \boldsymbol{e} \tag{4}$$

$$\boldsymbol{d^{pre}} = \mathbf{G}\boldsymbol{m}, \tag{5}$$

where $\boldsymbol{d^{obs}}$ and $\boldsymbol{d^{pre}}$ are vectors containing the observed and modelled in situ air temperatures, respectively, $\mathbf{G}$ is a matrix containing the various predictors, $\boldsymbol{m}$ is a vector containing regression coefficients, and $\boldsymbol{e}$ is the fitting error.

The regression coefficients are found using damped least squares (Menke, 1989). The least squares method is used since the problem is generally over-determined, and the damping is added to limit effects of noisy data. The regression coefficients are thus given as:

$$\mathbf{G^{-g}} = (\mathbf{G^T G} + \varepsilon^2 \mathbf{I})^{-1} \mathbf{G^T} \tag{6}$$

$$\boldsymbol{m} = \mathbf{G^{-g}} \boldsymbol{d^{obs}}, \tag{7}$$

where $\mathbf{G^{-g}}$ is called the generalized inverse, ε is a damping factor and $\mathbf{I}$ is an identity matrix (with ones in the diagonal and zeros elsewhere). The superscript operator $\mathbf{T}$ denotes transposing and -1 denotes inversion. We have tested a range of damping factors to assess the relation to the error coefficients. A damping factor of 0.2 was chosen to avoid overfitting noise in the data, while keeping the error coefficients low.

5   The choice of predictors is based on current knowledge of the parameters that influence the relationship between IST_{skin} and $T2m_{insitu}$ (Adolph et al., 2018; Hall et al., 2008; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma and Pirazzini, 2005), limited by the available satellite data. Nielsen-Englyst et al. (2019) showed that the T2m-Tskin difference varies over the season with the smallest differences during the spring, fall, and summer in non-melting conditions. For that reason, we have also tested the effect of including a seasonal cycle as predictor. A total of 5 regression models with different predictors

10   have been tested (Høyer et al., 2018):

$\hat{I}ST_{skin}$: $\qquad\qquad\qquad T2m_{sat} = \alpha_0 + \alpha_1 IST_{skin\_L3}$ (8)

$\hat{I}ST_{skin}$SWd: $\qquad\qquad T2m_{sat} = \alpha_0 + \alpha_1 IST_{skin\_L3} + \alpha_2 SWd$ (9)

$\hat{I}ST_{skin}$WS: $\qquad\qquad T2m_{sat} = \alpha_0 + \alpha_1 IST_{skin\_L3} + \alpha_2 WS$ (10)

$\hat{I}ST_{skin}$Lat: $\qquad\qquad T2m_{sat} = \alpha_0 + \alpha_1 IST_{skin\_L3} + \alpha_2 Lat$ (11)

$\hat{I}ST_{skin}$Season: $\qquad T2m_{sat} = \alpha_0 + \alpha_1 IST_{skin\_L3} + \alpha_2 COS\left(\frac{t\cdot2\pi}{1\,yr}\right) + \alpha_3 SIN\left(\frac{t\cdot2\pi}{1\,yr}\right)$ (12)

The regression model in Eq. (8) is limited to an offset and a scaling of IST_{skin_L3}, where the latter term accounts for the synoptic and seasonal variations, which are the dominating factors in both the IST and T2m variability. This part is thus included in all regression models tested. The other regression models also have a third predictor, which is included to examine how to best

15   represent the residual variations in the T2m-IST difference. The model in Eq. (9) uses theoretical top of atmosphere shortwave radiation, Eq. (10) uses the wind forcing (from ERA-I and ERA5, respectively), Eq. (11) uses latitude variation, and Eq. (12) uses a seasonal variation. In the regression model in Eq. (12), the seasonal variation is assumed to be the shape of a cosine function, $A \cdot COS\left(\frac{t\cdot2\pi}{1\,yr} - \varphi\right)$, where A is the amplitude, φ is the phase and t is time. Since $COS(x_1 - x_2) = COS(x_1)COS(x_2) + SIN(x_1)SIN(x_2)$, the seasonal cycle can be rewritten to the form in Eq. (12) with $A = \sqrt{\alpha_2{}^2 + \alpha_3{}^2}$ and $\varphi = ARCTAN\left(\frac{\alpha_3}{\alpha_2}\right)$.

20   The training data have been used to calculate the regression coefficients for each regression model covering the land ice and sea ice, respectively. The performance of each regression model has been investigated using the training data and the results are shown in Table 3. The best performance is found by using the regression model where $T2m_{sat}$ is predicted from IST_{skin_L3} combined with a seasonal variation ($\hat{I}ST_{skin}$Season). This model predicts $T2m_{sat}$ better compared to the other regression models, with correlations above 96 % and RMS values of 3.25-3.28°C against training data for both surface types (Table 3). In the

25   following, we will use the regression model given in Eq. (12) with the seasonal term included and with separate regression

coefficients for land ice and sea ice (see Table 4). The phase corresponds to a maximum the 19th January and 12th February for land ice and sea ice, respectively. This is in agreement with Nielsen-Englyst et al. (2019) who found the strongest clear-sky inversion during the winter months (Dec-Feb) for all sites included in the analysis except from the ones located in the lower ablation zone (not included here), where pronounced surface melt takes place for long periods of time.

3.2 Uncertainty estimates for T2m$_{sat}$

Uncertainty estimates on the derived T2m$_{sat}$ are crucial to facilitate the usage of the data set in modelling and for monitoring purposes. The uncertainty estimates of the satellite-derived T2m$_{sat}$ data follow the approach in Bulgin et al. (2016) and Rayner et al. (2015), which has also been used for the AASTI data. The uncertainty on a single T2m$_{sat}$ estimate is divided into random,

10 locally correlated and systematic uncertainty components, with the total uncertainty μ$_{total\_t2m}$ given as the square root of the sum of the three squared components:

$$\mu_{total\_T2m} = \sqrt{\mu_{rnd\_T2m}{}^2 + \mu_{local\_T2m}{}^2 + \mu_{glob\_T2m}{}^2}$$

The random uncertainty component for the T2m$_{sat}$ belonging to a particular grid cell at a particular point in time is found by propagating the AASTI IST$_{skin\_L3}$ random uncertainty through the regression model:

15 $$\mu_{rnd\_T2m} = \sqrt{\left(\alpha_1 \mu_{rnd\_L3}\right)^2},$$

with μ$_{rnd\_L3}$ given as the aggregated μ$_{rnd\_L2}$:

$$\mu_{rnd\_L3} = \frac{\mu_{rnd\_L2}}{\sqrt{N}},$$

where N is the number of observations for each bin in the aggregation from L2 to L3. The $\sqrt{N}$ reduction applies because the random uncertainty of each L2 data point that goes into the L3 calculation is by definition independent from the other.

20 The L3 global uncertainty component does not average out in any aggregation and is thus transferred directly from the L2 uncertainty estimate and multiplied by α_1 to make up μ_{glob_T2m}:

$$\mu_{glob\_T2m} = \alpha_1 \mu_{glob\_L3} = \alpha_1 \cdot 0.1°C$$

The μ$_{local\_T2m}$ contains both the local uncertainty component of L2, a sampling error μ$_{lsamp\_L3}$ related to sampling errors in space and time due to the aggregation, a relationship error, cloud mask uncertainty etc. When aggregating from L2 to daily L3,

25 additional sources of uncertainty enter through the gridding process as IST$_{skin\_L3}$ can only be retrieved for clear sky pixels. This introduces a temporal and spatial sampling uncertainty. If all our satellite observations were obtained during all-sky conditions, we assume that the high polar temporal coverage is such that the temporal sampling uncertainty in the L3 files can be set to zero. However, this is not the case and using only clear-sky observations generally leads to a clear-sky bias in averaged IST$_{skin}$ satellite observations when compared to in situ observations (Hall et al., 2012; Nielsen-Englyst et al., 2019; Rasmussen et al.,

30 2018). The relationship error represents the standard deviation of the residuals calculated at in situ stations, where both skin and air temperatures are available, i.e. T2m$_{sat}$-T2m$_{insitu}$. Estimating all the different components that make up the μ$_{local\_T2m}$ is a

very challenging task and is out of the scope of this paper. Instead, we estimate the μ_{local_T2m} component using a simple regression model fitted to the satellite derived T2m and in situ T2m differences. Separate models have been chosen for the land ice and sea ice, due to the differences in the error characteristics. The variables to include in the uncertainty regression models have been chosen from a careful examination of the matchup data set. For land ice and sea ice the most relevant

5   variables were the IST_{skin_L3} itself and the number of 3 h time bins with observations in the L3, N_{bins}.

For land ice the regression model for μ_{local_T2m} is given as following:

$$\mu_{local\_T2m\_landice} = \beta_0 + \beta_1 IST_{skin\_L3} + \beta_2 N_{bins} ,$$

while the regression model for sea ice is given as:

$$\mu_{local\_T2m\_seaice} = \gamma_0 + \gamma_1 IST_{skin_{L3}} + \gamma_2 IST_{skin_{L3}}{}^2 + \gamma_3 N_{bins} .$$

10 The coefficients have been determined by fitting to the $T2m_{sat}$-$T2m_{insitu}$ standard deviations calculated for the training data with IST_{skin_L3} bin intervals of 2°C and N_{bins} interval of 1. The μ_{rnd_T2m} and μ_{glob_T2m} components have been removed from the standard deviations in each bin as well as an assumed in situ uncertainty of 0.1°C and an average sampling uncertainty of 0.5°C (Høyer et al., 2017a; Reeves Eyre and Zeng, 2017) before fitting the regression models. The optimal regression coefficients for each domain are listed in Table 5.

15 **4 Results**

In Sect. 3.1, we selected the best (Eq. 12) of the five different algorithms and used it together with the derived coefficients (Table 3 and 4) to retrieve T2m from satellite surface temperature estimates. The derived dataset consists of daily estimates of near surface air temperature on a 0.25 degree regular latitude-longitude grid, during the period 2000-2009 (Høyer et al., 2018; Kennedy et al., 2019). Days with clouds and few clear sky observations (as explained in Section 2.2) are not included in the

20 dataset. However, for those days when the satellite derived T2m product is available, it provides an estimate of the daily averaged all-sky T2m (see Sect. 5). Each temperature estimate is associated with three components of uncertainty on the 0.25 degree daily scale: a random uncertainty, a synoptic scale correlated uncertainty and a globally correlated uncertainty excluding uncertainties related to the masking of clouds. The three types of uncertainties are also gathered in a total uncertainty estimate (see Sect. 3.2). The land ice temperatures have been calculated for grid cells categorized as ice sheet by ETOPO1 global relief

25 model (Amante and Eakins, 2009), averaged to the 0.25 degree grid. Sea ice temperatures have been calculated for grid cells with sea ice concentrations above 30 %, according to OSISAF (Tonboe et al., 2016).

4.1 Validation of T2m$_{sat}$

The derived T2m$_{sat}$ product has been validated against independent in situ data (i.e. the validation subset described in Sect. 3.1). Figure 6 shows an example of the daily near surface air temperature coverage (from Jan 1st, 2008). Circles are in situ

30 T2m measurements from coincidence independent AWSs and buoys and there seems to be quite good agreement between these and T2m$_{sat}$ during this specific day. The overall model performance, when compared to all independent AWS and buoy

observations, is summarized in Table 6. The satellite derived air temperatures are about 0.3°C warmer than measured in situ air temperature for both land ice and sea ice. For the GrIS, the bias is partly explained by topographic effects (see Sect. 4.3). The correlations are above 95 % for both surface types and the RMS is 3.47°C and 3.20°C for land and sea ice, respectively. Note that the uncertainty of the in situ data is also included in these RMS values.

