High-resolution maps of Arctic surface skin temperature and type retrieved from airborne thermal infrared imagery collected during the HALO–$(\mathcal{AC})^3$ campaign

Joshua J. Müller[1], Michael Schäfer[1], Sophie Rosenburg[1], André Ehrlich[1], and Manfred Wendisch[1]

[1]Leipzig Institute for Meteorology, University of Leipzig, Leipzig, Germany

Correspondence: Joshua J. Müller (joshua.mueller@uni-leipzig.de)

Abstract. Two retrieval methods for the determination of Arctic surface skin temperature and surface type based on radiance measurements from the thermal infrared (TIR) imager VELOX (Video airbornE Longwave Observations within siX channels) ~~are introduced. VELOX captures~~ were developed. VELOX captured TIR radiances in terms of brightness temperatures ~~in the atmospheric window~~ for wavelengths from 7.7 µm to 12 µm in six spectral channels. ~~It~~ The imager was deployed on the High Altitude and LOng Range research aircraft (HALO) during the HALO–$(\mathcal{AC})^3$ ~~airborne~~ aircraft field campaign conducted in the framework of the Arctic Amplification: Climate Relevant Atmospheric and SurfaCe Processes and Feedback Mechanisms $(\mathcal{AC})^3$ research program. The measurements were taken over the Fram Strait and the central Arctic in March and April 2022. ~~To derive the surface skin temperature, radiative transfer simulations were performed~~ Radiative transfer simulations assuming cloud-free atmospheric conditions ~~, quantifying~~ were performed showing that the influence of water ~~vapour~~ vapor on the measured brightness temperature ~~. Since this influence was negligible,~~ can be neglected. Therefore it was possible to apply a single-channel retrieval ~~of~~ technique to obtain the surface skin temperature ~~. The derived surface skin temperatures~~ from the VELOX data. The retrieval results were compared with data from the MODerate-resolution Imaging Spectroradiometer (MODIS) ~~. Furthermore~~ showing an agreement within 2.0 K. Secondly, a pixel-by-pixel surface classification retrieval was developed using a random forest algorithm. It classifies surfaces into types of open water, sea ice water mixture, thin sea ice, and snow-covered sea ice ~~was developed using a random forest algorithm. When the~~. The resulting sea-ice concentrations ~~are~~ were compared with satellite data, yielding a mean absolute difference (MAD) of 5 % ~~is obtained~~. In addition, the classified pixels were aggregated into segments of the same surface type, providing different segment size distributions for all surface types. When grouped by the distance to the sea ice edge, the segment size distribution ~~shows a shift , favoring~~ showed a shift to fewer but larger floes in the direction of the pack ice.

1 Introduction

Arctic amplification comprises Arctic-specific processes and feedback mechanisms that cause ~~the observed~~ a number of obvious changes of the Arctic climate system, such as accelerated warming of the Arctic region as compared to the rest of the globe (Wendisch et al., 2023a). Another signature of Arctic amplification involves the transition to fewer, thinner, and more dynamic sea ice within the last decades (Kwok, 2018; Meier and Stroeve, 2022; Budikova, 2009; Notz and Community, 2020).

25 Therefore, observations of the current state an the changes of the Arctic sea ice are critical. Furthermore, the Arctic sea ice serves as a thermal insulator, regulating heat and moisture exchange between ocean and atmosphere (Maykut and Untersteiner, 1971; Qu et al., 2019). To quantify these exchange processes measurements of sea-ice surface skin temperature (IST) and open ocean sea surface temperature (SST) are crucial. In-situ measurements from buoys or ship-borne instruments, are sparse in the Arctic due to harsh conditions and logistical challenges in this area (Smith et al., 2019). As a consequence, remote sensing

30 techniques are used to determine IST and SST (Hall et al., 2004; Li et al., 2022; Nielsen-Englyst et al., 2023). To retrieve these properties, established approaches use information supplied by observations in the wavelength region range of the atmospheric window (7 μm – 14 μm) , where atmospheric absorption can mostly be neglected (McMillin and Crosby, 1984; Liu et al., 2006). Specifically, wavelength bands centered around 11 μm and 12 μm are commonly used to retrieve the temperature of prevailing surface features (Hall et al., 2004). However, in the Arctic, these surface features partly represent small-scale phenomena, such

35 as leads, which are narrow openings in sea ice with spatial extents ranging from meters to kilometers. Leads may account for a significant amount of net heat energy flux fluxes in the Arctic (Qu et al., 2019; Gryschka et al., 2023). Additionally, melt ponds, which form on sea ice due to melting processes, reduce the surface albedo by up to 45 % (Tao et al., 2024), thereby affecting the solar significantly the solar atmospheric radiative energy budget (Anhaus et al., 2021; Niehaus et al., 2023). Established close to the ground (Anhaus et al., 2021; Niehaus et al., 2023). Common satellite retrievals often lack the horizontal resolution

40 needed to discriminate the majority of narrow leads and small melt-ponds. For example, the horizontal resolution of the MODerate resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites (Willmes and Heinemann, 2015) (Willmes and Heinemann, 2015, MODIS) restricts its observations to features larger than 500 m (Hall et al., 2004). The heterogeneous spatial distribution of typical Arctic surface types, e.g., open water, thin sea ice, snow-covered sea ice, melt ponds, and transitional types plays an important role in the determination of the Arctic Radiative Energy Budget (REB) (Di Biagio

45 et al., 2021; Anhaus et al., 2021; Wendisch et al., 2023b).

To quantify spatial heterogeneity, surface classification algorithms have been developed, using empirically determined thresholds (Massom and Comiso, 1994; Jäkel et al., 2019b; Thielke et al., 2023), including supervised (Wright and Polashenski, 2018) and unsupervised statistical learning approaches (Paul and Huntemann, 2021). Massom and Comiso (1994) used different wavelengths of measurements at wavelengths in the thermal infrared (TIR) data from the Advanced Very High Resolution

50 Radiometer (Cracknell, 1997, AVHRR) to discriminate classify the surface into open water, new ice, young ice, and thick ice with a snow cover with a resolution of 1.1 km at nadir. The scene classification by Paul and Huntemann (2021) uses used MODIS TIR data with wavelengths similar to those in the approach of Massom and Comiso (1994), but the classification into open water, thin sea ice, thick sea ice, and clouds is were performed by a deep neural network instead of thresholds based on a histogram. As both approaches use have used TIR data, they can be also applied at polar-night. In contrast, Jäkel et al.

55 (2019b), Thielke et al. (2023), and Wright and Polashenski (2018) rely relied on high-spatial resolution airborne data rather than satellite imagery. While Thielke et al. (2023) used a thermal imager mounted to a helicopter during polar night, both Wright and Polashenski (2018) and Jäkel et al. (2019b) derived surface type classifications with airborne based measurements in the at visible wavelengths to distinguish open-water, melt-ponds, and sea-ice. While Thielke et al. (2023) retrieved IST to distinguish sea-ice and open water with a resolution of 1 m , and Wright and Polashenski (2018) used imagery on the decimeter

60 scale. In summary, results of satellite retrievals offer wide scene and consistent time coverage, but provide data with limited horizontal resolution, while airborne data offer images of high horizontal resolution, but with limited spatial and temporal coverage.

Therefore, we have developed a skin temperature retrieval algorithm applied to the TIR imager VELOX (Video airbornE Longwave Observations within siX channels; Schäfer et al., 2022)

65 and combined with a surface type classification using supervised machine learning techniques. A random forest algorithm (Breiman, 2001; Belgiu and Drăguţ, 2016; Wright and Polashenski, 2018) was used to classify the observed surface types pixel-by-pixel into four categories: Open-Water (OW), Ice-Water Mix (IWM), Thin Ice (TI), and Snow-Covered ice (SC). To sharpen the interpretation of spatial properties of the surface types, a segmentation was applied unifying neighboring pixels of the same surface type into segments. The article is structured as follows: the measurements from

70 the HALO–$(\mathcal{AC})^3$ campaign and satellite data used in this study are introduced in Sect. 2. The single-channel surface skin temperature retrieval method and the random forest algorithm used to classify surface types are described in Sect. 3. The data are used to investigate the spatial characteristics of the classified surface types in Sect. 4.

2 Measurements and instrumentation

2.1 Airborne campaign

75 The HALO–$(\mathcal{AC})^3$ aircraft campaign was conducted from 7 March to 12 April 2022, to investigate the evolution of air mass transformation processes during warm air intrusions and cold air outbreaks in the Arctic (Wendisch et al., 2024). In total, 59 flights with multiple research aircraft were realized, among them 17 flights with research aircraft HALO, which was based in Kiruna, Sweden. In Fig. 1, the locations of measurement are depicted, together with the campaign-averaged sea-ice concentration (SIC). In addition to all HALO tracks that were flown

80 during the campaign (limited to the map extent; for a full overview see Wendisch et al. (2024)). Due to HALO's range of up to 9000 km, the measurements capture Arctic surface and atmospheric parameters on a regional scale while ensuring high spatial resolution when compared to satellite sensors like MODIS, AVHRR or the Sentinel-3 Sea and Land Surface Temperature Radiometer (SLSTR; Donlon et al., 2012) The variability of different Arctic surface types (OW, IWM, TI and SC) is highest in the region between the ice-free open-ocean and the pack ice. Here,

85 we focus on small-scale variability of the surface skin temperature resulting from the inhomogeneous distribution of Arctic surface types. Therefore, the following analysis will be restricted to the marginal sea ice zone (MIZ), which is suitable for such investigations. The MIZ is defined as the region where the campaign averaged sea-ice concentration (SIC) average was between 10 % and 90 %. In addition, we include data where the SIC exceeded 90 % for less than 10 min.

