Improved machine-learning based open-water/sea-ice/cloud discrimination over wintertime Antarctic sea ice using MODIS thermal-infrared imagery

Stephan Paul[1,2] and Marcus Huntemann[3]

[1]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[2]Deutsches Geodätisches Forschungsinstitut (DGFI), Technical University Munich, Munich, Germany
[3]Department of Environmental Physics, University of Bremen, Bremen, Germany
Correspondence: Stephan Paul (stephan.paul@awi.de)

Abstract. The frequent presence of cloud cover in polar regions limits the use of the Moderate-Resolution Imageing Spectro-radiometer Spectroradiometer (MODIS) and similar instruments for the investigation and monitoring of sea-ice polynyas compared to passive-microwave-based sensors. The very low thermal contrast between present clouds and the sea-ice surface in combination with the lack of available visible and near-infrared channels during polar nighttime results in deficiencies in the MODIS cloud mask and dependent MODIS data products. This leads to frequent misclassifications of i) present clouds as sea ice/open water (false-negative) and ii) open- wateropen-water/thin-ice areas as clouds (false-positive), which results in an underestimation of actual polynya area and subsequently subsequent derived information. Here, we present a novel machine-learning based approach using a deep neural network that is able to reliably discriminate between clouds, sea-ice, and open-water/thin-ice areas in a given swath solely from thermal-infrared MODIS channels and additionally derived derived additional information. Compared to the reference MODIS sea-ice product for the year 2017, our data results in an overall increase of 31% 20 % in annual swath-based coverage for the Brunt Ice Shelf polynya, attributed to an improved cloud-cover discrimination . Overall, and the reduction of false-positive classifications. At the same time, the mean annual polynya area decreases by 44 % through the reduction of false-negative classifications of warm clouds as thin ice. Additionally, higher spatial coverage results in a an overall better sub-daily representation of thin-ice conditions that cannot be reconstructed with current state-of-the-art cloud-cover compensation methods.

1 Introduction

Information on cloud presence is of crucial importance when using thermal-infrared imagery. This is especially true for the polar regions, where the thermal contrast between clouds and the underlying snow and sea-ice surface can be low through persistent surface temperature inversion and low clouds (Welch et al., 1992). Furthermore, occurrences of warm clouds over cold sea ice as well as cold clouds over relatively warm and thin sea ice are both possible. Despite improvements (Liu et al., 2004; Frey et al., 2008; Holz et al., 2008; Liu and Key, 2014), the performance of the frequently used Moderate-Resolution
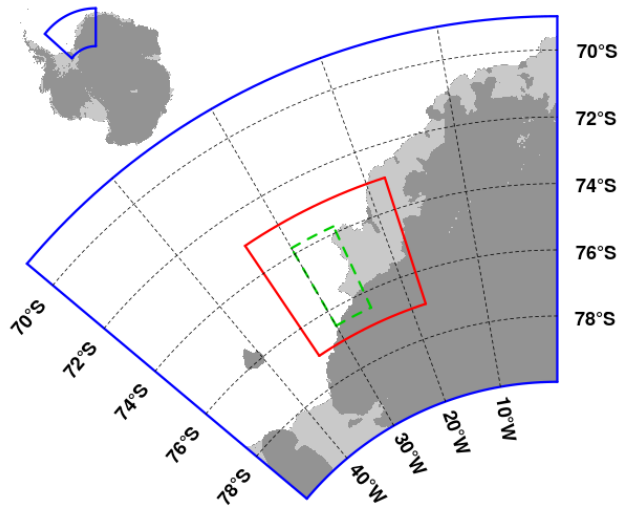
Figure 1. Location of the general (red) and focus (green) study area of the Antarctic Brunt Ice Shelf study area (red) in the south-eastern Weddell Sea (blackblue). Data of land ice (dark gray) and floating ice shelves (light gray) are retrieved from Rtopo-2 (Schaffer et al., 2016).

Imaging Spectroradiometer (MODIS) cloud mask product (MOD35/MYD35; Ackerman et al., 2015) is substantially reduced during polar nighttime compared to its performance during daytime conditions.

Nonetheless, several studies use MODIS thermal-infrared (TIR) data to monitor polyana polynya area and associated sea-ice

25 production in polynyas both in the Arctic as well as the Antarctic and compare well to or even outperform studies using passive-microwave satellite data in certain regions (e.g., Paul et al., 2015; Preußer et al., 2019)(e.g., Paul et al., 2015; Aulicino et al., 2018; Preußer et al., 2019). These studies generally utilize ice-surface temperature from the National Snow and Ice Data Center (NSIDC) sea-ice product (MOD/MYD29; Hall et al., 2004; Hall and Riggs, 2015a,b). The MOD/MYD29 product is derived from both MODIS sensors onboard the NASA polar orbiting Aqua and Terra satellites with the MOD/MYD35 cloud mask

30 product already applied (Riggs and Hall, 2015). However, especially positive temperature-anomaly features such as large warm open-water areas through sea-ice polynyas pose a problem for the MODIS cloud mask and result in frequent misclassification of these areas as cloud cover (Fraser et al., 2009). Additionally, other MODIS applications would potentially benefit from an improved wintertime cloud masking. These applications comprise composite generation (e.g., Fraser et al., 2010, 2020), merged optical and passive microwave sensor applications (e.g., Ludwig et al., 2019), basin-wide lead detection from

35 thermal-infrared data (e.g., Reiser et al., 2020), as well as sea-ice motion tracking through image cross-correlation.

In this study, we propose a new novel machine-learning based approach to discriminate between open-water/thin-ice, sea-ice and cloud-covered areas in MODIS TIR swaths. We evaluate and analyze the use of two different approaches – Random Forests (Breiman, 2001) and a Neural Network (e.g., Kohonen, 1988) – a deep neural network (e.g., Kohonen, 1988; Goodfellow et al., 2016) building upon a comprehensive set of newly generated labeled training data. The data set is derived using a combined approach of unsupervised

40 deep-learning, subsequent clustering, and manual screening from co-located full channel 1km resolution MOD/MYD02 product
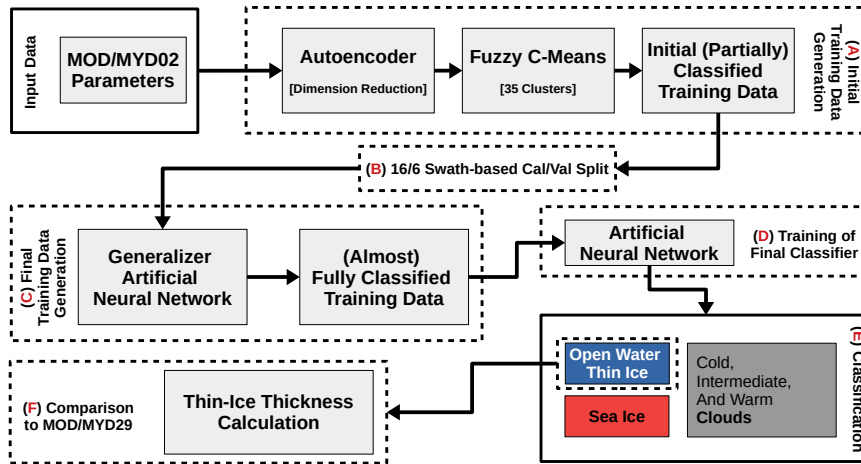
Figure 2. Flow chart summarizing all processing steps from the generation of the initial training data through manual classification to the training of the final classifier and its application for open-water/sea-ice/cloud discrimination.

data (MODIS Characterization Support Team (MCST), 2017a,b) accessed through the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) and Sentinel-1 A/B (S1-A/B) synthetic aperture radar (SAR) calibrated backscatter data ~~acessed~~ accessed through the Alaska Satellite Facility (ASF) DAAC as a cloud-independent reference.