5   Figure 7 shows the average seasonal variation in bias and standard deviation for land ice and sea ice, respectively. For both land ice and sea ice, there is a seasonal dependency in standard deviation with the largest values during the winter and smallest values during the summer. This is likely explained by a better cloud screening performance during sunlit periods (Karlsson and Dybbroe, 2010) and by the smaller natural thermal variability that is observed during summer conditions. Similar seasonality in performance is seen in five reanalysis products (including ERA-I/ERA5) for the GrIS (Zhang et al., 2021). As

10   shown in Fig. 7, the average seasonal variation in bias is largest over sea ice with the largest values in March and August. However, this seasonal tendency in bias over sea ice is only reflected at the beginning of the time period (i.e. 2000-2004). This can be seen in Fig. 8 which shows the seasonal averaged independent validation statistics for the entire period for land ice and sea ice, respectively. The figure also shows a quite stable performance over the time period for both land ice and sea ice.

As more satellite observations have become available over the time period, increased coverage of the surface temperature is

15   expected over time. Figure 9 shows the average number of filled 3 h bins per day for the GrIS and Arctic sea ice for 2000-2009. Both surface types show an increase in filled 3 h bins over time, with large seasonal variations. In most years, sea ice has 1-1.5 filled bins per day more during winter than summer, due to a more extensive cloud cover over sea ice during summer (Curry et al., 1996; Beesley and Moritz, 1999). The GrIS typically has fewer filled bins per day during the winter and summer, than spring and fall, which is also explained by differences in cloud coverage (Griggs and Bamber, 2008). Note that the increase

20   in the average number of filled 3 h bins from 2000 to 2009 is not reflected in the performance of the T2m product (Figure 8). Figure 10 shows $T2m_{sat}$-$T2m_{insitu}$ differences plotted as a function of AASTI L3 skin temperature for land ice and sea ice, respectively. Over land ice, the standard deviation decreases as a function of IST_{skin_L3}, while the bias is around zero for IST_{skin_L3} between -45°C and -10°C, positive for higher temperatures and negative for lower temperatures. For sea ice, the maximum standard deviation is found at skin temperatures of about -20°C, with smaller standard deviations for higher and

25   lower IST_{skin_L3}. Positive biases are found for very cold skin temperatures (< -25°C) and for temperatures around the melting point (> -4°C), while the intermediate temperatures have a slightly negative bias. This effect is included in the uncertainty estimates as presented in Sect. 3.2, which include IST_{skin_L3} as a predictor for both land ice and sea ice.

Figure 11 shows the validation results of the estimated uncertainties, where the $T2m_{sat}$-$T2m_{insitu}$ difference is plotted against the theoretical total uncertainties as obtained in Sect. 3.2 for land ice and sea ice, respectively. The dashed lines represent the

30   ideal uncertainty with the assumptions that the in situ observations have an uncertainty of 0.1°C and that the sampling uncertainty is 0.5°C. The estimated uncertainties show good agreement with the observed uncertainties when the error bars follow the dashed line, which is the case here for both land ice and sea ice.

4.2 Comparison with reanalyses

The performance of T2m$_{sat}$ has been compared to the performance of T2m from ECMWF's reanalysis ERA-I (T2m$_{ERA-I}$; Dee et al., 2011) and the replacement reanalysis, ERA5 (T2m$_{ERA5}$; Hersbach et al., 2020). Table 7 shows the performance of T2m$_{ERA-I}$ and T2m$_{ERA5}$ against the independent in situ T2m observations, which should be compared with the performance of the regression derived T2m$_{sat}$ as shown in Table 6. The comparison may not be truly independent as a number of stations and buoys have been assimilated into the ERA-I and ERA5 data products (Dee et al., 2011; Hersbach et al., 2020), which would favour the reanalysis products in the comparison. Yet, the bias is significantly lower for T2m$_{sat}$ than for both T2m$_{ERA-I}$ and T2m$_{ERA-5}$, while the other validation parameters are similar, with slightly better correlation and standard deviation, but slightly worse RMS results for T2m$_{ERA}$. Previous studies have also found that ERA-I suffers from a consistent warm bias in the Arctic (Lüpkes et al., 2010; Jakobson et al., 2012; Vihma et al., 2002; Batrak and Müller, 2019; Simmons and Poli, 2014), and recent studies suggest that the warm bias still exists in ERA5 over sea ice (Wang et al., 2019; Graham et al., 2019). Similarly, recent studies found no significant improvements in 2 m temperatures over the GrIS for ERA5 compared to ERA-I (Delhasse et al., 2020; Zhang et al., 2021). Note, however, that the NCEP-CFSR, which is based on a coupled atmosphere – sea ice – ocean model, has shown better performance than ERA-I for near-surface atmospheric variables over sea ice (Jakobson et al., 2012).

Figure 12 shows the RMS difference between in situ measured T2m and T2m$_{ERA-I}$, T2m$_{ERA5}$ and T2m$_{sat}$, respectively, for the individual validation sites and both surface types. Due to the large number of buoys these have been validated for each data source with all observations weighted equally. The last bars refer to the RMS obtained by validating all validation sites in one long time series weighting all daily observations equally. The total T2m$_{sat}$ agrees better with in situ observations for both surface types compared to both ERA-I and ERA5. For most land ice stations, the T2m$_{sat}$ outperforms ERA-I and ERA5. One exception is the ARM station (BAR), where a bias of 2.49°C gives rise to a relatively large RMS error for T2m$_{sat}$. This is likely explained by physical differences between the seasonal snow covered sites and the GrIS sites, which are not fully captured by the regression model. ERA5 is significantly better than ERA-I over the GrIS, but ERA5 performs worse than both ERA-I and T2m$_{sat}$ over sea ice. Over sea ice, T2m$_{ERA-I}$ agrees better with in situ observations from ECMWF data stream and *Polarstern*. However, these may be assimilated into both ERA-I and ERA5. The validation against *Polarstern* is relatively good even though the temperature measurements are made at 29 metres height. This is likely because the data are mainly from the summer, when the vertical temperature gradients in the boundary layer are mostly small, and the performance of the cloud screening algorithm reaches its maximum. The independent in situ observations by ACSYS, CRREL, DAMOCLES and FRAMZY are better reproduced by the satellite-derived T2m. The errors in the T2m$_{ERA-I}$/T2m$_{ERA5}$ and T2m$_{sat}$ data sets are expected to be independent and uncorrelated. For that reason, a combination of either T2m$_{ERA-I}$ or T2m$_{ERA5}$ and T2m$_{sat}$ can lead to an improved T2m estimate.

4.3 Topographic effects

The effects from topography over the GrIS have been assessed by introducing a new matchup data set that ensures that the elevation difference between satellite and in situ observations is less than 100 m over the GrIS. Excluding those AWSs (4 out of 23) with a larger elevation difference than 100 m results in a reduction of the training data set of 2,935 matchups (i.e. from

5   GC-net_JAR1 and PROMICE TAS_U) and a reduction in the validation data set of 560 matchups (i.e. from PROMICE QAS_U and UPE_U). The performance of the satellite derived T2m improves, in particular the bias, which decreases to 0.07ºC, while the standard deviation decreases to 3.41ºC over land ice. ERA-I and ERA5 show limited changes in performance, with slightly increased biases of 3.48ºC and 2.07ºC, and standard deviations of 3.14ºC and 3.08ºC, respectively, when introducing the new matchup data set over land ice. A similar good performance of the regression model is found when the 2 AWSs in the

10   validation subset are kept. Despite the increased performance of the regression model, we have included all observations in the training of the model to ensure a robust and spatial representative solution.

4.4 Analysis of T2m$_{sat}$

The monthly mean T2m$_{sat}$ is shown in Fig. 13 for March, June, September and December averaged over the period 2000-2009. The interior and northern part of the GrIS is typically colder than other parts of the Arctic in all months, while the warmest

15   regions are found along the sea ice marginal ice zone and the ablation zone of the GrIS. Limited spatial variability is seen over the Arctic sea ice during summer.

Figure 14 shows the monthly mean near surface air temperature estimates averaged over the GrIS for the period 2000-2009. The GrIS records a distinct annual cycle in near surface air temperature, with the maximum temperatures of around of -4°C during July and minimum temperatures of about -28°C during winter. The range in monthly mean air temperature is in

20   agreement with those reported by van As et al. (2011) at a number of PROMICE AWSs. The temporal variability is largest during winter due to a larger cloud radiative effect (compared to near-zero during summer) and a larger meridional temperature gradient resulting in a more vigorous atmospheric circulation in winter (Serreze et al., 1993). In addition, the temporal variability is lower during summer due to the fact that when the surface begins to melt, the sensible heat is used for melting and hence reducing surface air temperature variability (Steffen, 1995).

25   As illustrated in Fig. 15, T2m$_{sat}$ provides increasing coverage over the period 2000-2003 and quite stable coverage for the years 2003-2009. The average daily coverage is 84 % and 67 % for land ice and sea ice, respectively, for the stable 2003-2009 period and the 0.25 degree grid. When considering a 1 degree grid resolution, these numbers increase to 94 % and 81 %, respectively. Over land ice, the maximum coverage is during the spring and fall, while the sea ice coverage has a clear drop in coverage during the summer due to increased cloud cover (Curry et al., 1996; Beesley and Moritz, 1999).

5 Discussion

Due to the limited number of in situ observations in the Arctic, and especially over sea ice, gathering in situ observations for testing and validating the regression models is not a simple task. The lack of observations that represent all conditions and regions in the Arctic, and the resulting matching threshold of 15 km combined with the large topographical variations over the

5 GrIS, increase the uncertainty in the pixel-to-point comparison thereby complicating the derivation and validation of the regression models. Despite this, the validation against independent in situ observations and the comparison with ERA-I and ERA5 demonstrate the value of the $T2m_{sat}$ product in the Arctic.

Five regression models were tested, and the best regression model predicts $T2m_{sat}$ from daily satellite IST_{skin_L3} combined with a seasonal variation. The performance of the $T2m_{sat}$ product did not improve much when the wind speed information from

10 ERA-I or ERA5 (Table 3) was included despite the fact that previous studies have shown a strong dependency of wind speed for both land ice and sea ice (Adolph et al., 2018; Hudson and Brandt, 2005; Miller et al., 2013; Nielsen-Englyst et al., 2019). This was unexpected, at least for sea ice. The reason is likely that the quality of the wind speed fields is not adequate for use in the relationship model. In particular, accurately representing katabatic winds in numerical weather prediction (NWP) models is a challenging task due to the high resolution needed in the vertical direction (Grisogono et al., 2007; Steeneveld, 2014; Weng

15 and Taylor, 2003; Zilitinkevich et al., 2006). Furthermore, the representation of surface roughness and the processes of snow-surface coupling, radiation and turbulent mixing are hampered by limited resolution, while the relative importance of the processes varies with wind speed (Sterk et al., 2013). More accurate information on the wind speed is expected to improve the performance of the regression model which includes wind speed as predictor. In particular, the higher resolution NWP output may be very beneficial in the regions of the GrIS, where the local topography interacts with the wind through katabatic effects

20 (DuVivier and Cassano, 2013; Oltmanns et al., 2015; Renfrew, 2004). Regional high resolution reanalysis products are currently being developed within the Copernicus Arctic regional Reanalysis service C3S project (https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service). It is likely that such products will provide winds that can be used within a relationship model.

Since infrared satellites cannot measure the surface temperature during cloudy conditions, a cold clear-sky bias is often

25 observed in infrared satellite IST_{skin_L3} averages compared to all-sky temperature averages (see, e.g., Table 2; Hall et al., 2008; Koenig and Hall, 2010). When using satellite IST_{skin_L3} observations, it is thus important to be aware of the clear-sky bias, which moreover varies with different temporal averaging windows (Nielsen-Englyst et al., 2019). Here, when using an empirical statistical method, which is trained against daily averaged in situ T2m (obtained both in clear sky and cloudy conditions), the conversion from IST_{skin_L3} to $T2m_{sat}$ removes the systematic IST_{skin_L3} clear-sky bias effects that may be present

30 in the satellite data. As a result, we obtain a $T2m_{sat}$ estimate which performs similar or better than the IST_{skin_L3}, when validated against in situ observations. For the IST_{skin_L3}, the temporal sampling errors resulting from clouds have been minimized through a number of requirements. For short-lasting (<24 hours) cloudy conditions, the division into 3 h bin averages and the requirement of filled 3 h bins both during the night (between 18:00 and 6:00 local solar time) and day (between 6 and 18 local

solar time) ensure that the diurnal cycle is best resolved despite the gaps with clouds. For long-lasting ($\geq$ 24 hours) cloudy conditions, IST_{skin_L3} is not available, and we do not retrieve $T2m_{sat}$ for these days.