90 A set of remote sensing instruments was deployed on HALO (Ehrlich et al., 2024), of which only those relevant to the analysis are briefly introduced here. To capture two-dimensional (2D) fields of TIR radiances, the VELOX (Video airbornE Longwave Observations within
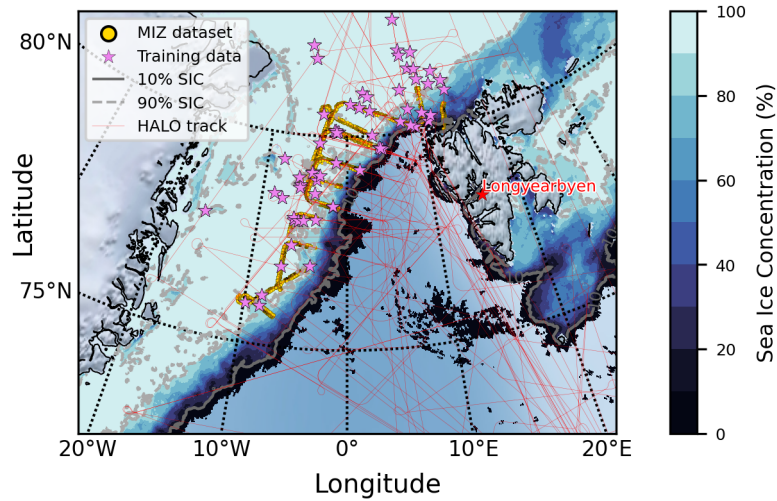
Figure 1. Overview of the data applied in this study, with the location of the data-points in orange and all flown HALO tracks in light red. The average SIC during the campaign is shown in blue contours, with a gray solid and dashed line indicating the 10 % and 90 % SIC contour, respectively. The data was provided by Spreen et al. (2008). Pink stars indicate the location of the training data used for the supervised classification.

siX channels) TIR imager was operated in a nadir-viewing configuration~~aboard HALO during the campaign~~. VELOX covers a spectral range of 7.7 to 12 µm, providing radiance measurements, which are converted to brightness temperatures (Schäfer et al., 2022). At a typical flight altitude of 10 km, the imager achieves a horizontal resolution of 10 m by 10 m per pixel, corresponding to a field of view (FOV) spanning an area of 5 km by 6 km. VELOX acquires images with a ~~high~~ temporal resolution of 100 Hz. The instrument is operated with six spectral filters resulting in six channels, of which two are redundant

Table 1. Spectral wavelength range and thermal noise uncertainty in terms of the net equivalent temperature difference (NETD) of VELOX (Video airbornE Longwave Observations within siX channels) adapted from Schäfer et al. (2022)

Channel	Wavelength range (µm)	Symbol	NETD (K)
1	7.7–12.0	$T_{B,1}$	0.048
2	8.7 ± 0.6	$T_{B,2}$	0.347
3	10.7 ± 0.4	$T_{B,3}$	0.605
4	7.7–12.0	$T_{B,4}$	0.048
5	11.7 ± 0.8	$T_{B,5}$	0.473
6	12.0 ± 0.5	$T_{B,6}$	0.442

broadband channels (channel 1 and 4). The remaining channels are narrow-band, each centered on specific wavelengths. The uncertainty in the measurements is characterized by the net equivalent temperature difference (NETD) for each channel. The

broadband channels have a NETD of 0.048 K, while the narrow-band channels show varying NETD values ~~, as~~ shown in Table 1. For ~~the airborne field campaign~~ HALO–$(\mathcal{AC})^3$, the corrected brightness temperature data, resampled to 1 s temporal resolution, ~~was~~ were provided by Schäfer et al. (2023). To retrieve cloud cover, the HALO Microwave Package (Mech et al., 2014, HAMP) and the water vapour differential absorption lidar WALES (Wirth et al., 2009) were installed on HALO. In addition, ~~more than~~ 330 dropsondes (George et al., 2024) were released during the campaign. We have restricted our analysis to cloud-free scenes in the MIZ. For this purpose, a cloud mask based on campaign-specific radar reflectivity and lidar backscatter coefficient thresholds was applied (Konow et al., 2019). To ~~further ensure high~~ ensure the data quality, each scene was visually examined to confirm the absence of clouds.

2.2 Satellite data

Independent measurements of surface skin temperature ~~are~~ were provied by MODIS sea-ice surface temperature (IST; Hall and Riggs, 2021) and sea surface temperature (SST; NASA, 2024). Both data sets ~~are~~ were based on a split-window retrieval algorithm, which determined surface skin temperature from the measured brightness temperatures. For the respective surface types, MODIS channels 1 (0.645 µm), 2 (0.865 µm), 4 (0.555 µm), 6 (1.64 µm), 31 (11 µm), 32 (12 µm) were used. The ~~datasets are provided~~ IST dataset were provided as swaths with a with a horizontal resolution of 1 km by 1 km, while the SST dataset is gridded with a with a horizontal resolution of 4 km by 4 km. Daily fields of SIC ~~are~~ were provided by the assimilated MODIS/AMSR-2 SIC product, derived from a synthesis from MODIS and AMSR-2 (Ludwig et al., 2019). Depending on the combination of MODIS and AMSR-2, the fields of SIC have a 5 km horizontal resolution for all conditions, and 1 km for cloud-free scenes. Satellite images in terms of spectral radiance with high horizontal resolution ~~are~~ were obtained from the Sentinel-2 multispectral imager (MSI, hereafter Sentinel-2) data. To characterize the surface reflectivity, the red (0.664 µm), green (0.559 µm), and blue (0.492 µm) (RGB) channels ~~are~~ were sufficient, which have a horizontal resolution of 10 m by 10 m. For the high latitudes reached by HALO-$(\mathcal{AC})^3$ observations, the revisit time of Sentinel-2 is about one day, which ~~enables~~ enabled daily observations and ~~allows~~ allowed for collocation of the satellite observations with VELOX images (Spoto et al., 2012). The Sentinel-2 data ~~are~~ were accessed via the Google Earth Engine (GEE; Gorelick et al., 2017).

3 Retrieval methods

3.1 Surface skin temperature

VELOX ~~measurements provide TIR brightness temperature~~ detects spectral radiances in the TIR wavelength range that are converted to TIR brightness temperatures, T_B ~~emitted by the surface (surface skin temperature). A significant contribution to the measured TIR T_B~~, characterizing the combined emission by atmospheric components and the surface. Actually, a significant contribution results from emission by atmospheric gases, although the spectral bands are located in the atmospheric window region. ~~Under Arctic conditions atmospheric emission results in a positive bias in the measured brightness temperature compared to the surface~~ This atmospheric contribution has to be removed from the signal to derive the surface temperature, also

called surface skin temperature. To correct for the atmospheric emission between the airplane and the surface, a split-window method (SW) is commonly applied (McMillin and Crosby, 1984; Li et al., 2013). Adjusted to VELOX measurements, this approach can be formulated as follows:

$$T_S = a_{sw} + b_{sw} \cdot T_{B,5} + c_{sw} \cdot (T_{B,5} - T_{B,6}). \tag{1}$$

Here, T_S represents the surface skin temperature, $T_{B,5}$ is the brightness temperature measured with VELOX channel 5 installed on HALO in about 10 km altitude, which is least affected by water vapour absorption. The coefficients a_{sw}, b_{sw}, and c_{sw} are empirically determined with a linear regression. Thus, $T_{B,5} - T_{B,6}$, represents the brightness temperatures difference between channels 5 and 6, serving as a proxy for water vapour absorption. Vincent et al. (2008) found that this brightness temperature difference observed in the Arctic region is also sensitive to other parameters, such as atmospheric inversion height or aerosol particles. Their proposed single-channel algorithm (SCA; Vincent, 2019) ~~can be~~ has been adapted to VELOX data as follows:

$$T_S = a_{sca} + b_{sca} \cdot T_{B,5}. \tag{2}$$

We have performed radiative transfer simulations (RTS) to constrain the contribution of atmospheric absorption to the surface skin temperature for both retrieval methods. The RTS were conducted with the radiative transfer library (*libRadtran*; Emde et al., 2016). The simulations were initialized with temperature and humidity profiles from dropsondes that were released from HALO during the campaign (Wendisch et al., 2024). The surface skin temperature ~~is~~ was provided by MODIS (Hall and Riggs., 2021), whereas ozone content ~~is~~ was given by the ERA5 reanalysis data (Hersbach et al., 2020). For the molecular absorption
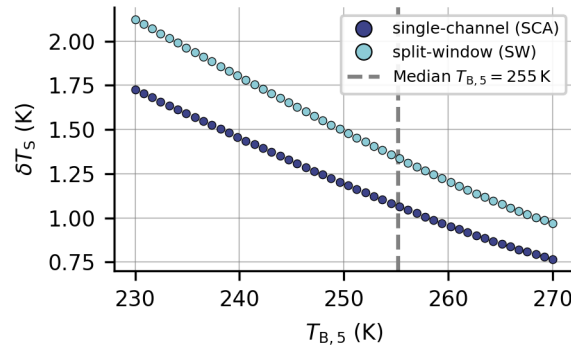
Figure 2. Empirically determined total uncertainty of the surface skin temperature retrieval δT_S as a function of VELOX measured brightness temperature $T_{B,5}$. In dark blue, the total uncertainty was calculated for the SCA, in light blue for the SW.

parameters, ~~*REPTRAN medium*~~ REPTRAN medium (Gasteiger et al., 2014) was chosen, along with the DIScrete ORdinate Radiative Transfer solvers ~~(DISORT; Stamnes et al., 2000)~~ (DISORT; Stamnes et al., 2000). To constrain the retrieval uncertainties as a function of the atmospheric total column water vapour concentration, the integrated water vapour (IWV) was varied from $0\,\mathrm{kg\,m^{-2}}$ to $50\,\mathrm{kg\,m^{-2}}$. During the HALO–$(\mathcal{AC})^3$ campaign, the integrated water ~~vapour~~ vapor (IWV) ~~can be~~ was confined to values less than $10\,\mathrm{kg\,m^{-2}}$ (Walbröl et al., 2023). To evaluate the two retrieval methods the total uncertainty δT_S ~~is~~

was calculated for both algorithms. Adapted from Brown and Minnett (1999), who formulated the uncertainties for the MODIS IST retrieval, the total uncertainty of the VELOX retrievals ~~can be~~ was formulated as follows:

$$\delta T_{\mathrm{S}}^{\mathrm{sw}} = \sqrt{(\delta T_{\mathrm{atm}}^{\mathrm{sw}})^2 + (\delta T_{\mathrm{VEL},i})^2}, \tag{3}$$

$$\delta T_{\mathrm{S}}^{\mathrm{sca}} = \sqrt{(\delta T_{\mathrm{atm}}^{\mathrm{sca}})^2 + (\delta T_{\mathrm{VEL},5})^2}, \tag{4}$$

$$\delta T_{\mathrm{vel},i} = \sqrt{(\delta T_{\mathrm{vel},i}^{\mathrm{sys}})^2 + (\delta T_{\mathrm{VEL},i}^{\mathrm{ran}})^2}, \tag{5}$$