45   The resulting classifier performance is then analyzed and evaluated based on wintertime estimates of resulting polynya area in comparison to the MOD/MYD29 reference product for the Brunt Ice Shelf (BIS) region in the Antarctic Weddell Sea in the year 2017 (Fig. 1). This region was chosen for its combination of high inter-annual polynya activity and high spatio-temporal coverage with Sentinel-1 data. Results are expected to be transferable to other polynya regions in the Antarctic.

50   In the following sections, we will first describe our methodology and input data starting with the employed basic methods and algorithms (Sect. 2.1) ~~as well as~~ followed by the used input data (Sect. 2.2), a detailed explanation of the initial training data generation scheme (Sect. 2.3) and the subsequent ~~training and evaluation of the initial supervised classifiers (Sect. ??&2.4). Next, we describe the~~ processing steps that lead to our final classifier (Sect. ~~2.5) and, finally,~~ 2.4 & 2.5). Finally, we describe and discuss our results (Sect. 3) in comparison to standard MOD/MYD29-derived estimates as well as using co-located S1-A/B SAR reference data. In the
55   end we provide a summary and an outlook to future applications (Sect. 4).

2   Data and Methods

In the following subsections we describe our methods and input data that lead to our deep neural network for the sea-ice, cloud, and open-water/thin-ice discrimination (Fig. 2).

2.1 Basic methods and algorithms

60 This section intends to provide a basic introduction to the methods used in this study. However, it would be beyond the scope of this article to provide an exhaustive review of these methods. Please refer to the provided references for more details For more details, additional references are provided.

All computations for this study were carried out using the R software (R Core Team, 2018) running on a commercially available laptop.

2.1.1 Gray-level co-occurrence matrices (GLCM)

Gray-level co-occurrence matrices (GLCM) are a tool to quantify spatial texture based on brightness values of a pixel neighborhood (Haralick et al., 1973; Haralick, 1979; Hall-Beyer, 2017; R: Zvoleff, 2019). The directional-dependent occurrence frequencies of brightness-value combinations are counted and normalized to probabilities. Subsequently, several statistical measures can be calculated from the GLCM as an additional descriptive statistic of the data.

70 Haralick et al. (1973) proposed 14 different metrics, however, not all were commonly adopted and implemented into modern software. For R, eight different measures are implemented (Zvoleff, 2019), from which we utilized three four: GLCM mean, GLCM variance, and Homogeneity Contrast and Entropy (Tab. 1).

Hall-Beyer (2017) showed that GLCM variance can be associated with edges of different class patches, while GLCM mean and Homogeneity Contrast, and Entropy correspond well to patch-interior texture.

75 In general, the use of GLCM texture metrics is suitable for cloud detection and classification in polar regions using visual, near-, and thermal infrared satellite data (Welch et al., 1992). However, as the size of each GLCM per pixel in a sliding-neighborhood window corresponds and increases proportionally to the image bit depth, computational cost increases rapidly for i) large sliding windows and ii) a large number of gray-levels in the input data. For our study, all MOD/MYD02 channel-based input parameters for the GLCM computations were re-scaled to 32 gray-levels, using a 3 7×3 7 sliding-neighborhood 80 window with horizontal, vertical as well as diagonal directional pixel relationships.

2.1.2 Fuzzy c-means clustering (FCM)

For initial clustering of our initial training data, we utilize an unsupervised procedure called fuzzy c-means clustering (FCM; Dunn, 1973; Bezdek et al., 1984; R: Meyer et al., 2019).

The FCM is comparable to a classic k-means clustering approach (MacQueen, 1967; Hartigan and Wong, 1979), with the 85 addition of providing cluster membership probabilities for each pixel. This type of clustering is also referred to as 'soft' clustering. In contrast to 'hard'-clustering approaches such as k-means, FCM allows for a pixel to belong into several clusters with a certain probability. This type of clustering is also referred to as 'soft' clustering.

For this type of unsupervised clustering, it is necessary to preselect the number of clusters in which the input data should be separated into. Without a-priori knowledge about potential relationships and correlations between predictors, it is standard

90 common practice to choose a large number of initial clusters and manually merge similar clusters afterwards to the desired number of classes.

In this study, we always use a setup of 35 clusters and stop the clustering process after 30 iterations.

2.1.3 Artificial neural networks (NN)

An artificial neural network (NN) generally consists of several neurons organized in hierarchical sequential layers in which each
95 neuron of a layer is fully interconnected to all neurons in the adjacent two layers through weighted paths. These neurons respond to the weighted input of the preceding neurons and pass on their output to the adjacent neurons, modulated based on a type of activation function (Kohonen, 1988; Lee et al., 1990; Welch et al., 1992; Atkinson and Tatnall, 1997; LeCun et al., 2015; Schmidhuber, 2015 Kohonen, 1988; Lee et al., 1990; Welch et al., 1992; Atkinson and Tatnall, 1997; LeCun et al., 2015; Schmidhuber, 2015; Goodfellow et al., 2016; R: Lammers, 2019 Allaire and Chollet, 2020).
100 Once trained, NN are powerful tools for fast and efficient processing of large amounts of remote sensing data and have been shown to be more accurate, e.g., in classification tasks, than other techniques (Kohonen, 1988; Lee et al., 1990; Atkinson and Tatnall, 1997).

Furthermore, NN can represent complex and non-linear functions without formal description through learning from labeled training data. In contrast to statistical methods, NN allow to incorporate data from different sources and require no knowledge
105 or assumptions about its parametric distributions. Hence, in contrast to statistical approaches, NN solely depend on to provided input data (Lee et al., 1990; Atkinson and Tatnall, 1997; LeCun et al., 2015).