The $T2m_{sat}$ product derived here provides increasing coverage over the period 2000-2003 and stable coverage for 2003-2009. The coverage varies with the season with the minimum coverage over sea ice in the period from July to September due to

5   extensive cloud cover over the Arctic sea ice during summer (Curry et al., 1996; Beesley and Moritz, 1999). Nevertheless, the average daily coverage is 84 % and 67 % for land ice and sea ice, respectively, for the stable 2003-2009 period. The high percentages in coverage demonstrate that the gaps due to cloudy days are limited (except for over sea ice in the summer) and that the data set contains a significant amount of information on the all-sky daily T2m even though it is based on clear sky satellite observations.

10   Atmospheric models using data assimilation or statistical techniques may be applied to fill in the gaps due to clouds. This has already been done in the EUSTACE project by using an advanced statistical model to combine in situ observed and clear sky satellite derived T2m estimates (over land, lakes, ocean and ice), including uncertainty estimates, into a global and gap-free daily analysis of surface air temperatures from 1850 to 2015 (Morice et al., 2019; Rayner et al., 2020). The $T2m_{sat}$ product derived in this paper is used as input to the EUSTACE surface air temperature analysis for the GrIS and the Arctic sea ice.

15   The $T2m_{sat}$ data set developed here only covers the Arctic, but the AASTI satellite data set also covers the Antarctica. This implies that similar statistical methods can be derived for the Antarctic ice sheet and sea ice. Preliminary investigations indicate that a T2m product can be derived for the Antarctic ice sheet with similar performance to GrIS, whereas the Southern Ocean sea ice is challenging due to very few in situ observations (Morice et al., 2012). For both Southern regions, more in situ observations are needed to repeat the work performed for the Arctic and to determine a reliable statistical model. This product

20   can also be extended to seasonal snow and ice, but it requires a dynamic surface mask and that the derivation of the regression model is repeated. However, similar efforts have already been made within EUSTACE to cover seasonal snow (Good, 2015; Morice et al., 2019; Rayner et al., 2020).

The AASTI version builds on the Clara version 1 data set from the CM-SAF. A version 2 of the data set is now available (Karlsson et al., 2017), which facilitates the production of an AASTI version 2 data set that covers the period 1982 up to

25   present. With consistency in the retrieval algorithm and data sets, it will be possible to use the relationship model to produce a satellite-based climate data record of T2m from 1982 to today.

Including other available satellite products such as MODIS IST observations (Hall et al., 2004) or the (A)ATSR data set (Ghent et al., 2017) may improve the quality of the $T2m_{sat}$ product. However, adding new data requires detailed knowledge of the characteristics of the data set such as sampling frequency and uncertainty of the IST observations. In addition, determination

30   of the relationship model is needed again. At the same time, adding more satellite overpasses to the daily estimates may not reduce the uncertainty of the products. This is evident when comparing Fig. 7 and 8 where the variation in the number of satellite observations during the record (Fig. 8) is not reflected in a similar variation in the performance of the product (Fig. 7). The uncertainty in the beginning of the record is comparable to the uncertainty at the end of the record, despite an almost doubling of the observed 3-hourly averages throughout the day.

6 Conclusions

The surface air temperature is one of the key indicators for Arctic climate change, and it can easily be compared with climate change indicators from other regions. This study introduces a methodology for using satellite skin temperatures for estimating air temperatures to compensate for the lack of in situ measurements and as a supplement to reanalysis products in the Arctic.

Daily near surface air temperatures (T2m) have been estimated based on daily clear sky satellite Level 3 (L3) observations of ice surface skin temperatures (IST_{skin_L3}), using the Arctic and Antarctic ice Surface Temperatures from thermal Infrared satellite sensors (AASTI) reanalysis. A regression based method has been used and tuned against in situ observed T2m using IST_{skin_L3} observations covering both Arctic sea ice and the Greenland Ice Sheet (GrIS). In general, there is a good correlation between T2m and IST_{skin_L3} due to the seasonal cycle in both IST and T2m. Different models have been tested to examine how to best capture the variability in the T2m-IST difference. The highest correlation and lowest RMS were found using a model where $T2m_{sat}$ is predicted from daily satellite IST_{skin_L3} combined with a seasonal variation, assumed to have the shape of an annual harmonic. This model has been used to derive daily T2m on a 0.25 degree regular latitude-longitude grid from the clear sky AASTI IST_{skin_L3} over the Arctic during the time period 2000-2009 (Kennedy et al., 2019), using different regression coefficients for land ice and sea ice. Days with clouds or limited clear sky observations have been excluded from the analysis. Considering a 1 degree regular latitude-longitude grid, the average daily coverage of the $T2m_{sat}$ product is 94 % over the GrIS and 81 % for sea ice for the years 2003-2009. The days when the $T2m_{sat}$ is available, the T2m estimate can be considered as a daily averaged all-sky T2m, since it has been tuned against all-sky in situ observations.

The estimated $T2m_{sat}$ data show average biases of 0.30°C and 0.35°C and average root mean square errors of 3.47°C and 3.20°C for land ice and sea ice, respectively, when validated against independent in situ observations. All daily $T2m_{sat}$ estimates include a total uncertainty estimate divided into a random, locally systematic, and large-scale systematic uncertainty component. The total uncertainty of $T2m_{sat}$ shows good validation results when validated against independent in situ observations. A comparison with two of ECMWF's reanalyses (i.e. ERA-I and ERA5) shows that $T2m_{sat}$ validates similar or better than both of these even though the reanalyses actively assimilate available in situ observations. The $T2m_{sat}$ product is independent of the quality of the NWP forecasts, and thus, it represents an important supplement to the model based T2m. The errors in NWP products (e.g. $T2m_{ERA-I}$ or $T2m_{ERA5}$) and the errors in the product derived here ($T2m_{sat}$) are expected to be independent and uncorrelated, and a combination of a NWP product and the $T2m_{sat}$ data can therefore lead to an even better T2m estimate. The regression models presented here both work on satellite observations that are available from reprocessed records but opens up for a near real time estimation of T2m from satellites. The results obtained for the ice covered areas show that there is a large potential for using satellite observed surface temperatures to estimate near surface air temperatures. These estimates are not supposed to replace the already existing air temperature measurements or reanalyses, but rather to supplement these e.g. in areas where no in situ observations are currently available.

7. Data availability

The derived surface air temperatures from satellite surface skin temperatures over ice can be downloaded from http://dx.doi.org/10.5285/f883e197594f4fbaae6edebafb3fddb3 (Kennedy et al., 2019). The PROMICE data can be accessed through http://www.promice.dk (last access: 16 November 2018). The ARM data are available at https://www.archive.arm.gov/discovery/\#v/results/s/s::co (last access: 21 December 2018). GC-Net data can be found through doi:10.5067/6S7UHUH2K5RI (Greenland Climate Network (GC-Net) Radiation for Arctic System Reanalysis, Version 1., 2016). Data from CRREL mass balance buoys are available from: http://imb-crrel-dartmouth.org (last access: 24 November 2016), while *Polarstern* data can be downloaded at https://dship.awi.de/Polarstern.html (last access: 24 November 2016. FRAMZY data are available from doi:10.1594/WDCC/UNI_HH_MI_FRAMZY2002 (Brümmer et al., 2012b), http://dx.doi.org/doi:10.1594/WDCC/UNI_HH_MI_FRAMZY2007 (Brümmer et al., 2011b), and http://dx.doi.org/doi:10.1594/WDCC/UNI_HH_MI_FRAMZY2008 (Brümmer et al., 2011c), while ACSYS data are found here: DOI: 10.1594/WDCC/UNI_HH_MI_ACSYS2003. Damocles data can be found here: doi:10.1594/wdcc/uni_HH_MI_DAMOCLES2007 (Brümmer et al., 2011a). The traditional buoy and ship data obtained from ECMWF are distributed through the World Meteorological Organization's (WMO) Global Telecommunication System (GTS) and available for members at the ECMWF Meteorological Archival and Retrieval System (MARS). Finally, the AASTI IST_{skin_L2} data are available from http://dx.doi.org/10.5285/60b820fa10804fca9c3f1ddfa5ef42a1 (Høyer et al., 2019).

8 Author contribution

Pia Nielsen-Englyst, Kristine S. Madsen and Gorm Dybkjær compiled and quality checked the in situ data. Pia Nielsen-Englyst, Jacob L. Høyer and Kristine S. Madsen designed and developed the regression model and estimated uncertainties. Gorm Dybkjær, Jacob L. Høyer and Rasmus Tonboe developed the AASTI IST_{skin_L2} data. Sotirios Skarpalezos did the ERA5 matchup. Pia Nielsen-Englyst prepared the manuscript with contributions from all authors.

9 Competing interests

The authors declare that they have no conflict of interest.

10 Acknowledgements

This study was carried out as a part of the European Union Surface Temperatures for All Corners of Earth (EUSTACE), which is financed by the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171. The aim of EUSTACE is to provide a spatially complete daily field of air temperatures since 1850 by combining satellite and in situ observations. The author would also like to thank the data providers.

References

Abermann, J., Hansen, B., Lund, M., Wacker, S., Karami, M., and Cappelen, J.: Hotspots and key periods of Greenland climate change during the past six decades, Ambio, 46, 3–11, https://doi.org/10.1007/s13280-016-0861-y, 2017.

Ackerman, T. P. and Stokes, G. M.: The Atmospheric Radiation Measurement Program, Phys. Today, 56, 38–44, https://doi.org/10.1063/1.1554135, 2003.

Adolph, A. C., Albert, M. R., and Hall, D. K.: Near-surface temperature inversion during summer at Summit, Greenland, and its relation to MODIS-derived surface temperatures, The Cryosphere, 12, 907–920, https://doi.org/10.5194/tc-12-907-2018, 2018.

Ahlstrøm, A., van As, D., Citterio, M., Andersen, S., Fausto, R., Andersen, M., Forsberg, R., Stenseng, L., Lintz Christensen, E., and Kristensen, S. S.: A new Programme for Monitoring the Mass Loss of the Greenland Ice Sheet, Geol. Surv. Den. Greenl. Bull., 15, 61–64, 2008.

Amante, C. and Eakins, B. W.: ETOPO1 Global Relief Model converted to PanMap layer format, https://doi.org/10.1594/PANGAEA.769615, 2009.

van As, D.: Warming, glacier melt and surface energy budget from weather station observations in the Melville Bay region of northwest Greenland, J. Glaciol., 57, 208–220, https://doi.org/10.3189/002214311796405898, 2011.

van As, D., Fausto, R. S., Ahlstrøm, A., Andersen, S., Citterio, M., Edelvang, K., Graversen, P., Machguth, H., Nick, F., Nielsen, S., and Weidick, A.: Programme for Monitoring of the Greenland Ice Sheet (PROMICE): first temperature and ablation record., Geol. Surv. Den. Greenl. Bull., 73–76, 2011.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nat. Commun., 10, 4170, https://doi.org/10.1038/s41467-019-11975-3, 2019.

Beesley, J. A. and Moritz, R. E.: Toward an explanation of the annual cycle of cloudiness over the Arctic Ocean, J. Clim., 12, 395–415, 1999.

Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F.-J. W., Brown, R., Bhatt, U. S., Euskirchen, E. S., Romanovsky, V. E., Walsh, J. E., Overland, J. E., Wang, M., Corell, R. W., Meier, W. N., Wouters, B., Mernild, S., Mård, J., Pawlak, J., and Olsen, M. S.: Key indicators of Arctic climate change: 1971–2017, Environ. Res. Lett., 14, 045010, https://doi.org/10.1088/1748-9326/aafc1b, 2019.