155

where:

$$\delta T_{\mathrm{VEL},i}^{\mathrm{sys}} = A + B \cdot T_{\mathrm{B},i}, \tag{6}$$

$$\delta T_{\mathrm{VEL},i}^{\mathrm{ran}} = \mathrm{NETD}_{\mathrm{B},i}. \tag{7}$$

160 The ~~total~~ overall uncertainty of the surface skin temperature δT_{S} ~~is~~ was quantified as the square root of the sum of the squared uncertainties from the atmospheric correction, and the uncertainty introduced by VELOX $\delta T_{\mathrm{VEL},i}$, which ~~can be~~ was split into a random part $\delta T_{\mathrm{VEL},i}^{\mathrm{ran}}$ equivalent to the NETD, and a systematic part $\delta T_{\mathrm{VEL},i}^{\mathrm{sys}}$. The systematic uncertainty was parametrized based on the measured brightness temperature $T_{\mathrm{B},i}$ in channel i. The superscripts 'sca' and 'sw' correspond to the respective algorithms, while the index subscript i indicates VELOX channel i. The total uncertainty for both retrievals, depending on $T_{\mathrm{B},5}$
165 and assuming a constant δT_{atm}, is shown in Fig. 2. Due to difference in considering only the NETD of channel 5 for the SCA and both NETDs of channel 5 and 6 for the SW, the SCA retrieval has a lower error across all temperature ranges. To evaluate
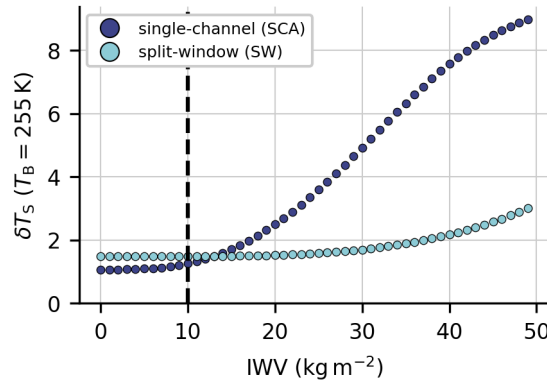
Figure 3. Comparison of surface temperature retrieval uncertainty (δT_{S}) as a function of integrated water vapor (IWV, $\mathrm{kg\,m^{-2}}$) for single-channel algorithm (dark blue) and split-window method (light blue).

the sensitivity of the retrievals to IWV, the total uncertainty of both retrievals as a function of IWV is shown in Fig. 3. Below the IWV threshold of $10\,\mathrm{kg\,m^{-2}}$, the SCA outperforms the SW, due of the reduced NETD of using only one channel. Above this threshold, atmospheric absorption dominates the total uncertainty favoring the SW algorithm. In summary, the single-channel
170 algorithm has a lower total uncertainty for IWV values below $10\,\mathrm{kg\,m^{-2}}$, while the split-window algorithm is a more suitable

for more humid atmospheres. Therefore, the SCA is applicable in the Arctic region when low IWV concentrations are present.

As a consequence, we continue with the derivation of the regression coefficients $a_{\mathrm{sc}}, b_{\mathrm{sc}}$ of the SCA algorithm. For this
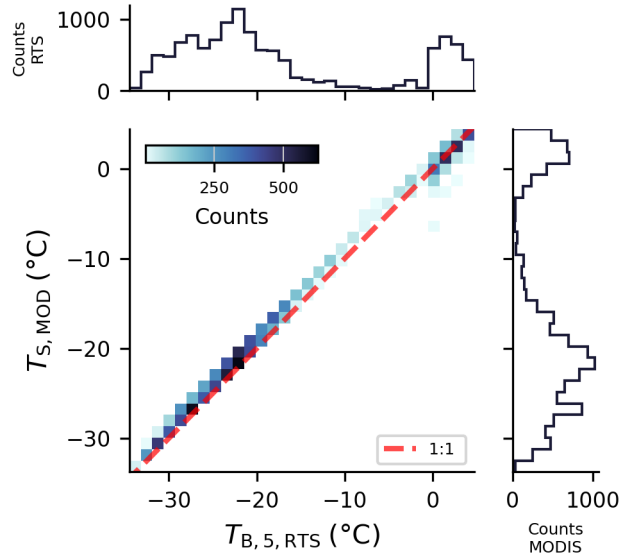
Figure 4. MODIS skin temperature $T_{\mathrm{S,MOD}}$ is compared to the simulated brightness temperature at HALO flight altitude for VELOX channel 5 (11.5 μm) $T_{\mathrm{RTS,B5}}$.

purpose, the RTS were performed with a temporal resolution of one second, resulting in simulated brightness temperature values for VELOX channel 5 $T_{\mathrm{B,5,RTS}}$ at HALO flight altitude. These simulated brightness temperature values at flight altitude are linearly regressed against MODIS surface skin temperature. The resulting fit parameters for slope and offset serve as the single-channel coefficients:

$$a_{\mathrm{sc}} = 9.051\,\mathrm{K} \tag{8}$$

$$b_{\mathrm{sc}} = 0.967\,\mathrm{K}^{-1}. \tag{9}$$

In Fig. 4, MODIS surface skin ~~temperature~~ temperatures $T_{\mathrm{S,MOD}}$ ~~is plotted against the~~ were plotted against simulated brightness temperature $T_{\mathrm{B,5,RTS}}$. The regression shows a coefficient of determination of $R^2 = 0.99$ and a root mean square error (RMSE) of 0.47 K. Substituting $\delta^{\mathrm{sc}}_{\mathrm{atm}} = 0.47\,\mathrm{K}$ and $\mathrm{NETD} = 0.473\,\mathrm{K}$ into Eq. 4, the ~~total~~ overall uncertainty of the SCA algorithm using RTS with ERA5 IWV data was computed ~~to be~~ as:

$$\delta T_{\mathrm{S}} = 1.1 \pm 0.3\,\mathrm{K}, \tag{10}$$

where the range of δT_{S} reflects the error for different measurement conditions.

3.2 Surface type classification

To distinguish different surface types, we have adapted established definitions (Miao et al., 2015; Wright and Polashenski, 2018; Jäkel et al., 2019a). As no melt-ponds were observed during HALO-$(\mathcal{AC})^3$, this surface type was omitted. To illustrate the surface types, a Sentinel-2 true color image is analyzed in Fig. 5. All surface types applied in this study are present in this scene and characterized in Tab. 2.
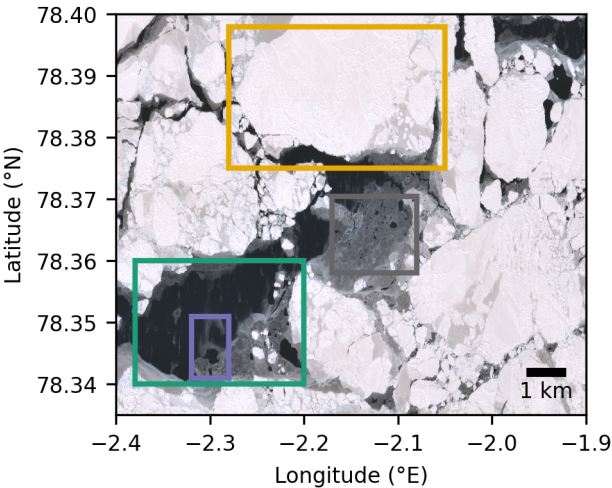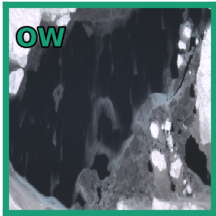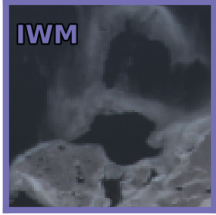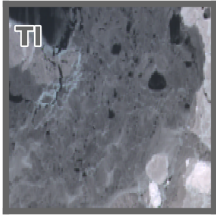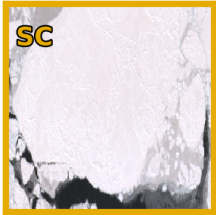
Figure 5. True color image provided by Sentinel-2 on 04.04.2022 at 13:35:00 local time. The four colored rectangles represent the surface types selected for this study: sea-ice free open-water (green), ice-water mix (purple), thin sea ice (gray), and snow-covered sea ice (yellow)

The image analysis consists of three steps. First, the VELOX 2D-images are preprocessed. Next, a random forest (RFA) classification algorithm is applied for pixel-wise surface ~~typing~~classification. Finally, a segmentation algorithm is used to identify and summarized areas of same surface type.

3.2.1 Preprocessing images

Since the temporal sampling rate of the VELOX data ~~applied here from Schäfer et al. (2023) is~~ was 1 Hz, and the typical cruise speed of HALO ~~is~~ was about 200 m s^{-1}, it ~~is~~ was possible to construct pushbroom-like images (PLI) of the corresponding nadir strips at each timestep. With this technique, the effect of the viewing ~~azimuth~~ zenith angle (VZA) on the measured brightness temperature can be neglected. Furthermore, georeferencing ~~is~~ was performed for each data point of the PLI, providing crucial information about the geographic location of the measurement. This process incorporates the geographical position, flight altitude, and attitude data from HALO, as well as the calculated viewing azimuth and zenith angles for the applied lens and detector combination of VELOX, and the measured mounting direction of VELOX.

Table 2. Surface types with short characterization, corresponding abbreviations, and representative images from Fig. 5.

Abbr.	Surface Type	Image
OW	Open Water: sea-ice free surfaces of open ocean water, including leads.	
IWM	Sea Ice-Water Mixture: unconsolidated frazil and grease ice, mixed with open ocean water.	
TI	Thin Sea Ice: freshly formed sea ice (nilas), appearing dark or grey in optical wavelengths.	
SC	Snow-Covered Sea Ice: sea-ice covered with a snow layer.	