In their simplest form, a so-called 'shallow' NN consist of an input layer, a hidden layer and an output layer. Input-layer neurons correspond to the number of input features/predictors, whereas output layer neurons correspond in our case case of classification tasks to the number of classes the input data should be categorized into. With an increasing number of hidden
110 layers, so-called 'deep' NN can handle even more complex problems (Atkinson and Tatnall, 1997; Schmidhuber, 2015).

In this study, we experiment with different numbers of hidden layers, activation functions, and the number of neurons per hidden layer. While some general suggestions for the NN architecture exist, solutions are often found empirically by maximizing the accuracy through minimizing/maximizing the loss function/accuracy for both calibration and validation data classification without overfitting the model. This process is described in the following subsections. For all NN in this study, we use a logarithmic loss function and the adam optimizer (Kingma and Ba, 2014) as well as
115 either rectified linear unit (ReLU) or hyberbolic tangent (TanH) activation functions (LeCun et al., 2015).

In addition to these general NN, we work with a second type called an autoencoder (AE). An AE is a specialized variant of a NN used for anomaly detection and dimension reduction (Cao et al., 2018; Dong et al., 2018; R: Lammers, 2019 Allaire and Chollet, 2020).

In a typical AE, the output or target data is equal to the input data. However, all information is forced through a bottle-neck
120 hidden layer. The result relies on the capability of the bottle-neck hidden-layer neurons to extract relevant features information from the training data to enable the AE to reconstruct the input image with minimized error (Cao et al., 2018).

This is achieved by constructing two branches of symmetric hidden layers of neurons (called the encoder and the decoder, respectively) around a bottle-neck neuron layer generally consisting of very few neurons (Cao et al., 2018). The resulting encoder part of the AE can then be used for dimension reduction.

125   In this study, we trained the AE in a **39-23-10-3**-10-23-39 fully-coupled hidden-layer setup with ReLU activation functions in 35 epochs, using a batch size of 100 and a pseudo-huber loss function (Charbonnier et al., 1997). The used parameters for the AE training and its application as well as for the NN are summarized in Table 1.

2.1.4   Random forests (RF)

Random Forests (RF) are an ensemble supervised machine-learning method based on multiple decision trees (Breiman, 2001; R: Liaw and Wiener, 2002). Each decision tree in a RF is a rather simple statistical tool to predict data categories based on thresholds used in several steps to split the input data. When visualized, a decision tree resembles a tree with
130   an increasing number of branches between nodes, leading to the final categories.

During the RF training, all single decision trees are fitted to the input training data following three rules/constraints (Breiman, 2001): i) for each decision tree in a RF, training data of the same size as the input data is randomly sampled with replacement from the input data. ii) for M input parameters ideally a fixed number $m \ll M$ is specified and randomly selected out of M. The best split on these selected parameters m is then used to split the training data at a specific node. Throughout the growth of the RF, the value of m is held constant. iii) each tree is grown out fully with no pruning applied (i.e., data is split until each data point is categorized).

135   In contrast to single decision trees that tend to overfit (i.e., match data too precisely and therefore fail for any additional data), RF do not overfit and are also capable of dealing with unbalanced data sets (Breiman, 2001).

An additional benefit of the RF is its internally calculated importance measure for each provided predictor. In other words, it counts the frequency of use of each predictor in the growth process of each tree as well as the loss in overall accuracy when dropping this predictor. Furthermore, similar to the NN approach, RF are also capable to provide probabilistic classification in addition to a standard binary classification (Liaw and Wiener, 2002).

140 ## 2.2   Input data

In total, we use four different types of data sets for the year 2017:

1. MODIS Level 1B Calibrated Radiances obtained from the MODIS sensors on-board the polar-orbiting NASA satellites Terra and Aqua (MOD/MYD02; MODIS Characterization Support Team (MCST), 2017a,b; retrieved from the LAADS DAAC at: https://ladsweb.modaps.eosdis.nasa.gov/) with a spatial resolution of $1\,\text{km} \times 1\,\text{km}$ at nadir and swath dimen-
145   sions of $1354\,\text{km}$ (across track) $\times\ 2030\,\text{km}$ (along track),

2. Sentinel-1 A/B Level 1 calibrated backscatter data (S1-A/B; retrieved from the ASF DAAC at: https://asf.alaska.edu/ and processed by ESA) with a spatial resolution of $20\,\text{m} \times 20\,\text{m}$,

3. NSIDC MODIS Sea Ice product (MOD/MYD29, Hall et al., 2004; Riggs and Hall, 2015) in the same resolution as the MOD/MYD02 data but comprising a pre-computed and MODIS cloud-masked applied ice-surface temperature data
150   set (IST) data set, as well as

4. ECMWF ERA-Interim atmospheric reanalysis data (Dee et al., 2011) featuring a spatial resolution of $0.75\,^\circ$ and a temporal resolution of $6\,\text{h}$.

Table 1. Summary of all used parameters, their source product/sensor as well as their application in this study. These parameters comprise brightness temperatures (BT) from the selected MOD/MYD02 channel subset ($\star$) , as well as BT differences (ΔBT) to the neighboring swaths ($\bullet$; i.e. previous or next swath covering the same area) as well as normalized BT (BT_{norm}; $\diamond\bullet$). Furthermore, ice-surface temperatures (IST) from MOD/MYD02 together with the IST from neighboring swaths ($IST_{Neighbors}$) and the time-normalized IST change ($IST_{\Delta t}$) between them as well as IST from the MOD/MYD29 product. The texture metrics calculated from GLCM (Mean, Variance, Contrast, and HomogeneityEntropy), as well as the calibrated backscatter (σ^0) from Sentinel-1 A/B as reference (R). Finally, the atmospheric parameters taken from the ERA-Interim reanalysis necessary for the calculation of thin-ice thickness (TIT). The applications comprise primarily their use in Neural Network the neural network (NN) or Random Forest (RF) classifier and autoencoder (AE) training.Bold NN marks the parameters used for the final classifier.

Symbol/Abbreviation	Parameter	S
BT [$\star$]	Brightness Temperatures	M
ΔBT $\star$/$\bullet$Brightness Temperature Differences MOD/MYD02 AE BT_{norm} [$\star$/$\diamond$]	Normalized Brightness Temperatures	M
IST	Ice-Surface Temperature	M
$IST_{Neighbors}$	Ice-Surface Temperature of neighbor swaths	M
$IST_{\Delta t}$	Time-normalized Ice-Surface Temperature difference to neighbor swaths	M
$GLCM_{Mean}$ [$\star$]	Mean of the GLCM	M
$GLCM_{Var}$ [$\star$]	Variance of the GLCM	M
$GLCM_{Hom}$ $GLCM_{Con}$ [$\star$]	Homogeneity Contrast of the GLCM	M
$GLCM_{Ent}$ [$\star$]	Entropy of the GLCM	M
IST	Ice-Surface Temperature	M
σ^0	Calibrated Backscatter	S
$T2m$	2 m Temperature	E
T_d2m	2 m Dew-Point Temperature	E
$mslp$	Mean Sea-Level Pressure	E
$u10m$	10 m u Wind Component	E
$v10m$	10 m v Wind Component	E

AE/NN=Autoencoder/Neural Network; R=Reference; TIT=Thin-Ice-Thickness Calculation

$\star$=Calculated/Derived for MODIS channels: 20, 25, 31, and 33;

$\bullet$=Normalized through swath-wide mean and standard deviation: $BT_{norm} = (BT - \overline{BT}) \times \sigma_{BT}^{-1}$

An overview of all used input parameters with their respective source as well as their application is provided in Table 1.