Brümmer, B., Müller, G., Haller, M., Kriegsmann, A., Offermann, M., and Wetzel, C.: DAMOCLES 2007-2008 - Hamburg Arctic Ocean Buoy Drift Experiment: meteorological measurements of 16 autonomous drifting ice buoys, https://doi.org/10.1594/wdcc/uni_HH_MI_DAMOCLES2007, 2011a.

Brümmer, B., Müller, G., Lammert-Stockschläder, A., Jahnke-Bornemann, A., and Wetzel, C.: FRAMZY 2007 - Third Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of the research aircraft Falcon, 16 autonomous ice buoys and 13 autonomous water buoys, https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2007, 2011b.

Brümmer, B., Müller, G., and Wetzel, C.: FRAMZY 2008 - Fourth Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of 7 autonomous ice buoys, https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2008, 2011c.

Brümmer, B., Launiainen, J., Müller, G., Kirchgaessner, A., and Wetzel, C.: ACSYS 2003 - Arctic Atmospheric Boundary Layer and Sea Ice Interaction Study north of Spitsbergen: meteorological measurements of the research aircraft Falcon, 11 autonomous ice buoys and radiosoundings at the research vessels Aranda and Polarstern, https://doi.org/10.1594/WDCC/UNI_HH_MI_ACSYS2003, 2012a.

5 Brümmer, B., Launiainen, J., Müller, G., and Wetzel, C.: FRAMZY 2002 - Second Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of the research aircraft Falcon, 15 autonomous ice buoys and radiosoundings at the research vessel Aranda, https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2002, 2012b.

Bulgin, C. E., Embury, O., and Merchant, C. J.: Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data, Remote Sens. Environ., 177, 287–
10 294, https://doi.org/10.1016/j.rse.2016.02.021, 2016.

Cappelen (ed), J.: Greenland—DMI Historical Climate Data Collection 1768-2020., DMI Rep. 21-02, Copenhagen, Denmark, Danish Meteorological Institute, Copenhagen, Denmark, 2021.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., and Wehner, M.: Long-term Climate Change: Projections, Commitments
15 and Irreversibility, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1029–1136, https://doi.org/10.1017/CBO9781107415324.024, 2013.

Cowtan, K. and Way, R.: Update to "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature
20 trends". Reconciling global temperature series., https://doi.org/10.13140/RG.2.1.4334.8564, 2014.

Curry, J. A., Schramm, J. L., Rossow, W. B., and Randall, D.: Overview of Arctic Cloud and Radiation Characteristics, J. Clim., 9, 1731–1764, https://doi.org/10.1175/1520-0442(1996)009<1731:OOACAR>2.0.CO;2, 1996.

Davy, R. and Outten, S.: The Arctic Surface Climate in CMIP6: Status and Developments since CMIP5, J. Clim., 33, 8047–8068, https://doi.org/10.1175/JCLI-D-19-0990.1, 2020.

25 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc.,
30 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delhasse, A., Kittel, C., Amory, C., Hofer, S., van As, D., S. Fausto, R., and Fettweis, X.: Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet, The Cryosphere, 14, 957–965, https://doi.org/10.5194/tc-14-957-2020, 2020.

DuVivier, A. K. and Cassano, J. J.: Evaluation of WRF Model Resolution on Simulated Mesoscale Winds and Surface Fluxes
35 near Greenland, Mon. Weather Rev., 141, 941–963, https://doi.org/10.1175/MWR-D-12-00091.1, 2013.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Thresholds and Radiative Transfer Modeling. Part I: Algorithm Description, J. Appl. Meteorol., 44, 39–54, https://doi.org/10.1175/JAM-2188.1, 2005a.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Thresholds and Radiative Transfer Modeling. Part II: Tuning and Validation, J. Appl. Meteorol., 44, 55–71, https://doi.org/10.1175/JAM-2189.1, 2005b.

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

Dybkjær, G., Høyer, J. L., Tonboe, R., and Olsen, S. M.: Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST., NACLIM Deliverable D32.28, 2014.

Dybkjær, G., Eastwood, S., Borg, A. L., Høyer, J. L., and Tonboe, R.: Algorithm theoretical basis document (ATBD) for the OSI SAF Sea and Sea Ice Surface Temperature L2 processing chain. OSI205a and OSI205b, http://osisaf.met.no/docs/osisaf_cdop2_ss2_pum_ice-conc_v1p4.pdf, 19 February 2018.

Ghent, D. J., Corlett, G. K., Göttsche, F.-M., and Remedios, J. J.: Global Land Surface Temperature From the Along-Track Scanning Radiometers: Global LST from the ATSRs, J. Geophys. Res. Atmospheres, 122, 12,167-12,193, https://doi.org/10.1002/2017JD027161, 2017.

GHRSST Science Team: The Recommended GHRSST Data Specification (GDS) 2.0, document revision 4, available from the GHRSST International Project Office, 2011, pp 123., 2010.

Good, E.: Daily minimum and maximum surface air temperatures from geostationary satellite data: Satellite min and max air temperatures, J. Geophys. Res. Atmospheres, 120, 2306–2324, https://doi.org/10.1002/2014JD022438, 2015.

Good, E. J., Ghent, D. J., Bulgin, C. E., and Remedios, J. J.: A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series, J. Geophys. Res. Atmospheres, 122, 9185–9210, https://doi.org/10.1002/2017JD026880, 2017.

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., and Hudson, S. R.: Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer, J. Clim., 32, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1, 2019.

Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., and Svensson, G.: Vertical structure of recent Arctic warming, Nature, 451, 53–56, https://doi.org/10.1038/nature06502, 2008.

Griggs, J. A. and Bamber, J. L.: Assessment of Cloud Cover Characteristics in Satellite Datasets and Reanalysis Products for Greenland, J. Clim., 21, 1837–1849, https://doi.org/10.1175/2007JCLI1570.1, 2008.

Grisogono, B., Kraljević, L., and Jeričević, A.: The low-level katabatic jet height versus Monin-Obukhov height, Q. J. R. Meteorol. Soc., 133, 2133–2136, https://doi.org/10.1002/qj.190, 2007.

Hall, D., Box, J., Casey, K., Hook, S., Shuman, C., and Steffen, K.: Comparison of satellite-derived and in-situ observations of ice and snow surface temperatures over Greenland, Remote Sens. Environ., 112, 3739–3749, https://doi.org/10.1016/j.rse.2008.05.007, 2008.

Hall, D. K., Key, J. R., Case, K. A., Riggs, G. A., and Cavalieri, D. J.: Sea ice surface temperature product from MODIS, IEEE Trans. Geosci. Remote Sens., 42, 1076–1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Hall, D. K., Comiso, J. C., DiGirolamo, N. E., Shuman, C. A., Key, J. R., and Koenig, L. S.: A Satellite-Derived Climate-Quality Data Record of the Clear-Sky Surface Temperature of the Greenland Ice Sheet, J. Clim., 25, 4785–4798, https://doi.org/10.1175/JCLI-D-11-00365.1, 2012.

Hanna, E., Cappelen, J., Fettweis, X., Mernild, S. H., Mote, T. L., Mottram, R., Steffen, K., Ballinger, T. J., and Hall, R. J.: Greenland surface air temperature changes from 1981 to 2019 and implications for ice-sheet melt and mass-balance change, Int. J. Climatol., 41, https://doi.org/10.1002/joc.6771, 2021.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, https://doi.org/10.1029/2010RG000345, 2010.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holland, M. M. and Bitz, C. M.: Polar amplification of climate change in coupled models, Clim. Dyn., 21, 221–232, https://doi.org/10.1007/s00382-003-0332-6, 2003.

Høyer, J. L., Alerskans, E., Nielsen-Englyst, P., Thejll, P., Dybkjær, G., and Tonboe, R.: Detailed investigation of the uncertainty budget for non-recoverable IST observations and their SI traceability (ESA Tech. Rep. FRM4STS OP-70), http://www.frm4sts.org/wp-content/uploads/sites/3/2018/08/OP-70-FRM4STS_option3_report_v1-signed.pdf, 2017a.

Høyer, J. L., Lang, A. M., Tonboe, R., Eastwood, S., Wimmer, W., and Dybkjær, G.: Towards field inter-comparison experiment (FICE) for ice surface temperature (ESA Tech. Rep. FRM4STS OP-40)., http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/OFE-OP-40-TR-5-V1-Iss-1-Ver-1-Signed.pdf, 2017b.

Høyer, J. L., Good, E., Nielsen-Englyst, P., Madsen, K. S., Woolway, I., and Kennedy, J.: Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice, ice sheets, and lakes, 2018.

Høyer, J. L., Dybkjær, G., Eastwood, S., and Madsen, K. S.: EUSTACE/AASTI: Global clear-sky ice surface temperature data from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000-2009, Centre for Environmental Data Analysis, 2019.

Hudson, S. R. and Brandt, R. E.: A Look at the Surface-Based Temperature Inversion on the Antarctic Plateau, J. Clim., 18, 1673–1696, https://doi.org/10.1175/JCLI3360.1, 2005.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 1535 pp., https://doi.org/10.1017/CBO9781107415324, 2013.

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., and Jaagus, J.: Validation of atmospheric reanalyses over the central Arctic Ocean, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL051591, 2012.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010:, J. Geophys. Res. Atmospheres, 117, https://doi.org/10.1029/2011JD017139, 2012.

Karlsson, K.-G. and Dybbroe, A.: Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations, Atmospheric Chem. Phys., 10, 1789–1807, https://doi.org/10.5194/acp-10-1789-2010, 2010.

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J. F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., and Wolters, E.: CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmospheric Chem. Phys., 13, 5351–5367, https://doi.org/10.5194/acp-13-5351-2013, 2013.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmospheric Chem. Phys., 17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017, 2017.

Kennedy, J. J., Capponi, F., Ghent, D., Good, E. J., Høyer, J. L., Kent, E. C., Madsen, K. S., Mitchelson, J. R., Nielsen-Englyst, P., and Tonboe, R. T.: EUSTACE: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995-2016, https://doi.org/10.5285/f883e197594f4fbaae6edebafb3fddb3, 2019.

Key, J. R., Collins, J. B., Fowler, C., and Stone, R. S.: High-latitude surface temperature estimates from thermal satellite data, Remote Sens. Environ., 61, 302–309, https://doi.org/10.1016/S0034-4257(97)89497-7, 1997.

Greenland Climate Network (GC-Net) Radiation for Arctic System Reanalysis, Version 1.:

Knust, R.: Polar Research and Supply Vessel POLARSTERN operated by the Alfred-Wegener-Institute, J. Large-Scale Res. Facil. JLSRF, 3, https://doi.org/10.17815/jlsrf-3-163, 2017.

Koenig, L. S. and Hall, D. K.: Comparison of satellite, thermochron and air temperatures at Summit, Greenland, during the winter of 2008/09, J. Glaciol., 56, 735–741, https://doi.org/10.3189/002214310793146269, 2010.

König-Langlo, G., Loose, B., and Bräuer, B.: 25 Years of Polarstern Meteorology, WDC-MARE Rep., 4(CD-ROM), 1–137, 2006.

Langen, P. L., Mottram, R. H., Christensen, J. H., Boberg, F., Rodehacke, C. B., Stendel, M., van As, D., Ahlstrøm, A. P., Mortensen, J., Rysgaard, S., Petersen, D., Svendsen, K. H., Aðalgeirsdóttir, G., and Cappelen, J.: Quantifying Energy and Mass Fluxes Controlling Godthåbsfjord Freshwater Input in a 5-km Simulation (1991–2012)., J. Clim., 28, 3694–3713, https://doi.org/10.1175/JCLI-D-14-00271.1, 2015.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP Uncertainty Model, J. Geophys. Res. Atmospheres, 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Lindsay, R., Wensnahan, M., Schweiger, A., and Zhang, J.: Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic, J. Clim., 27, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1, 2014.

Lüpkes, C., Vihma, T., Jakobson, E., König-Langlo, G., and Tetzlaff, A.: Meteorological observations from ship cruises during summer to the central Arctic: A comparison with reanalysis data, Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL042724, 2010.