3.2.2 Random forest classification of surface types

To determine the surface type, a random forest (RFA) was implemented in a pixel-by-pixel fashion, i.e., each pixel ~~is analyzed individually~~ was individually classified. The RFA ~~is~~ comprises a supervised machine learning method that constructs ensembles of decision trees, which are fitted to user-defined ground-truth data. It combines the interpretability of decision trees with the robustness to noise characteristics of other ensemble methods (Breiman, 2017; James et al., 2023). For the implementation of the RFA, the machine learning library *autogluon* (Erickson et al., 2020) was used, allowing for a comparison of multiple machine learning methods. Compared to other supervised learning algorithms, the RFA demonstrated comparable accuracy, while significantly reducing computation time~~, suggesting it as a preferable choice (not shown)~~.

Sentinel-2 images classified manually were used as the ground truth. For labeling these images, the Computer Vision An-

Table 3. Input parameters as processed from VELOX measurements and used in the pixel-wise RF surface type classification.

Variable	Description
$T_{B,1}$	VELOX channel 1 (7.7 μm to 12 μm).
$\Delta T_{B,2\text{-}5}$	Brightness temperature difference (BTD) between channels centered at 8.54 μm and 11.7 μm.
$\Delta T_{B,3\text{-}5}$	BTD between channels centered at 10.7 μm and 11.7 μm.
$\Delta T_{B,5\text{-}6}$	BTD between channels centered at 11.7 μm and 12 μm.
$\lvert \nabla T_{B,1} \rvert$	Magnitude of the horizontal gradient of broadband brightness temperature as a measure of horizontal inhomogeneity.
$\overline{T_{B,1}}$	Mean of $T_{B,1}$ in a 5×5 pixel neighborhood.
$\sigma T_{B,1}$	Standard deviation of $T_{B,1}$ in a 5×5 pixel neighborhood.

210 notation Tool (Sekachev et al., 2020, CVAT) was ~~used~~applied. In total, 58 VELOX images from 10 ~~different~~ research flights were labeled, resulting in ~~over~~ 13 million labeled pixels. The locations of the training data are depicted as pink stars in Fig. 1. The training data were sampled randomly from the available data (Sentinel-2 image available, cloud free) and subsequently filtered to resemble all latitudes equally. Seven input features, as defined in Tab. 3, are applied to the RFA. All parameters are calculated from VELOX brightness temperature data.

215 The accuracy of a multi-class classification problem can be expressed by the ratio of correct to all predictions. When validated in five-fold cross-validation setup, the RF showed an accuracy of 87 % with respect to the test data. To further assess the per-
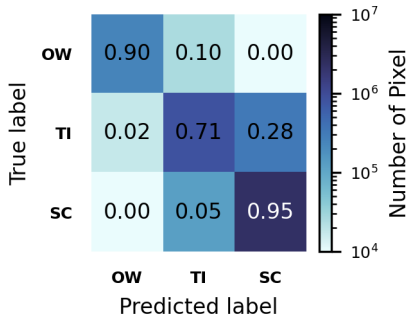
Figure 6. Confusion matrix of the RFA prediction, showing the percentage of the correctly predicted pixels on the diagonal. The off-diagonal elements represent the false positive and false negative values.

formance of the RFA, a confusion matrix is shown in Fig. 6. The highest accuracy is achieved on the SC surface type (95 %), followed by the OW surface type (90 %). The TI surface type achieves a lower overall accuracy with 71 %, due to transitional nature of this surface type. In Fig. 7 b) the initial RFA classification for an example scene is shown. A common challenge

220 when using RFA for image classification are the speckles, as seen in this figure. To address this issue, segmentation is required, which is described in the next section.

3.2.3 Segmentation

To assign the predefined surface types ~~from Sect. 3.2~~ to the retrieved fields of surface skin temperature, the PLIs are subjected to the open-source image segmentation algorithm *segment-anything* ~~(Kirillov et al., 2023, SAM;)~~ segment-anything (SAM; Kirillov et al., 202

225 . The SAM algorithm image segments on the basis of color-gradients and points that are placed by the user. The ~~initiall~~ initial segmentation of an exemplary scene is shown in Fig 7c~~)~~. Although the model was not fine-tuned, i.e., not trained with a specific user dataset, it proves a high capability to segment previously unseen data in a zero-shot fashion (Wu and Osco, 2023; Ren et al., 2024). This offers an advantage over training ~~for~~ a segmentation algorithm, which is ~~often~~ demanding in terms of data points and computational time.

230 To automatically generate a segmentation mask with SAM, a grid of points is placed on the PLI and then recursively shifted to avoid over-segmentation of the images. This means, that initially a grid is constructed on the image, and the algorithm searches for segments close to the gridpoints. To ensure stable segmentation, the grid is divided into smaller subgrids, which are then shifted relative to the initial grid points. This process is repeated three times. Since some over-segmentation still occurs, resulting in smaller predicted segments than those identified by humans, information from surface classification is added to the

235 segmentation. First, each segment identified by SAM is subjected to a majority vote, meaning the most frequently occurring surface type within a particular segment is assigned to that segment. Finally, the segments are obtained by merging neighboring segments of the same surface class. This results in a natural image segmentation, which is illustrated in the lowest panel of Fig. 7 e). ~~To further ensure the quality of Open-Water classified pixels , a threshold is applied sorting all pixels cooler than -2.5~~ This merging step can connect large, contiguous areas of a single surface type (e.g., thin ice), which can influence feature

240 size statistics. Therefore, the initial, finer-grained segmentation from SAM (prior to merging) is retained for a sensitivity analysis (see Sect. 4.3). To allow for full transparency and further exploration by the community, both the initial and final merged segmentation masks were provided in our public dataset (Müller et al., 2025). In a final post-processing step, the IWM class is identified to ensure the physical consistency of the final product. This step addresses instances where pixels were classified as OW despite having temperatures well below the physical freezing point of seawater. Specifically, any

245 OW pixel with a surface skin temperature cooler than -3 ~~°C into the Ice-Water Mix Class~~ °C were reclassified as IWM. This threshold was chosen to represent a sub-pixel sea-ice fraction of greater than 33 %, assuming the sub-pixel surface skin temperature of sea-ice to be -15 °C and the surface skin temperature of OW close to the freezing point of -1.7 °C (Skogseth et al., 2009; De La Rosa et al., 2011).
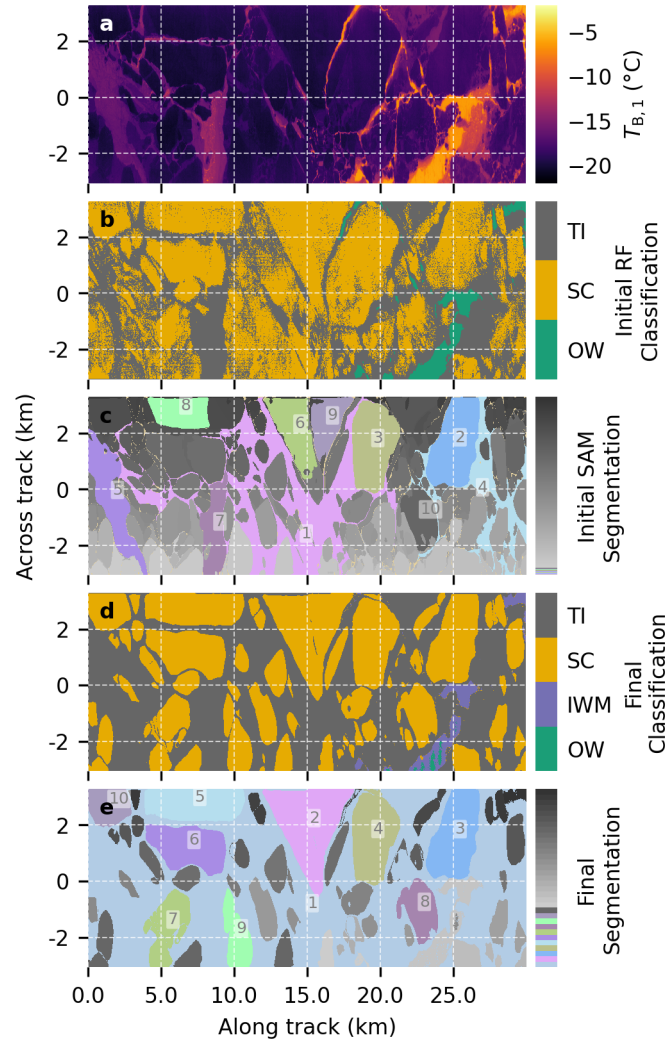
Figure 7. Overview of surface classification and segmentation results for the pushbroom-like image captured on 04. April 2022 from 13:36:14 to 13:38:31 UTC. a) Broadband brightness temperature $T_{B,1}$ (7.7 µm–12 µm) as pushbroom-like image b) Initial surface type classification using the random forest algorithm (RFA), identifying open-water, thin ice, and snow-covered ice. c) Initial segmentation using the *segment-anything* model (SAM), with numbered segments representing the ten largest areas for illustration. d) Final surface type classification: the most common surface type within each segment from (c) was assigned, and a surface skin temperature threshold was used to sort the ice-water mix class (IWM) form OW. (e) Final segmentation, where new segments were assigned to all connected regions of the same surface type derived from (d), with the largest segments again highlighted by their respective numbers.

4 Results and Discussion

4.1 Surface skin temperature~~analysis~~

In Fig. 8, the VELOX retrieved surface skin temperature $T_{S,VELOX}$ is compared to the MODIS surface skin temperature $T_{S,MOD}$, obtained from Hall and Riggs. (2021) and NASA (2024), showing the coefficient of determination R^2 to be equal to 0.96.
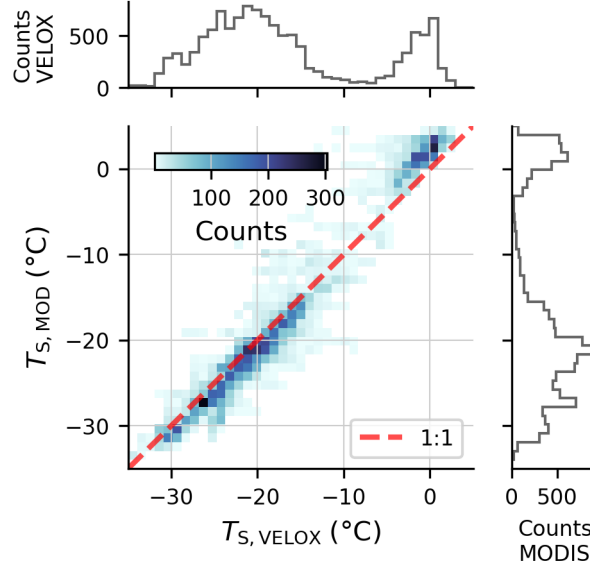
Figure 8. Scatter plot of MODIS $T_{S,MOD}$ against VELOX retrieved $T_{S,VEL}$, with VELOX data averaged to match the MODIS pixel size. The appended frequency distributions show the corresponding surface skin temperature distributions for both datasets.