All MODIS and ERA-Interim data are resampled to a common equi-rectangular grid of the Brunt Ice Shelf (BIS) area with an average spatial resolution of 1 km × 1 km and an extent from 34 °W to 18 °W and 77 °S to 73 °S using a nearest-neighbor approach. For visual reference, the S1-A/B data is also resampled to an equi-rectangular grid with the same extent but a

spatial resolution of 25 m. Through the decreasing distance between meridians towards the pole, the per-pixel spatial area also decreases. This results from the constant latitudinal distance between grid points in this type of projection. Ice-shelf areas are excluded from our analysis based on Rtopo-2 data (Schaffer et al., 2016).

2.2.1 MOD/MYD02 L1b calibrated radiances

Our goal for the later discrimination algorithm was for it to solely rely on MODIS-channel data, without the need for any auxiliary data.

Brightness temperatures (BT) were calculated from calibrated radiances comprising MODIS channels 20, 22, 23, 24, 25, 29, 31, 32, and 33 , and 35 following Toller et al. (2009). This channel subset allows to distinguish between sea-ice, open-water/thin-ice, and cloud pixels through a high inter-channel variability.

Furthermore, we computed additional parameters from these BT comprising , while reducing the impact of stripes in the MODIS data. Additionally, channel 32 data is used for the calculation of the ice-surface temperature (IST; following (Riggs and Hall, 2015)as well as . Furthermore, we computed image-texture parameters using GLCM (Tab. 1). For this we use MODIS collection 6.1 data.

We generally limited our swaths to study to swaths featuring sensor incidence angles ≤60 50 ° in 65 % of the study area (to minimize spatial distortion towards the swath edgesand a coverage) and a total coverage of our study area >90 %of our reference . In order to aid the manual categorization by providing favorable geometries, the MODIS colocation swath to the S1-A/B reference data needs to feature sensor incidence angles ≤35 ° in 65 % of the study area.

2.2.2 MOD/MYD29 sea-ice product

For a later comparison based on cloud coverage , and polynya area, and sea-ice production rates, we extracted and use the IST from the reference NSIDC MOD/MYD29 sea-ice product produced from MODIS collection 6 data, which offers an overall accuracy of 1–3 K under ideal (i.e., clear-sky) conditions (Hall et al., 2004; Riggs and Hall, 2015).

Both IST (MOD/MYD02 and MOD/MYD29) are derived based on a constant emissivity for snow/ice (Hall et al., 2015), but with the MODIS cloud mask already applied to the MOD/MYD29 product.

2.2.3 S1-A/B L1 calibrated backscatter

In order to reliably identify polynyas independent of cloud-cover or other atmospheric disturbances, we selected a total of eight 22 S1-A/B swaths featuring an active polynya in front of the BIS.

These S1-A/B swaths together with co-located and at least partially cloud-free MOD/MYD02 data are used for training calibration and validation of the algorithm. S1-A/B swath acquisition times are temporarily distributed over the 2017 Antarctic winter, with all additional information summarized in Table 2.

Table 2. List of used S1-A/B swaths where the top six were used for calibration/training, validation, and bottom two only for a detailed analysis (Fig. ??6).

Satellite	Product	Acquisition in UTC
Calibration/Training		
S1-B	IW_GRDH_1SSH	2017-04-02 03:49:42
S1-A	EW_GRDM_1SSH	2017-04-10 23:23:10
S1-A	EW_GRDM_1SSH	2017-04-07 22:58:51
S1-A	EW_GRDM_1SSH	2017-05-18 23:06:59
S1-A	EW_GRDM_1SSH	2017-05-28 00:19:56
S1-A	EW_GRDM_1SSH	2017-05-28 23:23:13
S1-A	EW_GRDM_1SSH	2017-06-21 23:23:14
S1-A	EW_GRDM_1SSH	2017-07-03 23:23:15
S1-A	EW_GRDM_1SSH	2017-07-08 00:28:10
S1-A	EW_GRDM_1SSH	2017-07-08 23:31:21
S1-A S1-B	IW_GRDH_1SSH	2017-07-13 2017-07-31 03:50:25 49:48
S1-A	IWEW_GRDHGRDM_1SSH	2017-08-06 03:50:27 2017-08-08 23:23:17
S1-B	IW_GRDH_1SSH	2017-08-19 23:30:41
S1-A	IW_GRDH_1SSH	2017-09-01 23:23:18
S1-A	EW_GRDM_1SSH	2017-09-20 00:11:57
S1-A	EW_GRDM_1SSH	2017-09-25 23:23:19
Validation		
S1-A	EW_GRDM_1SSH	2017-04-07 22:58:51
S1-B	EW_GRDM_1SSH	2017-04-09 00:27:24
S1-A	EW_GRDM_1SSH	2017-05-11 00:11:50
S1-A	EW_GRDM_1SSH	2017-07-20 00:28:11
S1-A	IW_GRDH_1SSH	2017-08-06 03:50:27
S1-A	IW_GRDH_1SSH	2017-09-11 03:50:28
Example		
S1-A	EW_GRDM_1SSH	2017-05-16 23:23:12
S1-A	EW_GRDM_1SSH	2017-05-18 23:06:59