Masson-Delmotte, V., Swingedouw, D., Landais, A., Seidenkrantz, M.-S., Gauthier, E., Bichet, V., Massa, C., Perren, B., Jomelli, V., Adalgeirsdottir, G., Hesselbjerg Christensen, J., Arneborg, J., Bhatt, U., Walker, D. A., Elberling, B., Gillet-Chaulet, F., Ritz, C., Gallée, H., van den Broeke, M., Fettweis, X., de Vernal, A., and Vinther, B.: Greenland climate change: from the past to the future: Greenland climate change, Wiley Interdiscip. Rev. Clim. Change, 3, 427–449, https://doi.org/10.1002/wcc.186, 2012.

Menke, W.: Geophysical Data Analysis: Discrete Inverse Theory, Elsevier, New York, 1989.

Merchant, C. J. and Le Borgne, P.: Retrieval of Sea Surface Temperature from Space, Based on Modeling of Infrared Radiative Transfer: Capabilities and Limitations, J. Atmospheric Ocean. Technol., 21, 1734–1746, https://doi.org/10.1175/JTECH1667.1, 2004.

Merchant, C. J., Harris, A. R., Murray, M. J., and Závody, A. M.: Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison, J. Geophys. Res. Oceans, 104, 23565–23578, https://doi.org/10.1029/1999JC900105, 1999.

Merchant, C. J., Matthiesen, S., Rayner, N. A., Remedios, J. J., Jones, P. D., Olesen, F., Trewin, B., Thorne, P. W., Auchmann, R., Corlett, G. K., Guillevic, P. C., and Hulley, G. C.: The surface temperatures of Earth: steps towards integrated understanding of variability and change, Geosci. Instrum. Methods Data Syst., 2, 305–321, https://doi.org/10.5194/gi-2-305-2013, 2013.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. F., and Wagner, W.: Uncertainty information in climate data records from Earth observation, Earth Syst. Sci. Data, 9, 511–527, https://doi.org/10.5194/essd-9-511-2017, 2017.

Miller, N. B., Turner, D. D., Bennartz, R., Shupe, M. D., Kulie, M. S., Cadeddu, M. P., and Walden, V. P.: Surface-based inversions above central Greenland, J. Geophys. Res. Atmospheres, 118, 495–506, https://doi.org/10.1029/2012JD018867, 2013.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set: The HADCRUT4 data set, J. Geophys. Res. Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.

Morice, C. P., Capponi, F., Kennedy, J. J., Killick, R. E., Lindgren, F., Mitchelson, J. R., Rayner, N. A., and Winn, J. P.: EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850-2015, v1.0, Centre for Environmental Data Analysis, 2019.

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R., Dybkjær, G., and Alerskans, E.: In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic, The Cryosphere, 13, 1005–1024, https://doi.org/10.5194/tc-13-1005-2019, 2019.

Oltmanns, M., Straneo, F., Seo, H., and Moore, G. W. K.: The Role of Wave Dynamics and Small-Scale Topography for Downslope Wind Events in Southeast Greenland, J. Atmospheric Sci., 72, 2786–2805, https://doi.org/10.1175/JAS-D-14-0257.1, 2015.

Østby, T. I., Schuler, T. V., and Westermann, S.: Severe cloud contamination of MODIS Land Surface Temperatures over an Arctic ice cap, Svalbard, Remote Sens. Environ., 142, 95–102, https://doi.org/10.1016/j.rse.2013.11.005, 2014.

Overland, J., Dunlea, E., Box, J. E., Corell, R., Forsius, M., Kattsov, V., Olsen, M. S., Pawlak, J., Reiersen, L.-O., and Wang, M.: The urgency of Arctic change, Polar Sci., https://doi.org/10.1016/j.polar.2018.11.008, 2018.

Perovich, D. K., Richter-Menge, J. A., and Polashenski, C. M.: Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice, 2016.

5   Pielke, R. A., Davey, C. A., Niyogi, D., Fall, S., Steinweg-Woods, J., Hubbard, K., Lin, X., Cai, M., Lim, Y.-K., Li, H., Nielsen-Gammon, J., Gallo, K., Hale, R., Mahmood, R., Foster, S., McNider, R. T., and Blanken, P.: Unresolved issues with the assessment of multidecadal global land surface temperature trends, J. Geophys. Res., 112, D24S08, https://doi.org/10.1029/2006JD008229, 2007.

Pithan, F. and Mauritsen, T.: Arctic amplification dominated by temperature feedbacks in contemporary climate models, Nat.
10 Geosci., 7, 181–184, https://doi.org/10.1038/ngeo2071, 2014.

Rapaić, M., Brown, R., Markovic, M., and Chaumont, D.: An Evaluation of Temperature and Precipitation Surface-Based and Reanalysis Datasets for the Canadian Arctic, 1950–2010, Atmosphere-Ocean, 53, 283–303, https://doi.org/10.1080/07055900.2015.1045825, 2015.

Rasmussen, T. A. S., Høyer, J. L., Ghent, D., Bulgin, C. E., Dybkjaer, G., Ribergaard, M. H., Nielsen-Englyst, P., and Madsen,
15 K. S.: Impact of Assimilation of Sea-Ice Surface Temperatures on a Coupled Ocean and Sea-Ice Model, J. Geophys. Res. Oceans, 123, 2440–2460, https://doi.org/10.1002/2017JC013481, 2018.

Rayner, N., Good, S., and Block, T.: SST CCI Product User Guide, Project Document, SST_CCI-PUG-UKMO-201., http://www.esa-sst-cci.org/PUG/pdf/SST_CCI-PUG-UKMO-201-Issue_1-signed.pdf, 12 2015.

Rayner, N. A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth
20 century, J. Geophys. Res., 108, https://doi.org/10.1029/2002JD002670, 2003.

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M. A., Ghent, D., Good, E., Høyer, J. L., Kennedy, J. J., Kent, E. C., Killick, R. E., van der Linden, P., Lindgren, F., Madsen, K. S., Merchant, C. J., Mitchelson, J. R., Morice, C. P., Nielsen-Englyst, P., Ortiz, P. F., Remedios, J. J., van der Schrier, G., Squintu, A. A., Stephens, A., Thorne, P. W., Tonboe, R. T., Trent, T., Veal, K. L., Waterfall, A. M., Winfield, K., Winn, J., and
25 Woolway, R. I.: The EUSTACE Project: Delivering Global, Daily Information on Surface Air Temperature, Bull. Am. Meteorol. Soc., 101, E1924–E1947, https://doi.org/10.1175/BAMS-D-19-0095.1, 2020.

Reeves Eyre, J. E. J. and Zeng, X.: Evaluation of Greenland near surface air temperature datasets, The Cryosphere, 11, 1591–1605, https://doi.org/10.5194/tc-11-1591-2017, 2017.

Renfrew, I. A.: The dynamics of idealized katabatic flow over a moderate slope and ice shelf, Q. J. R. Meteorol. Soc., 130,
30 1023–1045, https://doi.org/10.1256/qj.03.24, 2004.

Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., and Ortmeyer, M.: Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover, Ann. Glaciol., 44, 205–210, https://doi.org/10.3189/172756406781811727, 2006.

Richter-Menge, J. A., Overland, J., Mathis, J. T., and Osborne, E.: Eds., Arctic Report Card 2017, 2017.

35 Screen, J. A. and Simmonds, I.: Increasing fall-winter energy loss from the Arctic Ocean and its role in Arctic temperature amplification, Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL044136, 2010.

Serreze, M. C., Box, J. E., Barry, R. G., and Walsh, J. E.: Characteristics of Arctic synoptic activity, 1952-1989, Meteorol. Atmospheric Phys., 51, 147–164, https://doi.org/10.1007/BF01030491, 1993.

Shuman, C. A., Steffen, K., Box, J. E., and Stearns, C. R.: A Dozen Years of Temperature Observations at the Summit: Central Greenland Automatic Weather Stations 1987–99, J. Appl. Meteorol., 40, 741–752, https://doi.org/10.1175/1520-0450(2001)040<0741:ADYOTO>2.0.CO;2, 2001.

Shuman, C. A., Hall, D. K., DiGirolamo, N. E., Mefford, T. K., and Schnaubelt, M. J.: Comparison of Near-Surface Air Temperatures and MODIS Ice-Surface Temperatures at Summit, Greenland (2008–13), J. Appl. Meteorol. Climatol., 53, 2171–2180, https://doi.org/10.1175/JAMC-D-14-0023.1, 2014.

Simmons, A. J. and Poli, P.: Arctic warming in ERA-Interim and other analyses: Arctic Warming in ERA-Interim and Other Analyses, Q. J. R. Meteorol. Soc., 141, 1147–1162, https://doi.org/10.1002/qj.2422, 2014.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), J. Clim., 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Stamnes, K., Ellingson, R. G., Curry, J. A., Walsh, J. E., and Zak, B. D.: Review of Science Issues, Deployment Strategy, and Status for the ARM North Slope of Alaska–Adjacent Arctic Ocean Climate Research Site, J. Clim., 12, 46–63, https://doi.org/10.1175/1520-0442-12.1.46, 1999.

Steeneveld, G.-J.: Current challenges in understanding and forecasting stable boundary layers over land and ice, Front. Environ. Sci., 2, https://doi.org/10.3389/fenvs.2014.00041, 2014.

Steffen, K.: Surface energy exchange at the equilibrium line on the Greenland ice sheet during onset of melt, Ann. Glaciol., 21, 13–18, https://doi.org/10.3189/S0260305500015536, 1995.

Steffen, K. and Box, J.: Surface climatology of the Greenland Ice Sheet: Greenland Climate Network 1995-1999, J. Geophys. Res. Atmospheres, 106, 33951–33964, https://doi.org/10.1029/2001JD900161, 2001.

Sterk, H. A. M., Steeneveld, G. J., and Holtslag, A. A. M.: The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over Arctic sea ice, J. Geophys. Res. Atmospheres, 118, 1199–1217, https://doi.org/10.1002/jgrd.50158, 2013.

Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L., and Kern, S.: The EUMETSAT sea ice concentration climate data record, The Cryosphere, 10, 2275–2290, https://doi.org/10.5194/tc-10-2275-2016, 2016.

Vihma, T. and Pirazzini, R.: On the Factors Controlling the Snow Surface and 2-m Air Temperatures Over the Arctic Sea Ice in Winter, Bound.-Layer Meteorol., 117, 73–90, https://doi.org/10.1007/s10546-004-5938-7, 2005.

Vihma, T., Uotila, J., Cheng, B., and Launiainen, J.: Surface heat budget over the Weddell Sea: Buoy results and model comparisons, J. Geophys. Res., 107, 3013, https://doi.org/10.1029/2000JC000372, 2002.

Vihma, T., Jaagus, J., Jakobson, E., and Palo, T.: Meteorological conditions in the Arctic Ocean in spring and summer 2007 as recorded on the drifting ice station Tara, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008GL034681, 2008.

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J. H., Menne, M. J., Peterson, T. C., Reynolds, R. W., Smith, T. M., Williams, C. N., and Wuertz, D. B.: NOAA's Merged Land–Ocean Surface Temperature Analysis, Bull. Am. Meteorol. Soc., 93, 1677–1685, https://doi.org/10.1175/BAMS-D-11-00241.1, 2012.

Wang, C., Graham, R. M., Wang, K., Gerland, S., and Granskog, M. A.: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019, 2019.

Weng, W. and Taylor, P. A.: On Modelling the One-Dimensional Atmospheric Boundary Layer, Bound.-Layer Meteorol., 107, 371–400, https://doi.org/10.1023/A:1022126511654, 2003.

Wesslén, C., Tjernström, M., Bromwich, D. H., de Boer, G., Ekman, A. M. L., Bai, L.-S., and Wang, S.-H.: The Arctic summer atmosphere: an evaluation of reanalyses using ASCOS data, Atmospheric Chem. Phys., 14, 2605–2624, https://doi.org/10.5194/acp-14-2605-2014, 2014.