For this comparison, the instantaneous FOVs of the single VELOX pixels were combined by averaging to fit the MODIS pixel-size, allowing for a direct comparison between their two datasets. The RMSE was determined to be ~~2.4~~2.0 K with a bias of ~~-2.4~~0.51 K ~~, indicating a underestimation~~ and a mean average difference of 1.55 K. Furthermore, Fig. 8 indicates slightly higher values of surface skin temperature ~~by VELOX, $T_{S,VEL}$~~derived from VELOX, $T_{S,VELOX}$, with respect to MODIS, $T_{S,MOD}$ over sea-ice and an lower values over open-water. As the data set comprises multiple days, it is essential to provide information on the location of the data. To simplify this spatial information into a scalar, the data are grouped by its distance to the sea-ice edge (positive direction into the internal ice-zone). As the individual pixels have been georeferenced, their relative distance to the nearest sea ice edge (defined by campaign-averaged SIC values between 9-11 %) is computed. For this, the distance of each spatial segment center to the temporally closest available AMSR-2/MODIS SIC pixel is calculated. The mean surface skin temperature colored by surface types is plotted in Fig. 9a against the distance to the sea-ice edge. In Fig. 9b, the mean surface skin temperature of all segments $\overline{T_S}$, weighted by their size is shown. A clear separation between the $\overline{T_S}$ of
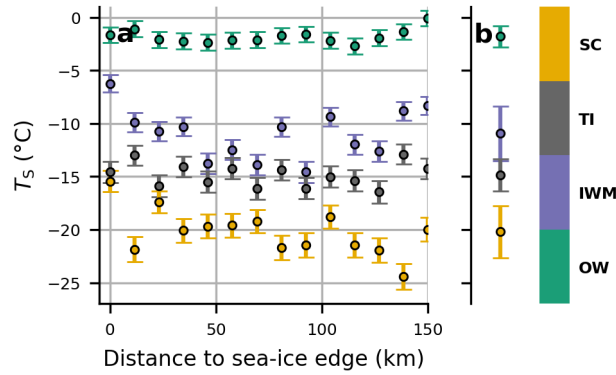
Figure 9. a) Mean T_S of different surface type segments, weighted by segment size and aggregated over 10 km bins from the sea-ice edge into the internal ice zone. b) Mean T_S over all Marginal sea-ice zone segments.

different surface types is observed as expected. The values from Fig. 9b are displayed together with the corresponding error range in Tab. 4.

4.2 Spatial analysis of surface types

From the segmentation, we retrieve the corresponding segment size, mean temperature, and standard deviation of each segment. The results are illustrated in Fig. 10, showing the spatial distance of each segment center to the nearest sea-ice edge plotted against the corresponding surface type. The fraction of the open water surface type-fraction decreases from 40 % to below 5 % in the first 20 km, while the fraction of the snow-covered surface type-fraction becomes increasingly dominant when approaching the pack-ice. The thin-ice and ice-water-mix surface types maintain relatively constant fractions of occurrence across all considered distances from the ice edge, with no clear trend observed. When compared with the provided SIC from MODIS/AMSR-2, the computed RMSE and mean absolute difference (MAD) are 8 % and 5 %, respectively. For this comparison, the nearest available SIC data from the satellite product were matched with a similar FOV of the VELOX PLI. The errors result from the temporal mismatch between both data sets, as MODIS/AMSR-2 SIC is only available as daily gridded product. When comparing the bias, i.e., the difference between VELOX SIC and MODIS/AMSR-2 SIC, an underestimation of 3 % or an overestimation of 5 % is observed, depending on whether only open-water classified pixels are considered as open water or if both open-water and ice-water-mix classified pixels are included. As shown in Fig. 10, the open-water fraction and the area fractions derived from VELOX agree within the given error range.

4.3 Segment size distribution

Since the segmentation enables the measurement of individual segment sizes, an analysis of the spatial structure of the data is performed. Here, we extend the concept of the floe size distribution (FSD; Rothrock and Thorndike, 1984; Herman, 2010;
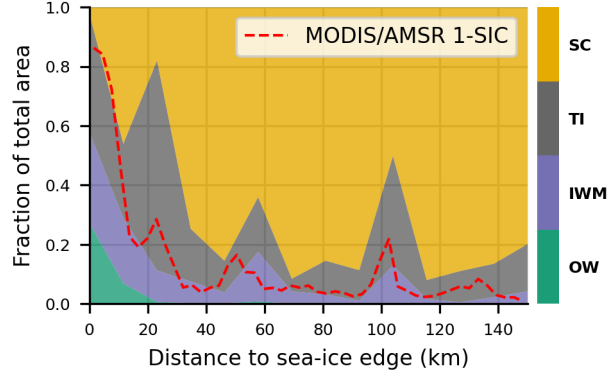
Figure 10. Fraction of total area for the four surface classes as a function of the distance to the closest sea-ice edge. The red dashed line indicates the open water fraction (1.0 - SIC) from MODIS/AMSR-2 (Ludwig et al., 2019).

Table 4. Summary of mean surface temperature $\overline{T_S}$, power-law exponents β, and goodness of fit R^2 for the corresponding N_{SSD}, for different surface types.

	$\overline{T_S}$ (C)	β	R^2
OW	-3.2 ± 1.1	-1.68 ± 0.04	0.987
IWM	-12.1 ± 2.3	-1.60 ± 0.02	0.992
TI	-17.0 ± 1.4	-1.25 ± 0.02	0.996
SC	-22.2 ± 2.0	-1.50 ± 0.02	0.992

Bateson et al., 2022) to the segment size distribution N_{SSD}, resulting in the following description:

$$N_{\text{SSD}}(x_{\text{SEG}}) = C \cdot (x_{\text{SEG}})^{\beta}. \tag{11}$$

Here, x_{SEG} represents the segment size in units of m^2, C is an empirical constant, and β is the dimensionless power-law exponent describing the scaling of the distribution. The closer the exponent is to zero, the more N_{SSD} favors large segments. This approach simplifies the complex spatial heterogeneity of the MIZ by expressing the scaling of N_{SSD} using β, a single scalar value In Fig. 11, the segment size density N_{SSD} (in units of km^{-2}) for different surface types is displayed in a double-logarithmic graph as a function of the segment size size, x_{SEG}. In addition, the individual distributions are fitted with a linear model. The slope of each linear fit corresponds to the exponent of the power-law distribution, β. The different β_i computed for each surface type are shown in Tab 4. We conclude that, in addition to the different sea-ice types, the ice-water-mix and open-water surface types also follow a power-law distribution. The computed β for, e.g. the snow-covered sea-ice type are in the range of corresponding literature data, with values ranging from -0.91 to -2.9 (Herman, 2010). A key characteristic observed during our flights over the MIZ was the presence of large, contiguous areas of thin ice or snow-covered ice, which can appear as very large segments in our final classification (e.g., Fig. 7e). We interpret these as genuine physical features of
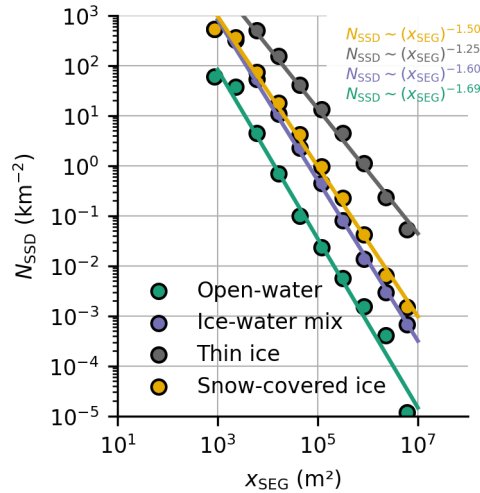
Figure 11. Double-logarithmic graph of the segment size density (colored dots) for all four surface types as a function of the segment size. The surface types are color-coded, indicating open water (green), ice-water-mix (purple), thin ice (grey), and snow-covered ice (yellow). Linear fits (colored lines) are added in the respective surface type color, providing the exponents listed at the top right.

newly forming ice in the MIZ at the time of observation rather than as segmentation artifacts. However, we acknowledge that the definition of a "segment" is sensitive to the processing methodology and that the scale of these large features can influence the resulting feature size distribution (FSD). To quantify this sensitivity, we performed an additional analysis by calculating the FSD statistics on the initial, finer-grained segmentation generated by SAM, before our final step of merging adjacent segments of the same class (an example of this initial stage is shown in Fig. 7c). The results, summarized in Tab 5, demonstrate that the power-law coefficients change by 4 to 8%, depending on the surface type.

Table 5. Power-law coefficients for the four surface types, with "Connected segments" denoting the final segmentation where same surface type segments are merged and "Broken segments" for the finer SAM segmentation.

Surface type	OW	IWM	TI	SC
Connected segments	-1.69	-1.60	-1.25	-1.50
Broken segments	-1.84	-1.69	-1.32	-1.56
Relative difference (%)	8.9	5.6	5.6	4.0

To gain more insight into the spatial heterogeneity within the MIZ, we fit the N_{SSD} of the snow-covered segments to 10 km sized bins of distance to the sea-ice edge (in pack-ice direction). In Fig. 12, the size power-law exponent β is shown as a function of the distance to the sea-ice edge for the different surface types. A linear trend is fitted only to the TI data, suggesting significance with a $R^2 = 0.76$ and a p-value less than 0.001. The increase in β from -1.6 to -1.3 reflects a physical characteristic of the MIZ. Closer to the sea-ice edge, a higher number of smaller segments is observed (more negative β) due
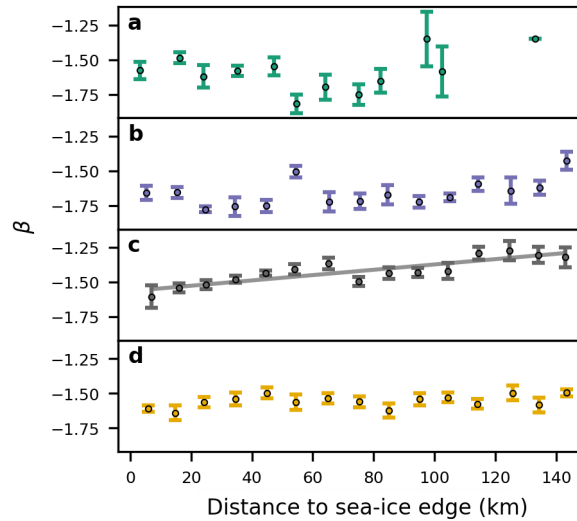
Figure 12. Power-law exponent β of the segment size distribution N_{SSD}, binned in 10 km steps, starting from the sea-ice edge into the direction of the internal ice zone by surface type. For panel a) open-water, b) ice-water mix, and d) snow-covered ice, no trend is observed. For panel c) thin ice, a significant linear trend is observed.

to intensified floe breakup, whereas larger floes (less negative β) become more prevalent further into the MIZ, where ocean wave propagation is more attenuated (Herman, 2010; Denton and Timmermans, 2022).