185 ### 2.2.4 ERA-Interim data and thin-ice retrieval

For a quantitative comparison between resulting sea-ice production rates and polynya area polynya area (i.e., the total area of pixels covered with a maximum ice thickness of 0.2 m), we calculate the thin-ice thickness (TIT) from MODIS IST for MOD/MYD02
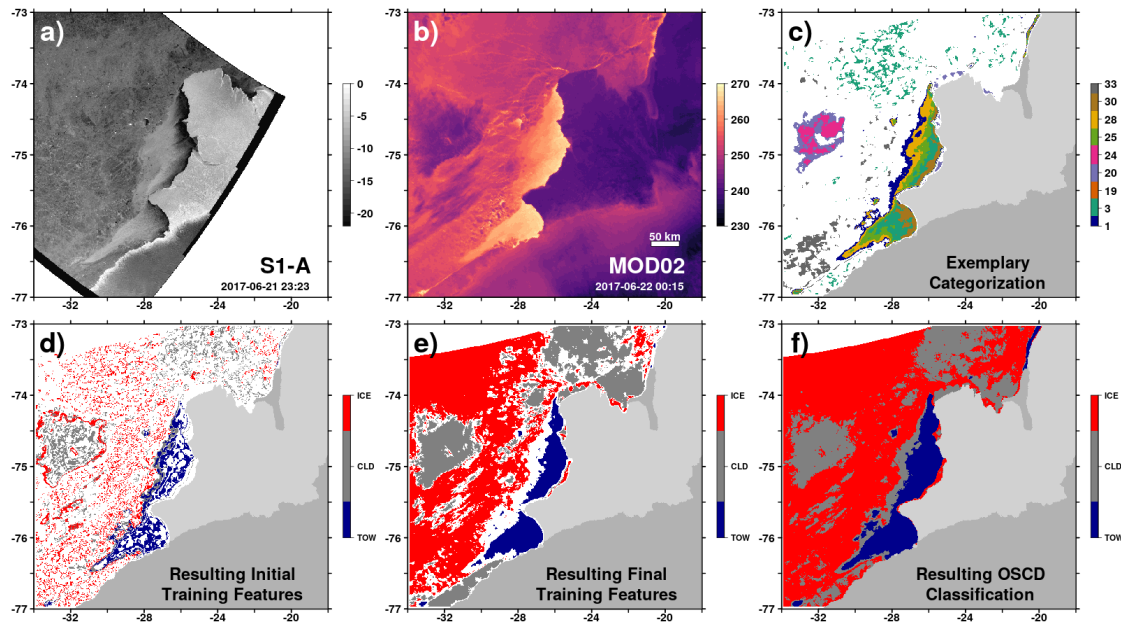
Figure 3. ~~Examplary~~ Exemplary generation of labeled training data with reference Sentinel-1 A/B calibrated backscatter image (in dB; a), MOD02 derived ice-surface temperature (in K; b), an exemplary subset of ~~exemplary results~~ nine clusters out of the 35 total clusters from the used Autoencoder and Fuzzy C-Means Clustering before (c) and after (d) manual categorization and quality control (d; thin-ice/open-water, TOW; clouds, CLD; sea ice, ICE), as well as the final training based on the generalizing NN (e), and the ~~corresponding ice-surface temperature of~~ resulting classification based on the ~~selected example cluster subsets~~ final OSCD algorithm (~~in ;~~ df). ~~Here~~ Land-ice (dark gray) and ice-shelf (light gray) overlays originate from Rtopo-2 (Schaffer et al., 2016). In (c), clusters ~~3, 7, 18,~~ 20 and ~~23~~ 24 were categorized as 'cloud'; clusters ~~13~~ 1, 3, 25, 28, and ~~14~~ 30 as 'open-water/thin-ice'; and ~~cluster~~ clusters 19 and 33 as 'sea ice'. ~~Land-ice (dark gray) and ice-shelf (light gray) overlays originate from Rtopo-2 (Schaffer et al., 2016).~~

~~und~~ and MOD/MYD29 data using a surface-energy-balance model together with the ERA-Interim 2 m air temperature, the 10 m wind-speed components, the mean sea-level pressure, and the 2 m dew-point temperature (Dee et al., 2011).

190   The surface-energy-balance model utilizes the inversely proportional relation between IST and the thickness of thin sea ice (Yu and Rothrock, 1996; Drucker et al., 2003). Generally, thinner ice features a higher IST than thicker ice as the influence from the warm ocean is diminished. The net positive flux towards the atmosphere between the warm ocean and the cold atmosphere is equalized from the conductive heat flux through the ice. From the conductive heat flux TIT is derived. A detailed description of the retrieval procedure as well as all equations and necessary assumptions are thoroughly described in Paul et al. (2015) as well as Adams
195   et al. (2013). For ice thicknesses between 0.0 m and 0.2 m, Adams et al. (2013) state an average uncertainty of ± 4.7 cm.

2.3 Initial training data generation

The availability and quality of labeled training data ~~is~~ are of utmost importance for the training of any supervised machine-learning algorithm. However, available spatio-temporal high-resolution cloud information over nighttime sea ice is practically non-existent. Therefore, we had to derive our own labeled training data using co-located MODIS and S1-A/B data to manually identify cloud, sea-ice as well as open-water/thin-ice pixels, respectively . (Fig. 2A).

To reduce manual effort and uncertainty to a minimum, we employ a mix of dimension reduction and unsupervised clustering before the final manual ~~classification~~categorization.

First, we selected up to 13 MODIS swaths in close temporal proximity for each of the ~~six~~ 22 S1-A/B reference swaths (Top six in Tab. 2), i.e. in a temporal range of $\pm_2 36\,$h around the S1-A/B swath. We chose the best temporal match based sensor zenith angle (65 % of the study area feature an angle $\leq 35\,°$) and swath coverage of our study area ($\geq 90\,$%). In this way, the data represent rather easy to distinguish configurations of cloud, sea-ice and open-water/thin-ice pixels with favorable geometries for manual categorization.

Secondly, in addition to the textural parameters from the GLCM (Tab. 1), we wanted to add a temporal component to the parameter mix. We therefore added the IST of two swaths acquired before and after the current swath, respectively. These four swaths were taken from the pool of selected MODIS swaths and arranged in temporal patterns before and after the best match. Additionally, we added the time-normalized IST difference between all these neighboring swaths.

From here, we take advantage of the AE dimension reduction capabilities (Fig. 2). Hence, instead A). Instead of using the total number of ~~39~~ 33 input parameters for the FCM with probably only mediocre results (Tab. 1), we cluster the encoded information from the three bottle-neck layer neurons for all pre-selected and pre-processed MODIS swathsswath-wise for all MODIS co-locations. Subsequently, the FCM soft-clusters similar pixels per swath into 35 clusters before we manually categorize these clusters into one of the three classes 'cloud', 'sea ice', or 'open water/thin ice'. An exemplary sequence of this procedure is shown in Figure 3. a-d.

For this task of dimension reduction, we trained and subsequently used the encoder part of our autoencoder based on a setting featuring a decreasing number of neurons per hidden layer of 32, 16, and 8 down to the bottle-neck layer containing three neurons. The decoder part is built symmetrically to the encoder but in reversed order. We used a mean squared error loss function and trained for 50 epochs using a batch size of 2048 with the adam optimizer (Kingma and Ba, 2014) for all available co-located MODIS/S1-A/B combinations. For a detailed explanation of these technical terms, please see Goodfellow et al. (2016).

In order to reduce uncertainty in the training data, we constrained the manual classification to 'obvious' cases (e.g. 'cold' continuous patches over otherwise 'warm' polynyas and adjacent sea-ice categorized as 'clouds'), which results in not every MOD/MYD02 swath being fully classified at this stage . (Fig. 3d).