Westermann, S., Langer, M., and Boike, J.: Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway, Remote Sens. Environ., 118, 162–167, https://doi.org/10.1016/j.rse.2011.10.025, 2012.

World Meteorological Organization: World Meteorological Organization (2014) Guide to Meteorological Instruments and Methods of Observation, 1128pp. (WMO-No.8). [preprint], http://hdl.handle.net/11329/365, 2014.

Zhang, W., Wang, Y., Smeets, P. C. J. P., Reijmer, C. H., Huai, B., Wang, J., and Sun, W.: Estimating near-surface climatology of multi-reanalyses over the Greenland Ice Sheet, Atmospheric Res., 259, 105676, https://doi.org/10.1016/j.atmosres.2021.105676, 2021.

Zilitinkevich, S., Savijärvi, H., Baklanov, A., Grisogono, B., and Myrberg, K.: Forthcoming Meetings on Planetary Boundary-layer Theory, Modelling and Applications, Bound.-Layer Meteorol., 119, 591–593, https://doi.org/10.1007/s10546-006-9069-1, 2006.

Table 1: Overview of in situ observations used in this study, covering the period 2000-2009.

	No. of sites, (AWS, buoys or ships)	No. of days with observations	Surface Type	Observation Type	Temperature measurements
ACSYS	7	280	Sea ice	Buoy	T2m
ARM	2	2,846	Seasonal snow	AWS	T2m, IST_{skin}
CRREL	10	1,031	Sea ice	Buoy	T2m
DAMOCLES	25	2,160	Sea ice	Buoy	T2m
ECMWF	196	27,235	Sea ice	Buoy	T2m
FRAMZY	11	251	Sea ice	Buoy	T2m
GC-NET	15	29,133	Land ice	AWS	T2m
POLARSTERN	1	189	Sea ice	Ship	T2m
PROMICE	8	2,685	Land ice	AWS	T2m, IST_{skin}

Table 2. Validation of daily AASTI v.1 Level 3 IST (IST_{skin_L3}) against in situ IST_{skin} (IST_{skin_insitu}) and T2m observations ($T2m_{insitu}$).
N: number of matchups, Corr: correlation, Std: standard deviation, RMS: root mean square difference, d is the matchup distance and Δh is the difference in elevation (AWS − Satellite).

Station	N	IST_{skin_L3} - IST_{skin_insitu}				IST_{skin_L3} − $T2m_{insitu}$				d (km)	Δh (m)
		Corr	Bias	Std	RMS	Corr	Bias	Std	RMS		
ARM_ATQ	1235	93.8	-2.47	3.69	4.44	93.7	-3.17	3.69	4.87	10.8	-
ARM_BAR	1594	94.1	-0.73	4.30	4.36	94.6	-1.14	4.02	3.86	6.1	-
PROMICE KAN-M	422	93.9	-3.65	3.37	4.96	94.6	-4.56	3.14	5.53	7.6	15
PROMICE KAN-U	239	93.9	-1.75	3.32	3.75	94.4	-3.39	3.17	4.64	14.6	21
PROMICE KPC-U	488	97.6	-1.31	2.62	2.92	98.2	-3.20	2.27	3.92	5.1	29
PROMICE NUK-U	296	77.7	-4.09	5.00	6.45	84.7	-7.19	4.01	8.23	14.4	64
PROMICE QAS-U	407	83.9	-1.65	4.20	4.51	86.3	-3.70	3.75	5.27	6.5	197
PROMICE SCO-U	403	91.5	-4.60	4.25	6.26	93.7	-7.55	3.75	8.43	4.2	20
PROMICE TAS-U	386	67.5	-1.03	5.43	5.52	79.5	-3.61	4.39	5.68	8.4	214
PROMICE UPE-U	125	88.2	-3.13	3.88	4.97	90.0	-5.49	3.50	6.50	3.0	110
All data	5595	92.9	-2.03	4.24	4.70	93.2	-3.36	4.12	5.32	8.1	83.8

Table 3: Statistics on the relation between observed and modelled temperatures for the training data. N: number of matchups used for testing, Corr: correlation, RMS: root mean square difference. Since, the training data are used for the regression, the bias is zero and thus the standard deviation equals RMS.

		N	Corr (%)	RMS (°C)
Land ice	$\hat{I}ST_{skin}$	13792	95.7	3.51
	$\hat{I}ST_{skin}SWd$	13792	96.2	3.28
	$\hat{I}ST_{skin}WS_{ERA-I}$	13792	95.8	3.47
	$\hat{I}ST_{skin}WS_{ERA5}$	13792	95.9	3.42
	$\hat{I}ST_{skin}Lat$	13792	95.8	3.48
	$\hat{I}ST_{skin}Season$	13792	96.3	3.28
Sea ice	$\hat{I}ST_{skin}$	15035	96.0	3.32
	$\hat{I}ST_{skin}SWd$	15035	96.0	3.32
	$\hat{I}ST_{skin}WS_{ERA-I}$	15035	96.0	3.32
	$\hat{I}ST_{skin}WS_{ERA5}$	15035	96.0	3.32
	$\hat{I}ST_{skin}Lat$	15035	96.1	3.28
	$\hat{I}ST_{skin}Season$	15035	96.2	3.25

Table 4: Model regression coefficients for $\ddot{I}ST_{skin}Season$.

	Offset, α_0 (°C)	IST_{skin_L3} factor, α_1	Amplitude, A	Phase, φ
Land ice	4.20	1.06	2.26	-0.33
Sea ice	1.46	0.89	1.83	-0.75

Table 5: Uncertainty model regression coefficients

Land ice	$\beta_0 = 3.82°C$	$\beta_1 = -0.24$	$\beta_2 = -0.03$	
Sea ice	$\gamma_0 = 2.01°C$	$\gamma_1 = -0.06$	$\gamma_2 = -0.12$	$\gamma_3 = -0.001$

Table 6: Statistics on the relation between satellite-derived and in situ measured temperatures for comparison with independent validation data. N: number of matchups used for validation, Corr: correlation, bias: $T2m_{sat} - T2m_{insitu}$ difference, Std: standard deviation, RMS: root mean square difference.

	N	Corr (%)	bias (°C)	Std (°C)	RMS (°C)
Land ice	20872	95.5	0.30	3.45	3.47
Sea ice	16111	96.5	0.35	3.18	3.20

Table 7: Statistics on the relation between ERA-I/ERA5 and in situ measured temperatures for independent test data. N: number of matchups used for validation, Corr: correlation, bias: T2m$_{ERA}$ – T2m$_{insitu}$ difference, Std: standard deviation, RMS: root mean square difference.

	N		Corr (%)	Bias (°C)	Std (°C)	RMS (°C)
Land ice	20872	ERA-I	96.4	3.41	3.18	4.66
		ERA5	97.1	2.03	3.08	3.69
Sea ice	16111	ERA-I	96.9	1.14	3.02	3.22
		ERA5	95.7	2.19	3.67	4.27

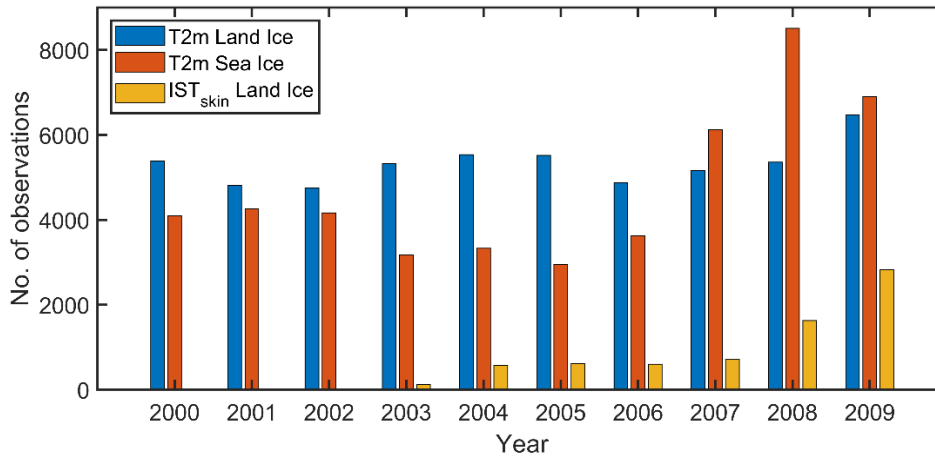



Figure 1: Total number of daily averaged in situ observations of T2m and IST$_{skin}$ over Arctic land ice and sea ice per year covering the period 2000-2009.

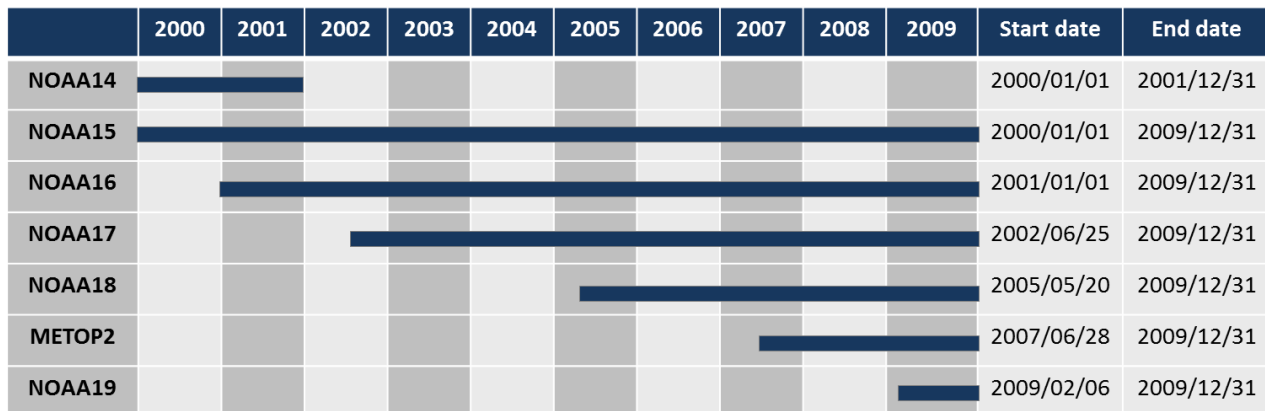



Figure 2: NOAA and Metop satellites carrying the AVHRR sensor, used for AASTI version 1.

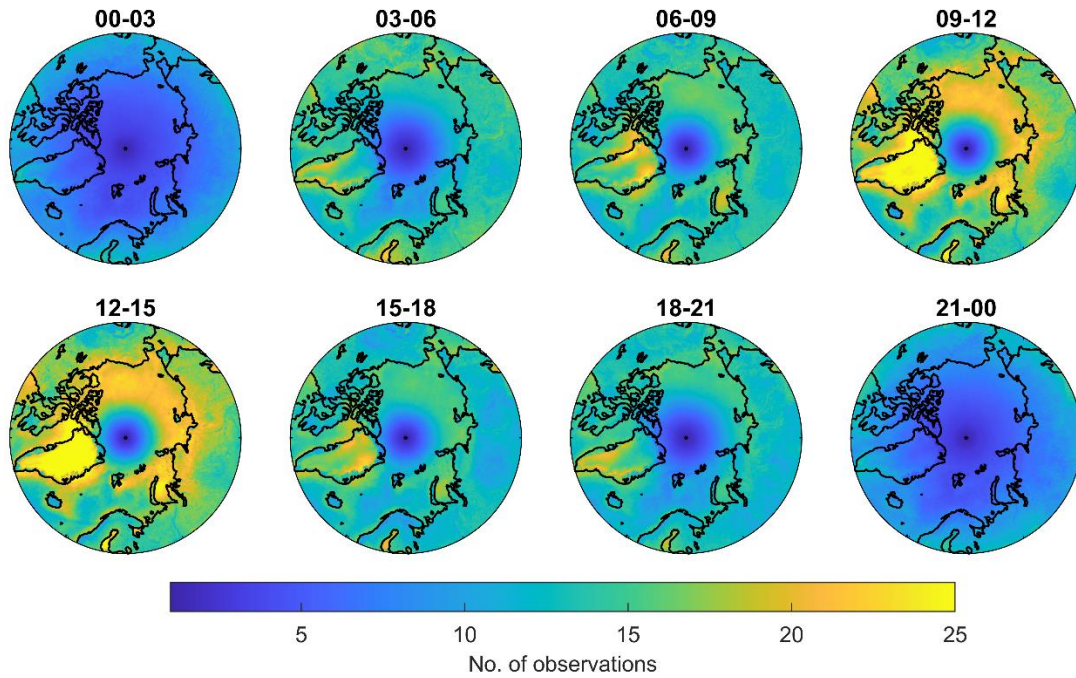



Figure 3: Mean number of observations per day in the L3 bins for each of the eight local solar time intervals, averaged for the period 2000-2009.