5   Summary and conclusions

310   During the HALO–$(\mathcal{AC})^3$ ~~airborne~~ aircraft field campaign, ~~conducted in the area from~~ covering the Fram Strait to the North Pole in March and April 2022, an extensive dataset of surface and atmospheric properties was ~~collected from~~ measured by a variety of instruments mounted on three research aircraft (Wendisch et al., 2024). Here we use data compiled by the High Altitude and LOng range research aircraft (HALO), which was instrumented with radar, lidar, a dropsonde launching facility, microwave radiometer, and various spectral imagers (Ehrlich et al., 2024). Here we have used data collected by the VELOX

315   (Video airbornE Longwave Observations within siX channels; Schäfer et al., 2022) thermal infrared (TIR) imaging system, which was installed On HALO in a nadir viewing direction. Due to its fast-spinning filter-wheel (100 Hz) equipped with multiple spectral band-pass and long-pass filters, a high spatial resolution of 10 m by 10 m pixel size for a target at 10 km distance is achieved with VELOX providing valuable high-resolution TIR spectral radiances expressed in brightness temperatures. Using VELOX data from HALO–$(\mathcal{AC})^3$, which are publicly available from Schäfer et al. (2023), a single-channel (SCA) sur-

320   face skin temperature retrieval based on linear coefficients derived from radiative transfer simulations (RTS) ~~is~~ was adapted. Comparisons with multiple-channel retrievals and surface skin temperature products from the MODerate resolution Imaging Spectroradiometer (MODIS; Hall et al., 2004; Hall and Riggs., 2021; NASA, 2024), ~~provide convincing agreement~~provided

agreements in the range of 2 K, with a coefficient of determination of $R^2 = 0.96$ and a bias of ~~-2.4~~0.5 K. To categorize the obtained surface skin temperature fields, a surface type classification algorithm ~~is~~ was developed based on publicly available

325 software tools combined with physically reasonable thresholds applied to regenerated push-broom images from the initial brightness temperature data. The resulting two-dimensional fields provide segment-vise information of the surface type, which can then be analyzed in combination with , e.g., the retrieved surface skin temperature. The data are classified into Open-Water (OW), Thin-Ice (TI), Ice-Water-Mix (IWM), and Snow-Covered-Ice (SC) surface type.

With the surface skin temperature and surface classification retrievals, important parameters were obtained with high spatial

330 resolution~~are obtained~~. When computing the resulting Sea-Ice Concentration (SIC) from the surface classification, ~~a reasonable agreement (bias ≈ 5~~an agreement with a bias of about 5 % ~~)~~ with the MODIS/AMSR-2 product ~~is achieved~~was obtained, if the IWM surface type is assigned to be "sea-ice-free". Additional sensitivity studies will be required to assess the influence of this surface type. The established classification serves as a promising foundation for these future investigations. The retrieved power-law segment size statistics are generally consistent with values reported in the literature (Denton and Timmermans,

335 2022). For the snow-covered surface type, these findings align with those of Herman (2010), who observed power-law exponents ranging from -0.91 to -2.9 and reported an increase in the exponent when transitioning from the sea-ice edge to the interior ice zone. Overall, although the temporal duration and spatial extent of the presented data set is limited, the agreement with other studies emphasizes its value for the sea-ice community. ~~Future analysis will focus on leads in the inner Arctic, using the presented methods.~~ Extending this analysis to different seasons will be crucial for capturing processes like melt pond

340 evolution, though this will require multi-sensor data fusion to resolve the increased complexity of surface types. Therefore, the primary value of this high-resolution methodology lies in providing "ground truth" for calibrating satellite retrievals and refining sub-grid-scale parameterizations in pan-Arctic models.

Code availability. The code for reproducing the analysis of this work can be accessed at https://github.com/FranzFlink/Mueller_et_al_2024_code.

345 *Data availability.* The 2D VELOX brightness temperature data can be accessed at https://doi.org/10.1594/PANGAEA.963401. The surface skin temperature, surface type and segment data can be accessed at https://doi.pangaea.de/10.1594/PANGAEA.974454. We thank the Institute of Environmental Physics, University of Bremen for the provision of the merged MODIS-AMSR2 sea-ice concentration data at https://data.seaice.uni-bremen.de/modis_amsr2 (17.12.2024).

Author contributions. JM, MS, SR, AE, and MW contributed to the conception and design of the study. JM elaborated the methods, per-

350 formed the analyses, created the figures, and prepared the original draft. All authors discussed the results, contributed to manuscript revision and approved the final submitted version.

Competing interests. Some authors are members of the editorial board of AMT.

Acknowledgements. We gratefully acknowledge the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 268020496 – TRR 172, within the framework of the Transregional Collaborative Research Center "ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms $(\mathcal{AC})^3$". The authors are thankful to AWI for providing and operating the two Polar 5 and Polar 6 aircraft. We thank the crews and the technicians of the three research aircraft for excellent technical and logistical support. The generous funding of the flight hours for the Polar 5 and Polar 6 aircraft by AWI, and for HALO by DFG, Max-Planck-Institut für Meteorologie (MPI-M), and Deutsches Zentrum für Luft- und Raumfahrt (DLR) is greatly appreciated. We are further grateful for funding of project grant number 316646266 by DFG within the framework of Priority Program SPP 1294 to promote research with HALO. This publication was supported by the Open Access Publishing Fund of Leipzig University.

References

Anhaus, P., Katlein, C., Nicolaus, M., Hoppmann, M., and Haas, C.: From Bright Windows to Dark Spots: Snow Cover Controls Melt Pond Optical Properties During Refreezing, Geophysical Research Letters, 48, https://doi.org/10.1029/2021GL095369, 2021.

Bateson, A. W., Feltham, D. L., Schröder, D., Wang, Y., Hwang, B., Ridley, J. K., and Aksenov, Y.: Sea ice floe size: its impact on pan-Arctic and local ice mass and required model complexity, The Cryosphere, 16, 2565–2593, https://doi.org/10.5194/tc-16-2565-2022, 2022.

Belgiu, M. and Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Breiman, L.: Classification and Regression Trees, Routledge, New York, ISBN 978-1-315-13947-0, https://doi.org/10.1201/9781315139470, 2017.

Brown, O. B. and Minnett, P. J.: MODIS Infrared Sea Surface Temperature Algorithm: Algorithm Theoretical Basis Document, Version 2.0, Tech. Rep. NAS5-31361, University of Miami, 1999.

Budikova, D.: Role of Arctic sea ice in global atmospheric circulation: A review, Global and Planetary Change, 68, 149–163, https://doi.org/10.1016/j.gloplacha.2009.04.001, 2009.

Cracknell, A. P.: The advanced very high resolution radiometer (AVHRR), Oceanographic Literature Review, 44, 526, 1997.

De La Rosa, S., Maus, S., and Kern, S.: Thermodynamic investigation of an evolving grease to pancake ice field, Annals of Glaciology, 52, 206–214, https://doi.org/10.3189/172756411795931787, 2011.

Denton, A. and Timmermans, M.: Characterizing the sea-ice floe size distribution in the Canada Basin from high-resolution optical satellite imagery, The Cryosphere, 16, 1563–1578, https://doi.org/10.5194/tc-16-1563-2022, 2022.

Di Biagio, C., Pelon, J., Blanchard, Y., Loyer, L., Hudson, S. R., Walden, V. P., Raut, J. C., Kato, S., Mariage, V., and Granskog, M. A.: Toward a Better Surface Radiation Budget Analysis Over Sea Ice in the High Arctic Ocean: A Comparative Study Between Satellite, Reanalysis, and local-scale Observations, Journal of Geophysical Research: Atmospheres, 126, e2020JD032 555, https://doi.org/https://doi.org/10.1029/2020JD032555, e2020JD032555 2020JD032555, 2021.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., and Sciarra, R.: The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission, Remote Sensing of Environment, 120, 37–57, https://doi.org/https://doi.org/10.1016/j.rse.2011.07.024, the Sentinel Missions - New Opportunities for Science, 2012.

Ehrlich, A., Crewell, S., Herber, A., Klingebiel, M., Lüpkes, C., Mech, M., Becker, S., Borrmann, S., Bozem, H., Buschmann, M., Clemen, H.-C., De La Torre Castro, E., Dorff, H., Dupuy, R., Eppers, O., Ewald, F., George, G., Giez, A., Grawe, S., Gourbeyre, C., Hartmann, J., Jäkel, E., Joppe, P., Jourdan, O., Jurányi, Z., Kirbus, B., Lucke, J., Luebke, A. E., Maahn, M., Maherndl, N., Mallaun, C., Mayer, J., Mertes, S., Mioche, G., Moser, M., Müller, H., Pörtge, V., Risse, N., Roberts, G., Rosenburg, S., Röttenbacher, J., Schäfer, M., Schaefer, J., Schäfler, A., Schirmacher, I., Schneider, J., Schnitt, S., Stratmann, F., Tatzelt, C., Voigt, C., Walbröl, A., Weber, A., Wetzel, B., Wirth, M., and Wendisch, M.: A comprehensive in-situ and remote sensing data set collected during the HALO–(AC)[3] aircraft campaign, Earth System Science Data Discussions, 2024, 1–49, https://doi.org/10.5194/essd-2024-281, 2024.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), Geoscientific Model Development, 9, 1647–1672, https://doi.org/10.5194/gmd-9-1647-2016, 2016.

Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., and Smola, A.: AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data, 2020.

400 Gasteiger, J., Emde, C., Mayer, B., Buras, R., Buehler, S., and Lemke, O.: Representative wavelengths absorption parameterization applied to satellite channels and spectral bands, Journal of Quantitative Spectroscopy and Radiative Transfer, 148, 99–115, https://doi.org/10.1016/j.jqsrt.2014.06.024, 2014.

George, G., Luebke, A. E., Klingebiel, M., Mech, M., and Ehrlich, A.: Dropsonde measurements from HALO and POLAR 5 during HALO-(AC)[3] in 2022, https://doi.org/10.1594/PANGAEA.968891, 2024.

405 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Gryschka, M., Gryanik, V. M., Lüpkes, C., Mostafa, Z., Sühring, M., Witha, B., and Raasch, S.: Turbulent Heat Exchange Over Polar Leads Revisited: A Large Eddy Simulation Study, Journal of Geophysical Research: Atmospheres, 128, e2022JD038 236, https://doi.org/https://doi.org/10.1029/2022JD038236, e2022JD038236 2022JD038236, 2023.

410 Hall, D. K. and Riggs., G. A.: MODIS/Aqua Sea Ice Extent 5-Min L2 Swath 1km, Version 61, https://doi.org/10.5067/MODIS/MYD29.061, 2021.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Sea Ice Extent 5-Min L2 Swath 1km, Version 61, https://doi.org/10.5067/MODIS/MOD29.061, 2021.

Hall, D. K., Key, J. R., Case, K. A., Riggs, G. A., and Cavalieri, D. J.: Sea ice surface temperature product from MODIS, IEEE Transactions

415 on Geoscience and Remote Sensing, 42, 1076–1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Herman, A.: Sea-ice floe-size distribution in the context of spontaneous scaling emergence in stochastic systems, Phys. Rev. E, 81, 066 123, https://doi.org/10.1103/PhysRevE.81.066123, 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., De Chiara, G., Dahlgren, P., Dee,

420 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., de Rosnay, P., Rozum, I., V., F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Jäkel, E., Stapf, J., Wendisch, M., Nicolaus, M., Dorn, W., and Rinke, A.: Validation of the sea ice surface albedo scheme of the regional

425 climate model HIRHAM–NAOSIM using aircraft measurements during the ACLOUD/PASCAL campaigns, The Cryosphere, 13, 1695–1708, https://doi.org/10.5194/tc-13-1695-2019, 2019a.

Jäkel, E., Stapf, J., Wendisch, M., Nicolaus, M., Dorn, W., and Rinke, A.: Validation of the sea ice surface albedo scheme of the regional climate model HIRHAM–NAOSIM using aircraft measurements during the ACLOUD/PASCAL campaigns, The Cryosphere, 13, 1695–1708, https://doi.org/10.5194/tc-13-1695-2019, 2019b.

430 James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J.: An Introduction to Statistical Learning: with Applications in Python, Springer Texts in Statistics, Springer International Publishing, Cham, ISBN 978-3-031-38746-3 978-3-031-38747-0, https://doi.org/10.1007/978-3-031-38747-0, 2023.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., and Lo, W.: Segment anything, arXiv preprint arXiv:2304.02643, 2023.

435 Konow, H., Jacob, M., Ament, F., Crewell, S., Ewald, F., Hagen, M., Hirsch, L., Jansen, F., Mech, M., and Stevens, B.: A unified data set of airborne cloud remote sensing using the HALO Microwave Package (HAMP), Earth System Science Data, 11, 921–934, https://doi.org/10.5194/essd-11-921-2019, 2019.

Kwok, R.: Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (1958–2018), Environmental Research Letters, 13, 105 005, https://doi.org/10.1088/1748-9326/aae3ec, 2018.

440 Li, Tang, B., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., and Sobrino, J. A.: Satellite-derived land surface temperature: Current status and perspectives, Remote Sensing of Environment, 131, 14–37, https://doi.org/10.1016/j.rse.2012.12.008, 2013.

Li, Z., Liu, M., Wang, S., Qu, L., and Guan, L.: Sea Surface Skin Temperature Retrieval from FY-3C/VIRR, Remote Sensing, 14, 1451, https://doi.org/10.3390/rs14061451, 2022.

Liu, A. Q., Moore, G. W. K., Tsuboki, K., and Renfrew, I. A.: The Effect of the Sea-ice Zone on the Development of Boundary-layer Roll
445 Clouds During Cold Air Outbreaks, Boundary-Layer Meteorology, 118, 557–581, https://doi.org/10.1007/s10546-005-6434-4, 2006.

Ludwig, V., Spreen, G., Haas, C., Istomina, L., Kauker, F., and Murashkin, D.: The 2018 North Greenland polynya observed by a newly introduced merged optical and passive microwave sea-ice concentration dataset, The Cryosphere, 13, 2051–2073, https://doi.org/10.5194/tc-13-2051-2019, 2019.

Massom, R. and Comiso, J. C.: The classification of Arctic sea ice types and the determination of surface temperature using advanced very
450 high resolution radiometer data, Journal of Geophysical Research: Oceans, 99, 5201–5218, https://doi.org/10.1029/93JC03449, 1994.

Maykut, G. A. and Untersteiner, N.: Some results from a time-dependent thermodynamic model of sea ice, Journal of Geophysical Research (1896-1977), 76, 1550–1575, https://doi.org/10.1029/JC076i006p01550, 1971.

McMillin and Crosby: Theory and Validation of the Multiple Window Sea Surface Temperature Technique, Journal of Geophysical Research, 89, 3655–3661, 1984.

455 Mech, M., Orlandi, E., Crewell, S., Ament, F., Hirsch, L., Hagen, M., Peters, G., and Stevens, B.: HAMP – the microwave package on the High Altitude and LOng range research aircraft (HALO), Atmospheric Measurement Techniques, 7, 4539–4553, https://doi.org/10.5194/amt-7-4539-2014, 2014.

Meier, W. and Stroeve, J.: An Updated Assessment of the Changing Arctic Sea Ice Cover, Oceanography, https://doi.org/10.5670/oceanog.2022.114, 2022.

460 Miao, X., Xie, H., Ackley, S. F., Perovich, D. K., and Ke, C.: Object-based detection of Arctic sea ice and melt ponds using high spatial resolution aerial photographs, Cold Regions Science and Technology, 119, 211–222, https://doi.org/https://doi.org/10.1016/j.coldregions.2015.06.014, 2015.

Müller, J., Schäfer, M., Rosenburg, S., Ehrlich, A., and Wendisch, M.: Aerial maps of Arctic surface skin temperature and surface type during the HALO-(AC)[3] campaign in March/April 2022, https://doi.pangaea.de/10.1594/PANGAEA.974454, 2025.

465 NASA: Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Level-2 11μm Day/Night Sea Surface Temperature, Version R2019.0 Data, NASA OB.DAAC, Greenbelt, MD, USA, https://doi.org/10.5067/AQUA/MODIS/L2/SST/R2019.0, accessed on 8/2/2024, 2024.

Niehaus, H., Spreen, G., Birnbaum, G., Istomina, L., Jäkel, E., Linhardt, F., Neckel, N., Fuchs, N., Nicolaus, M., Sperzel, T., Tao, R., Webster, M., and Wright, N.: Sea Ice Melt Pond Fraction Derived From Sentinel-2 Data: Along the MOSAiC Drift and Arctic-Wide, Geophysical
470 Research Letters, 50, e2022GL102 102, https://doi.org/https://doi.org/10.1029/2022GL102102, e2022GL102102 2022GL102102, 2023.

Nielsen-Englyst, P., Høyer, J. L., Kolbe, W. M., Dybkjær, G., Lavergne, T., Tonboe, R. T., Skarpalezos, S., and Karagali, I.: A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021, Remote Sensing of Environment, 284, 113 331, https://doi.org/10.1016/j.rse.2022.113331, 2023.

Notz, D. and Community, S.: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, https://doi.org/10.1029/2019GL086749, 2020.

475 Paul, S. and Huntemann, M.: Improved machine-learning-based open-water–sea-ice–cloud discrimination over wintertime Antarctic sea ice using MODIS thermal-infrared imagery, The Cryosphere, 15, 1551–1565, https://doi.org/10.5194/tc-15-1551-2021, 2021.

Qu, M., Pang, X., Zhao, X., Zhang, J., Ji, Q., and Fan, P.: Estimation of turbulent heat flux over leads using satellite thermal images, The Cryosphere, 13, 1565–1582, https://doi.org/10.5194/tc-13-1565-2019, 2019.

Ren, S., Luzi, F., Lahrichi, S., Kassaw, K., Collins, L. M., Bradbury, K., and Malof, J. M.: Segment Anything, From Space?, in: Proceedings
480 of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 8355–8365, 2024.

Rothrock, D. A. and Thorndike, A. S.: Measuring the sea ice floe size distribution, Journal of Geophysical Research: Oceans, 89, 6477–6486, https://doi.org/https://doi.org/10.1029/JC089iC04p06477, 1984.

Schäfer, M., Wolf, K., Ehrlich, A., Hallbauer, C., Jäkel, E., Jansen, F., Luebke, A. E., Müller, J., Thoböll, J., Röschenthaler, T., Stevens, B., and Wendisch, M.: VELOX – a new thermal infrared imager for airborne remote sensing of cloud and surface properties, Atmospheric
485 Measurement Techniques, 15, 1491–1509, https://doi.org/10.5194/amt-15-1491-2022, 2022.

Schäfer, M., Rosenburg, S., Ehrlich, A., Röttenbacher, J., and Wendisch, M.: Two-dimensional cloud-top and surface brightness temperature with 1 Hz temporal resolution derived at flight altitude from VELOX during the HALO-(AC)[3] field campaign, https://doi.pangaea.de/10. 1594/PANGAEA.963401, 2023.

Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOsmanov, Kruchinin, D., Zankevich, A.,
490 DmitriySidnev, Markelov, M., Johannes222, Chenuet, M., a andre, telenachos, Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Priya4607, Tehrani, R., Jeong, S., Skubriev, V., Yonekura, S., vugia truong, zliang7, lizhming, and Truong, T.: opencv/cvat: v1.1.0, https://doi.org/10.5281/zenodo.4009388, 2020.