We used a subset (n=11) of all available co-located MODIS swaths (n=52) in order to i) keep the manual categorization computationally and humanly manageable and ii) represent rather easy to distinguish configurations of cloud, sea-ice and open-water/thin-ice pixels with favorable geometries (e. g., low sensor zenith angles). To increase the variability of the feature space in our labeled training data set, we selected the neighboring four swaths randomly from a pool of MOD/MYD02 swaths with a maximum time

230 ~~difference of ±24 to the current MODIS swath. However, we manually classified two combinations of the randomly selected temporal neighbors and applied these results to up to 25 random combinations for each of the 11 swaths. This increased the overall temporal variability of our training data set substantially.~~

Finally, from our manual categorization, we only use pixels with ~~a~~ an FCM probability (i.e., the membership score) above 0.6 for 'open-water/thin-ice' pixels, 0.65 for 'sea-ice' pixels as well as 0.65 for 'cloud' pixels ~~.~~ (Fig. 3d). As 'sea-ice'/'cloud' pixels are harder to identify, we chose a stricter probability threshold for those two classes. Due to the large temperature
235 range present in Antarctic clouds, we arbitrarily separated our 'cloud' class internally into 'cold' ($<235\,$K), 'intermediate', and 'warm' ($>250\,$K) clouds. This separation lead to an improved general classification result ~~for the supervised learning algorithms~~through the neural network later on. All ice-shelf areas are excluded from our analysis to avoid any additional misclassifications due to the substantially different temperature regime.

Through this procedure, we created an initial labeled training data set consisting of about ~~8.2~~3.5×10^6 data points ~~.~~

240 **2.4** Training of classifier candidates

~~While some general rules of thumb exist for the setup of NN and RF, the process of finding the best setup is iteratively based on training with varying setups and subsequent assessing of the resulting accuracy based on validation data. Therefore, before training any classifier, we separated our 8.2× labeled training data points into a~~ for the 33 predictors (Tab. 1). For the purpose of training the NN, we divided the data into a training~~/~~calibration and a validation ~~part using randomly selected 75/25 of the total data points, respectively~~ data set (Fig. 2~~)~~.

245 ~~For classifier-candidate training, we use the following variations for the setup of our NN: the activation functions (ReLU~~B). As a random split would potentially lead to highly auto-correlated neighboring pixels, we decided for a swath-wise split with 16 swaths used for training/~~TanH; either constant for all hidden layers in a NN or combinations of them), the number of hidden layers (2-4) as well as the number of neurons (10, 20 or 25; either constant for all layers or decreasing per layer) in each hidden layer, and the number of trainingepochs (10-50). For the setup of the RF classifiers we change:~~

  – ~~the number of trees in each RF (25-40); while at the same time~~

250 – ~~using a constant m of 3 for selection of split parameters during decision-tree growth.~~

~~For both classifiers, we varied and limited the input parameters for both classifier types~~ calibration and six swaths used for validation plus additional two swaths for an additional analysis (Tab. ~~1~~2).

2.4 ~~Validation and final~~ Final **training data generation**

~~From assessing the classification accuracy on the validation data set as well as visual screening it became apparent that all of the classifier candidates showed biases towards either~~
255 ~~one or several of 'open-water/thin-ice' , 'sea-ice' or 'cloud' classes. This likely results from either a higher degree of complexity in the data that can not represented in the current classifier setups or a lack of representation of the overall variability in the data.~~

As mentioned, the initial training data set is based solely on 'obvious' cases that were manually categorized. This procedure lead to only few data points per swath (Fig. 3d). In order to ~~increase~~ (at least almost) fully classify all co-located MODIS swaths and thereby extend our training data ~~sample size to fully represent more different configurations of 'cloud', 'sea-ice', and 'open-water~~set,
260 two simple intermediate classifiers were trained to represent their respective initial training data set (i.e., calibration/~~thin-ice'~~

pixels, we selected eight classifier candidates with the highest accuracy (5×NN, 3×RF) and used them to classify all of our co-located MODIS swaths (n= 52). Here, we randomly selected three temporal neighbor combinations for each of the 52 swaths. validation) as best as possible (Fig. 2C).

In this way, With this, we are able to extend our training data set by identifying and classifying additional similar data points in the complete set of collocated MODIS swaths, that were previously not categorized. However, based on the class probabilities provided by the two NN and through visual screening, we excluded ambiguous pixels from the final training data set (Fig. 3e). In this way, we increased our total variability in the training data (while keeping it computationally manageable), as well as get a statistically substantiated classification of almost the complete swaths - in contrast to the partially categorized swaths through manual classification used before . Furthermore, we reduced the number of input predictors to 45 by removing consistently less important predictors through using the RF importance metric (Tab. 1 (Fig. 2C).

An ensemble combination of the best classifier results was used for the final training data, based on the probability estimates provided by the RF/NN classifiers as well as again manual visual screening.

All RF classifiers featured frequent rather close decisions in their resulting classifications, i.e. mid range probabilities. This likely originated from comparably few trees in the RF classifiers. However, larger numbers of tree could not be realized with the given computing power.

Through this procedure, we created our final labeled training data set of about 17.0 10.0× 10^6 and 3.1× 10^6 data points comprising 45 the 33 different predictors/parameters for calibration and validation, respectively (Tab. 1).

2.5 Training of the final classifier

Normalized confusion matrix for the final deep NN classifier between reference (R; columns) and prediction (P; rows) of the NN for the five classes of open-water/thin-ice pixels (OWT), sea-ice pixels (ICE) as well as the three temperature-dependent cloud sub-classes (CLD); (w)arm, (i)ntermediate, and (c)old. ↓P/R→ OWT CLD_{cold} CLD_{intm} CLD_{warm} ICE OWT **0.95** 0.00 0.00 0.01 0.01 CLD_c 0.00 **0.93** 0.00 0.00 0.00 CLD_i 0.00 0.05 **0.95** 0.02 0.04 CLD_w 0.02 0.00 0.01 **0.91** 0.03 ICE 0.03 0.02 0.04 0.06 **0.92**

This We used this final training data set , we again randomly split into calibration/validation data using 75/25 of the total data points, respectively to train our final classifier (Fig. 2).

As we could not increase the number of trees in the RF classifiers due to computing limitations, we settled for a NN as our final classifier type. All final classifier candidates are using ReLUactivation functions only, logarithmic loss function, a D). This NN consists of six hidden layers containing 20 neurons each with leaky rectified linear unit (leaky ReLU) activation functions, while using a fixed batch size of 100, and a total of 15 training epochs.

The only parameter we change for assessing the accuracy between different classifier candidates is the number of fully-coupled hidden layers (4-8), with each consisting 2048, a learning rate of 1×10^{-4}, a dropout rate of 20neurons per hidden layer.

The best result was achieved with a setup of six hidden layers (Table **??**) , featuring an overall accuracy of 93.2 % , i.e. as well as L2 regularization (Goodfellow et al., 2016). Furthermore, we used categorical cross-entropy loss and again the adam optimizer (Kingma and Ba, 2014).