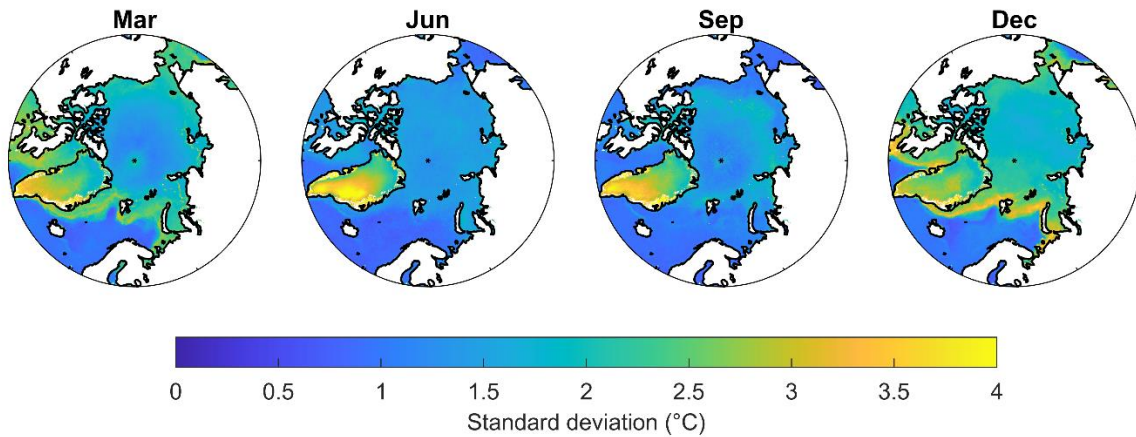



Figure 4: Standard deviations (°C) of daily satellite surface temperature observations for March, June, September and December of each year averaged for the years 2000-2009.

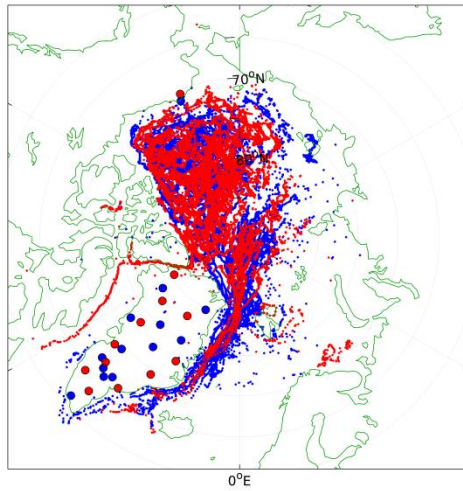



Figure 5: Positions of matchups on sea ice and land ice (red: training, blue: validation)

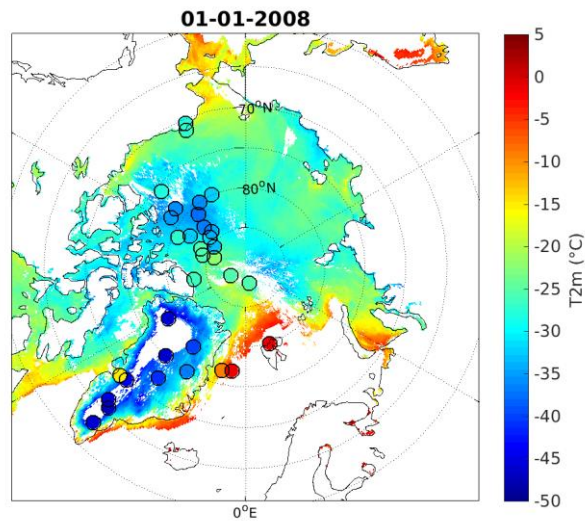



Figure 6: Daily mean 2 m air temperature over land ice and sea ice from January 1, 2008. Circles show in situ measurements.

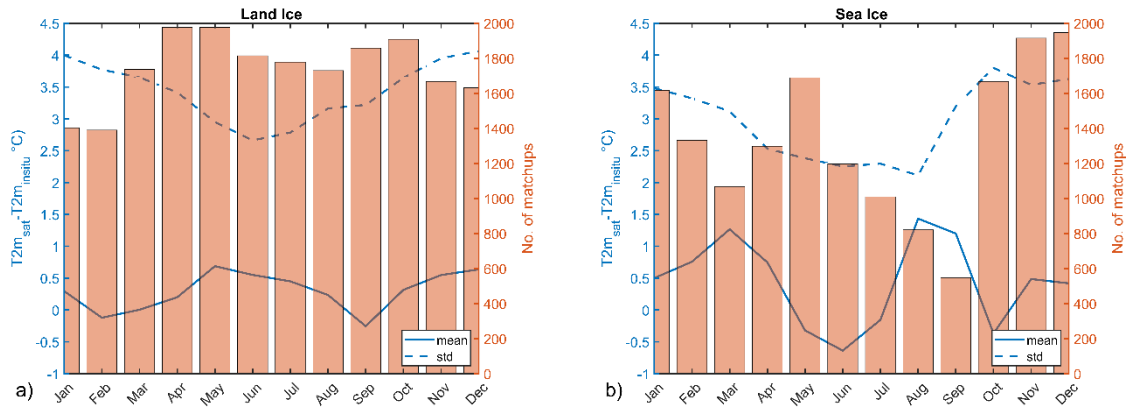
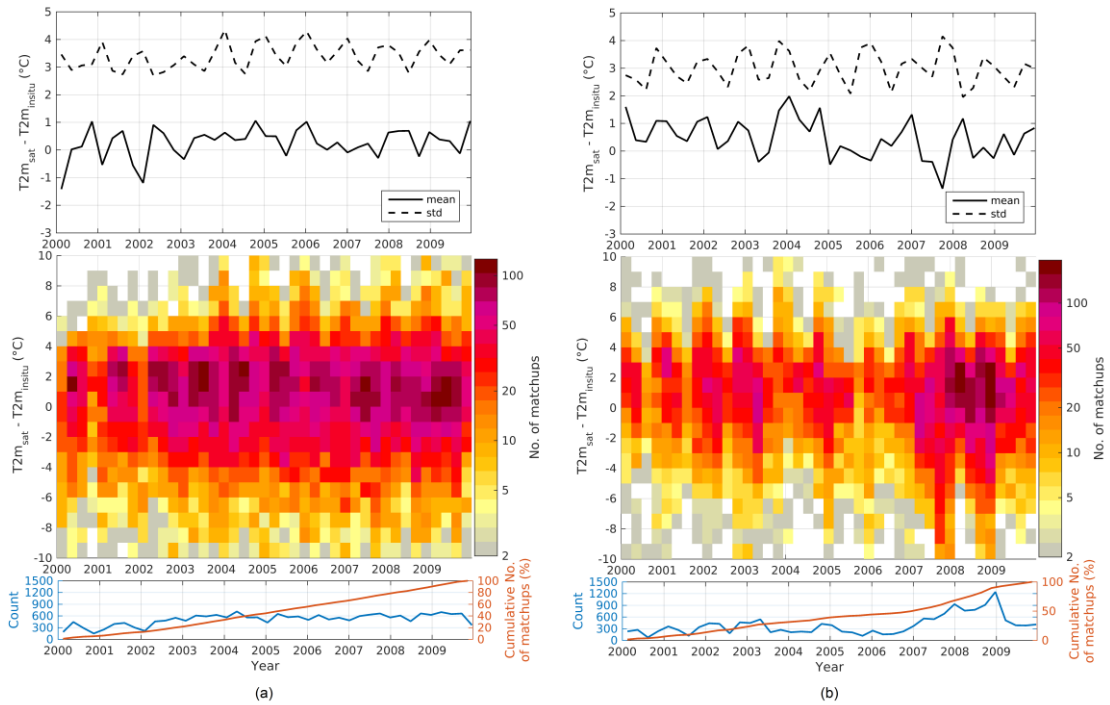



Figure 7: Estimated T2m minus observed T2m averaged for each month for (a) land ice and (b) sea ice. The dashed lines are standard deviations while the solid lines are biases. The bars show the average number of matchups for each month.

5   **Figure 8: Estimated T2m minus observed T2m (bin size of 1°C) for the full time period (bin size of 90 days) for (a) land ice and (b) sea ice. The dashed lines are standard deviations while the solid lines are bias in the upper figures. The surface plots in the middle figures show the number of matchups in each bin, while the bottom plots show the number of matchups (blue) and the cumulative percentage of matchups (red) in each time bin.**

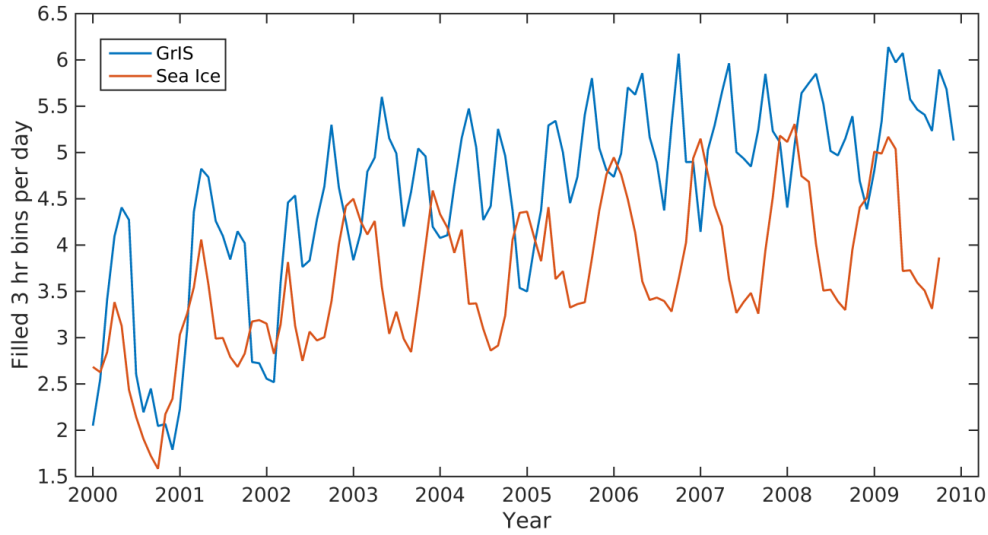



Figure 9: Average number of filled 3 h bins per day for the Greenland Ice Sheet and the Arctic sea ice.