Skogseth, R., Nilsen, F., and Smedsrud, L. H.: Supercooled water in an Arctic polynya: observations and modeling, Journal of Glaciology, 55, 43–52, https://doi.org/10.3189/002214309788608840, 2009.

495 Smith, G. C., Allard, R., Babin, M., Bertino, L., Chevallier, M., Corlett, G., Crout, J., Davidson, F., Delille, B., Gille, S. T., Hebert, D., Hyder, P., Intrieri, J., Lagunas, J., Larnicol, G., Kaminski, T., Kater, B., Kauker, F., Marec, C., Mazloff, M., Metzger, E. J., Mordy, C., O'Carroll, A., Olsen, S. M., Phelps, M., Posey, P., Prandi, P., Rehm, E., Reid, P., Rigor, I., Sandven, S., Shupe, M., Swart, S., Smedstad, O. M., Solomon, A., Storto, A., Thibaut, P., Toole, J., Wood, K., Xie, J., and Yang, Q.: Polar Ocean Observations: A Critical Gap in the Observing System and Its Effect on Environmental Predictions From Hours to a Season, Frontiers in Marine Science, 6,
500 https://doi.org/10.3389/fmars.2019.00429, 2019.

Spoto, F., Sy, O., Laberinti, P., Martimort, P., Fernandez, V., Colin, O., Hoersch, B., and Meygret, A.: Overview Of Sentinel-2, in: 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 1707–1710, https://doi.org/10.1109/IGARSS.2012.6351195, 2012.

Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-E 89-GHz channels, Journal of Geophysical Research: Oceans, 113, https://doi.org/https://doi.org/10.1029/2005JC003384, 2008.

505 Stamnes, K., Tsay, S., Wiscombe, W., and Laszlo, I.: DISORT, a general-purpose Fortran program for discrete-ordinate-method radiative transfer in scattering and emitting layered media: documentation of methodology, 2000.

Tao, R., Nicolaus, M., Katlein, C., Anhaus, P., Hoppmann, M., Spreen, G., Niehaus, H., Jäkel, E., Wendisch, M., and Haas, C.: Seasonality of spectral radiative fluxes and optical properties of Arctic sea ice during the spring–summer transition, Elementa: Science of the Anthropocene, 12, 00 130, https://doi.org/10.1525/elementa.2023.00130, 2024.

510 Thielke, L., Fuchs, N., Spreen, G., Tremblay, B., Birnbaum, G., Huntemann, M., Hutter, N., Itkin, P., Jutila, A., and Webster, M. A.: Preconditioning of Summer Melt Ponds From Winter Sea Ice Surface Temperature, Geophysical Research Letters, 50, e2022GL101 493, https://doi.org/10.1029/2022GL101493, e2022GL101493 2022GL101493, 2023.

Vincent: The Case for a Single Channel Composite Arctic Sea Surface Temperature Algorithm, Remote Sensing, 11, https://doi.org/10.3390/rs11202393, 2019.

515 Vincent, Marsden, R. F., Minnett, P. J., Creber, K. A. M., and Buckley, J. R.: Arctic waters and marginal ice zones: A composite Arctic sea surface temperature algorithm using satellite thermal data, Journal of Geophysical Research: Oceans, 113, https://doi.org/10.1029/2007JC004353, 2008.

Walbröl, A., Michaelis, J., Becker, S., Dorff, H., Gorodetskaya, I., Kirbus, B., Lauer, M., Maherndl, N., Maturilli, M., Mayer, J., Müller, H., Neggers, R., Paulus, F., Röttenbacher, J., Rückert, J., Schirmacher, I., Slättberg, N., Ehrlich, A., Wendisch, M., and Crewell, S.: Environ-

520 mental conditions in the North Atlantic sector of the Arctic during the HALO–$(\mathcal{AC})^3$ campaign, https://doi.org/10.5194/egusphere-2023-668, 2023.

Wendisch, M., Brückner, M., Crewell, S., Ehrlich, A., Notholt, J., Lüpkes, C., Macke, A., Burrows, J. P., Rinke, A., Quaas, J., Maturilli, M., Schemann, V., Shupe, M. D., Akansu, E. F., Barrientos-Velasco, C., Bärfuss, K., Blechschmidt, A.-M., Block, K., Bougoudis, I., Bozem, H., Böckmann, C., Bracher, A., Bresson, H., Bretschneider, L., Buschmann, M., Chechin, D. G., Chylik, J., Dahlke, S., Deneke,

525 H., Dethloff, K., Donth, T., Dorn, W., Dupuy, R., Ebell, K., Egerer, U., Engelmann, R., Eppers, O., Gerdes, R., Gierens, R., Gorodetskaya, I. V., Gottschalk, M., Griesche, H., Gryanik, V. M., Handorf, D., Harm-Altstädter, B., Hartmann, J., Hartmann, M., Heinold, B., Herber, A., Herrmann, H., Heygster, G., Höschel, I., Hofmann, Z., Hölemann, J., Hünerbein, A., Jafariserajehlou, S., Jäkel, E., Jacobi, C., Janout, M., Jansen, F., Jourdan, O., Jurányi, Z., Kalesse-Los, H., Kanzow, T., Käthner, R., Kliesch, L. L., Klingebiel, M., Knudsen, E. M., Kovács, T., Körtke, W., Krampe, D., Kretzschmar, J., Kreyling, D., Kulla, B., Kunkel, D., Lampert, A., Lauer, M., Lelli, L., von Lerber, A., Linke,

530 O., Löhnert, U., Lonardi, M., Losa, S. N., Losch, M., Maahn, M., Mech, M., Mei, L., Mertes, S., Metzner, E., Mewes, D., Michaelis, J., Mioche, G., Moser, M., Nakoudi, K., Neggers, R., Neuber, R., Nomokonova, T., Oelker, J., Papakonstantinou-Presvelou, I., Pätzold, F., Pefanis, V., Pohl, C., van Pinxteren, M., Radovan, A., Rhein, M., Rex, M., Richter, A., Risse, N., Ritter, C., Rostosky, P., Rozanov, V. V., Donoso, E. R., Saavedra Garfias, P., Salzmann, M., Schacht, J., Schäfer, M., Schneider, J., Schnierstein, N., Seifert, P., Seo, S., Siebert, H., Soppa, M. A., Spreen, G., Stachlewska, I. S., Stapf, J., Stratmann, F., Tegen, I., Viceto, C., Voigt, C., Vountas, M., Walbröl,

535 A., Walter, M., Wehner, B., Wex, H., Willmes, S., Zanatta, M., and Zeppenfeld, S.: Atmospheric and Surface Processes, and Feedback Mechanisms Determining Arctic Amplification: A Review of First Results and Prospects of the $(\mathcal{AC})^3$ Project, Bulletin of the American Meteorological Society, 104, E208–E242, https://doi.org/10.1175/BAMS-D-21-0218.1, 2023a.

Wendisch, M., Stapf, J., Becker, S., Ehrlich, A., Jäkel, E., Klingebiel, M., Lüpkes, C., Schäfer, M., and Shupe, M. D.: Effects of variable ice–ocean surface properties and air mass transformation on the Arctic radiative energy budget, Atmospheric Chemistry and Physics, 23,

540 9647–9667, https://doi.org/10.5194/acp-23-9647-2023, 2023b.

Wendisch, M., Crewell, S., Ehrlich, A., Herber, A., Kirbus, B., Lüpkes, C., Mech, M., Abel, S. J., Akansu, E. F., Ament, F., Aubry, C., Becker, S., Borrmann, S., Bozem, H., Brückner, M., Clemen, H.-C., Dahlke, S., Dekoutsidis, G., Delanoë, J., De La Torre Castro, E., Dorff, H., Dupuy, R., Eppers, O., Ewald, F., George, G., Gorodetskaya, I. V., Grawe, S., Groß, S., Hartmann, J., Henning, S., Hirsch, L., Jäkel, E., Joppe, P., Jourdan, O., Jurányi, Z., Karalis, M., Kellermann, M., Klingebiel, M., Lonardi, M., Lucke, J., Luebke, A. E., Maahn, M.,

545   Maherndl, N., Maturilli, M., Mayer, B., Mayer, J., Mertes, S., Michaelis, J., Michalkov, M., Mioche, G., Moser, M., Müller, H., Neggers, R., Ori, D., Paul, D., Paulus, F. M., Pilz, C., Pithan, F., Pöhlker, M., Pörtge, V., Ringel, M., Risse, N., Roberts, G. C., Rosenburg, S., Röttenbacher, J., Rückert, J., Schäfer, M., Schaefer, J., Schemann, V., Schirmacher, I., Schmidt, J., Schmidt, S., Schneider, J., Schnitt, S., Schwarz, A., Siebert, H., Sodemann, H., Sperzel, T., Spreen, G., Stevens, B., Stratmann, F., Svensson, G., Tatzelt, C., Tuch, T., Vihma, T., Voigt, C., Volkmer, L., Walbröl, A., Weber, A., Wehner, B., Wetzel, B., Wirth, M., and Zinner, T.: Overview: quasi-Lagrangian observations

550   of Arctic air mass transformations – introduction and initial results of the HALO–$(\mathcal{AC})^3$ aircraft campaign, Atmospheric Chemistry and Physics, 24, 8865–8892, https://doi.org/10.5194/acp-24-8865-2024, 2024.

Willmes, S. and Heinemann, G.: Pan-Arctic lead detection from MODIS thermal infrared imagery, Annals of Glaciology, 56, 29–37, https://doi.org/10.3189/2015AoG69A615, 2015.

Wirth, M., Fix, A., Mahnke, P., Schwarzer, H., Schrandt, F., and Ehret, G.: The airborne multi-wavelength water vapor differential absorption

555   lidar WALES: system design and performance, Applied Physics B, 96, 201–213, https://doi.org/10.1007/s00340-009-3365-7, 2009.

Wright, N. C. and Polashenski, C. M.: Open-source algorithm for detecting sea ice surface features in high-resolution optical imagery, The Cryosphere, 12, 1307–1329, https://doi.org/10.5194/tc-12-1307-2018, 2018.

Wu, Q. and Osco, L. P.: samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM), Journal of Open Source Software, 8, 5663, https://doi.org/10.21105/joss.05663, 2023.