Our final open-water/thin-ice, sea-ice, cloud discrimination (OSCD) classifier features an accuracy (the ratio of correctly classified pixels (3,961,579) to the total number of samples(4,252,193) in the) of 90.8 %/84.3 % on the calibration/validation data set, respectively. For our comparisons and the results, we always merged all cloud sub-classes to a single cloud class . (Fig. 2E and 3f).

Table 3. Summary of polynya area (PA; in km^2) estimates between S1-A/B (PA_{S1}), OSCD ($\text{PA}_{OSC\,OSCD}$), and MOD/MYD29 (PA_{M29}) data. PA estimates in parenthesis correspond to the PA retrieved from MODIS for the S1-A/B polygon in Figures 4/5.

Example	PA_{S1}	$\text{PA}_{OSC\,OSCD}$	PA_{M29}
Fig. 4a-d	~~4235~~ 903	6472 (4074~~106 (0)~~	5134 (1823~~714 (0)~~
Fig. 4e-h	2224	2586 (2129~~5620 (2136)~~	3146 (2131~~6897 (2122)~~
Fig. 4i-l	380	484 (366~~989 (355)~~	47 (18~~43 (16)~~
Fig. 5a-d	1093	1037 (880~~892 (601)~~	962 (920~~841 (577)~~
Fig. 5e-h	1448	1314 (1220~~1945 (858)~~	730 (722~~3366 (534)~~
Fig. 5i-l	1425	2172 (1414~~1748 (1407)~~	1019 (257~~1502 (245)~~

3 Results and Discussion

295   In the following, we describe and discuss the results from using our open-water/thin-ice, sea-ice, cloud discrimination (OSCD) product in comparison to the reference MOD/MYD29 sea-ice product on the basis of a thin-ice thickness (TIT) estimates on i) a swath-to-swath basis, ii) on the basis of daily composites of all available swaths per day, and iii) a comparison of overall achieved coverage over a year (Fig. 2F).

3.1   Swath-based comparison

300   In the following, we describe and discuss the results using the final NN classifier for our open-water/thin-ice, sea-ice, cloud discrimination (OSCD) product in comparison to the reference MOD/MYD29 sea-ice product.

Representative comparisons between resulting thin-ice thickness (TIT) TIT from OSCD and MOD/MYD29 swaths reveal sub- stantial substantial differences, especially in the high-temperature polynya and thin-ice areas (PA; Figs. 4&5)).

The S1-A/B reference data always feature a polynya signal in all our examples (Figs. 4a,e,i&5a,e,i) and these are (at least

305 partially) represented by a warm IST anomaly in the MODIS data (Figs. 4b,f,j&5b,f,j). While for some examples the difference in resulting TIT between OSCD and MOD/MYD29 is comparably small or negliable (Figs. 4g/h & 5c/d), substantial differences appear for other examples (Figs. 4c,k/4d,l & 5g,k/5h,l).

For a better comparison, the polynyas were hand-picked for the respective S1-A/B and MODIS data in Figures 4&5. The corresponding absolute polynya areas are summarized in Table 3. In addition to the corresponding respective numbers for each

310 polynya, the corresponding area covered in the S1-A/B extent is given in parenthesis. While there is a bit of some uncertainty due to the different grid resolution resolutions (25 m vs. 1 km) as well as acquisition-time difference and subsequent changes due to sea-ice drift, this allows for a good quantification of the impact of misclassified cloud cover erroneously classified cloud cover on the estimated TIT.

While there are correct and also corresponding cloud classifications in both MODIS products, the applied MODIS cloud

315 mask in the MOD/MYD29 product tends towards additionally masking out strong positive temperature anomalies (Figs. 4d
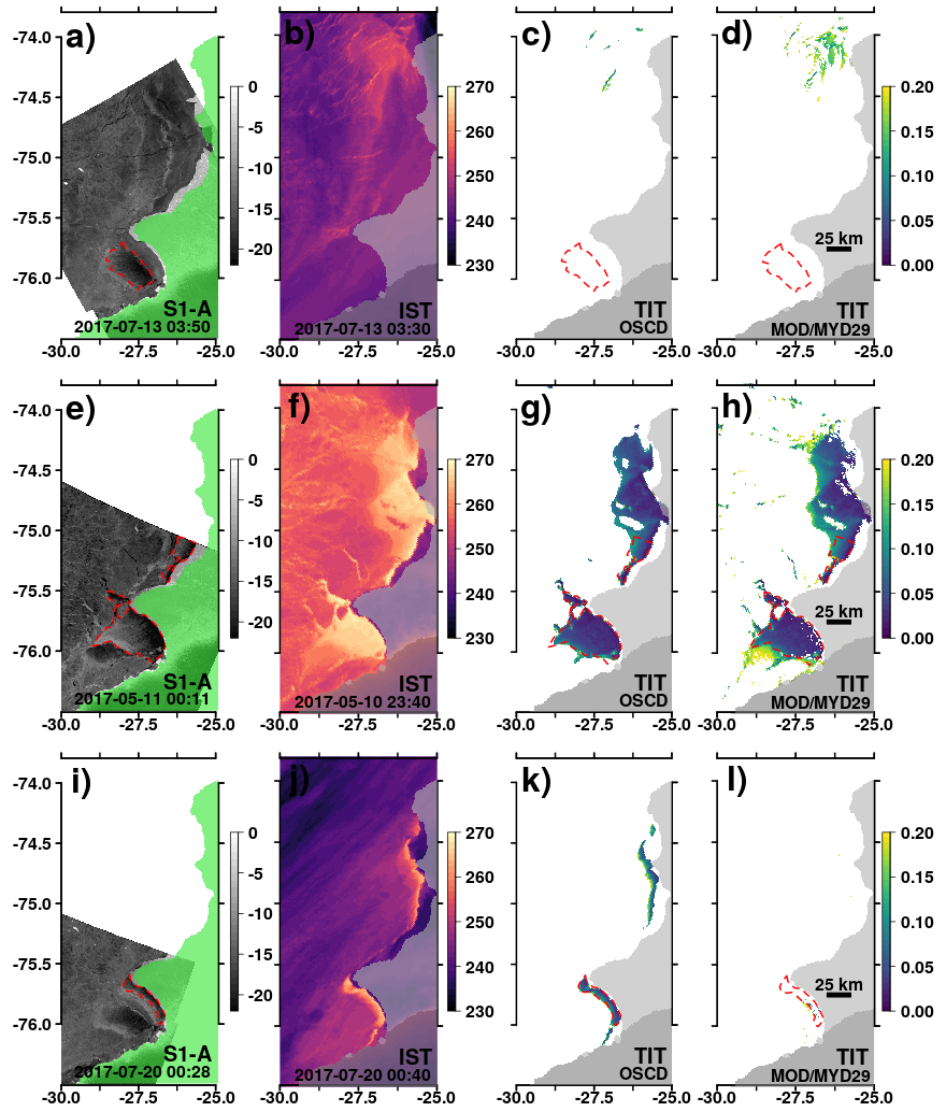
Figure 4. Compilation of exemplary co-located S1-A/B calibrated backscatter (in dB) and MODIS swaths of ice-surface temperature (IST; in K) and derived thin-ice thickness (TIT; in m) data (Tab. 3). Gray/green overlays highlight the ice-shelf extent. Manually picked S1-A/B reference polynya extent is outlined by a dashed red line in all panels.

l&5h,l). This happens frequently in the center of the primary polynya around 27.4 °W and 76 °S and leads to substantial differences in PA estimates (Tab. 3).