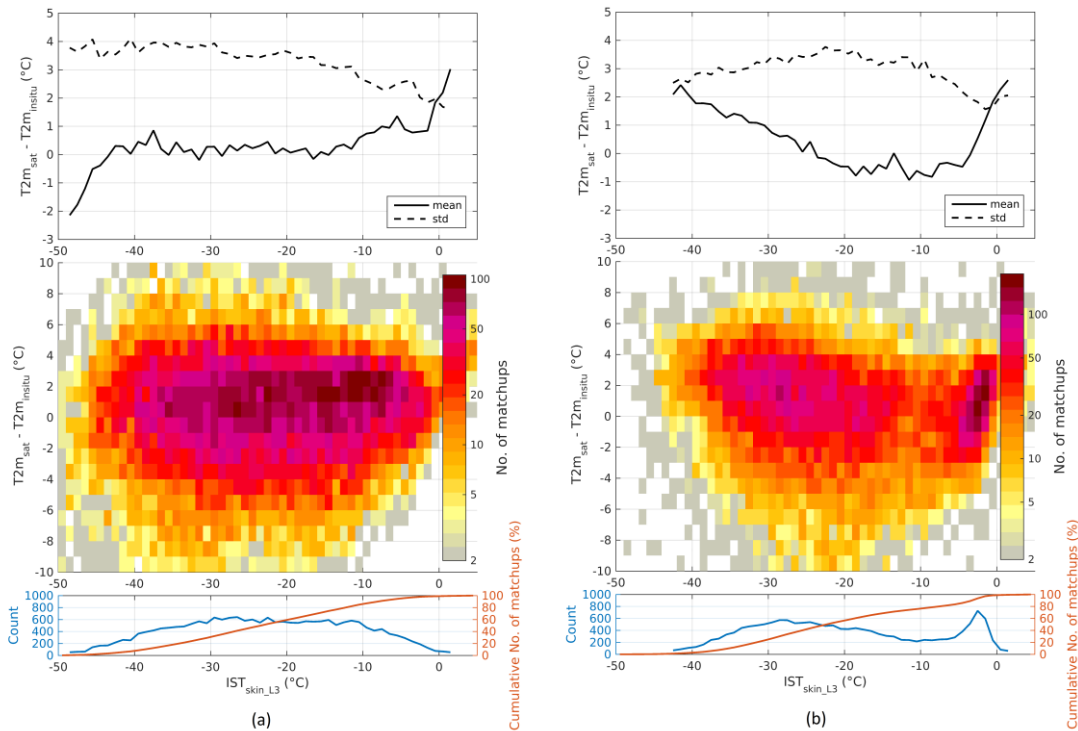



Figure 10: Estimated T2m minus observed T2m (bin size of 1°C) as a function of binned (bin size of 1°C) satellite IST$_{skin\_L3}$ for (a) land ice and (b) sea ice. The dashed lines are standard deviations while the solid lines are bias in the upper figure. The surface plots in the middle figures show the number of matchups in each bin while the bottom plots show the number of matchups (blue) and the cumulative percentage of matchups (red) in each IST$_{skin\_L3}$ bin.

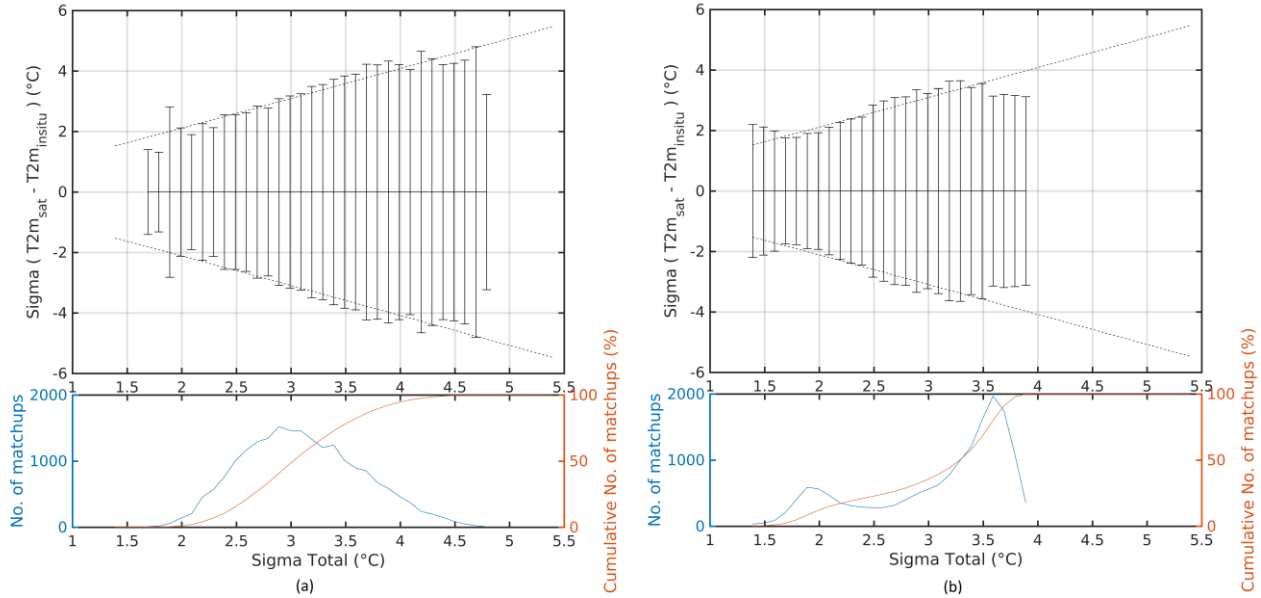



Figure 11: Satellite estimated T2m uncertainty validation with respect to independent in situ T2m for (a) land ice and (b) sea ice. Dashed lines show the modelled uncertainty accounting for uncertainties in the in situ T2m and the sampling error. Solid black lines show one standard deviation of the estimated minus in situ differences for each 0.1 °C bin. The bottom plots show the number of matchups (blue) and the cumulative percentage of matchups for each bin (red).

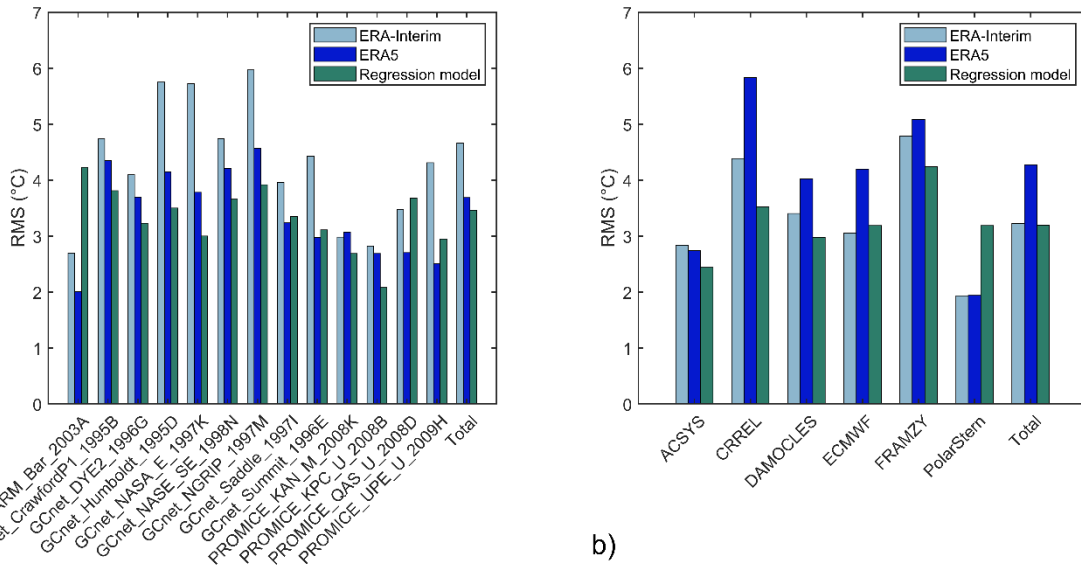



Figure 12: Root mean square (RMS) differences calculated for the (a) land ice sites and (b) sea ice sites using T2m from ERA-Interim, ERA5 and the regression model, respectively. Only buoys with more than 200 observations are included. The last two bars listed as "total" are the RMS obtained by using all validation data.

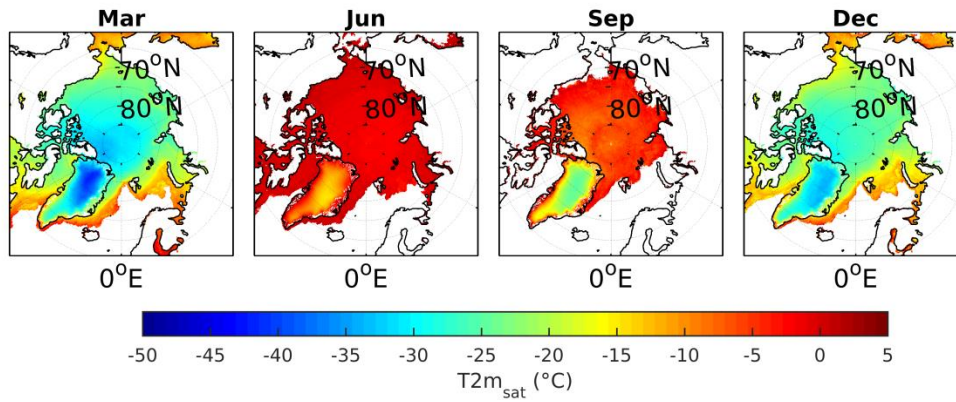



Figure 13: Monthly mean T2m$_{sat}$ during March, June, September and December, averaged for the period 2000-2009.

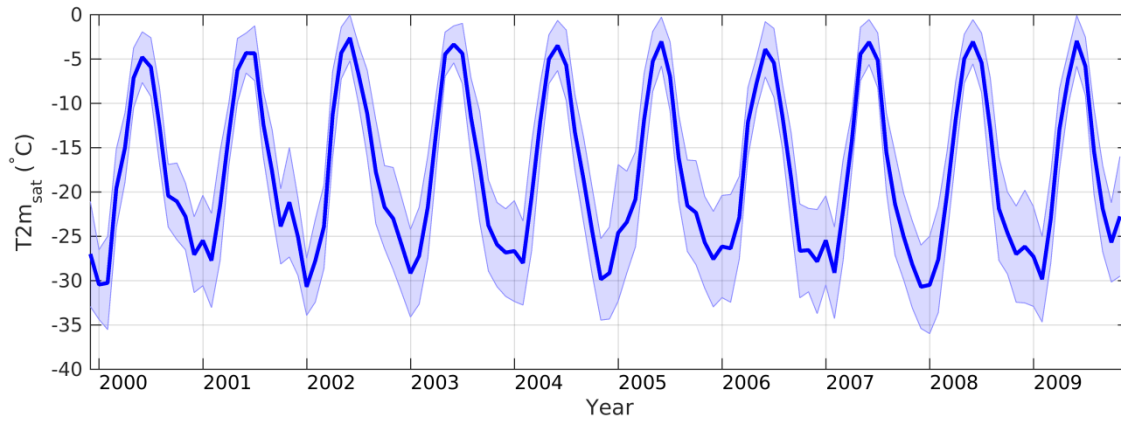



Figure 14: Monthly mean T2m$_{sat}$ for the Greenland Ice Sheet. The shading represents the variability.

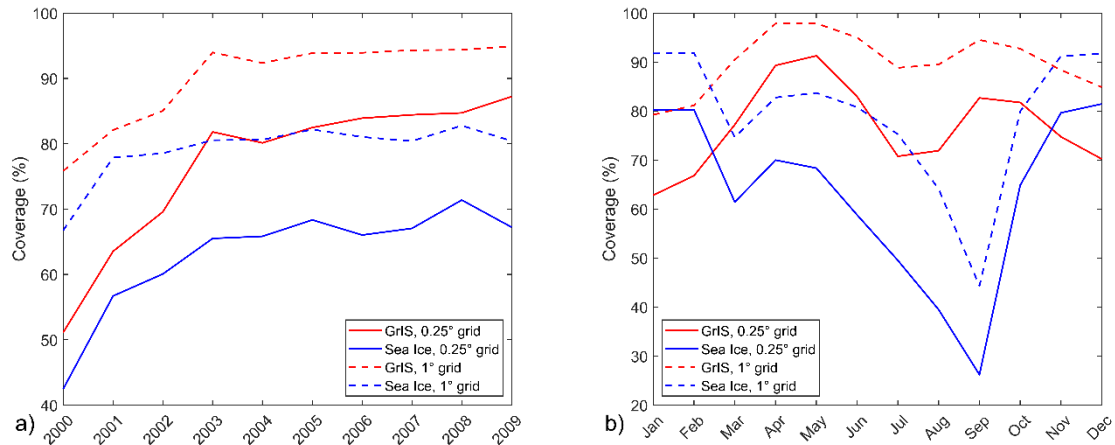



Figure 15: T2msat coverage averaged for a) each year and b) each month for land ice and sea ice, and using a grid resolution of 0.25 degree and 1 degree, respectively.