Due to the strong temperature gradient between the warm ocean and the cold atmosphere, turbulent exchange of sensible and latent heat is large and can potentially lead to the formation of sea fog and thin, low cloud cover (Gultepe et al., 2003).
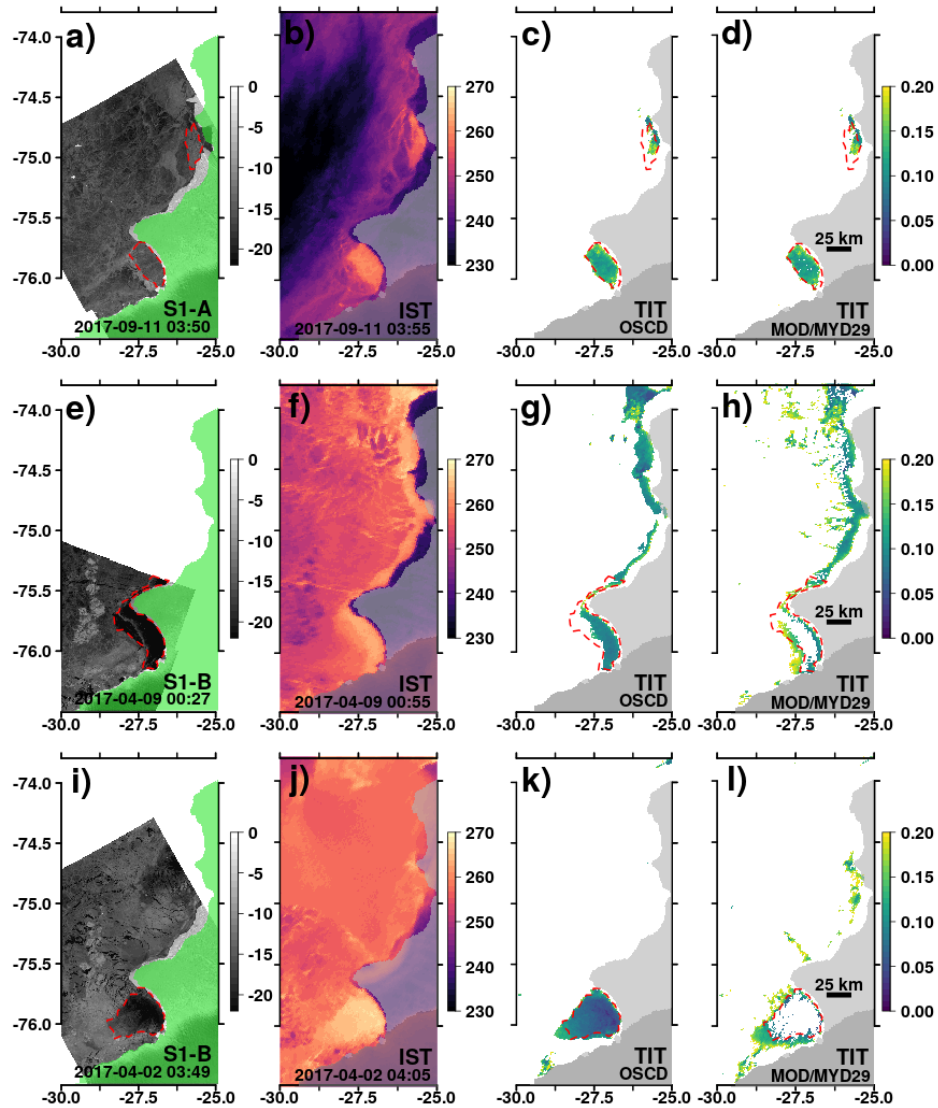
Figure 5. Additional compilation of exemplary co-located S1-A/B and MODIS swaths in the same setup as Figure 4. Gray/green overlays highlight the ice-shelf extent. Manually picked S1-A/B reference polynya extent is outlined by a dashed red line in all panels.

(Gultepe et al., 2003; Fraser et al., 2009). However, the temperature texture in the open-water/thin-ice areas appear to be homogeneous, and is likely not to be affected by either sea fog or clouds to the extent suggested by the MOD/MYD29 product through the MODIS cloud mask.
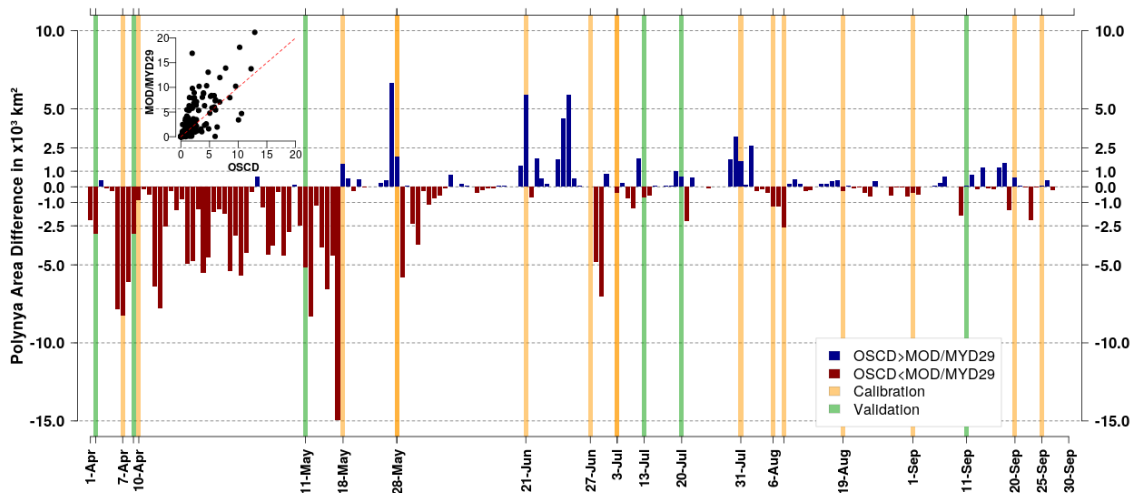
Figure 6. Daily polynya area difference in $\times 10^3 \mathrm{km}^2$ using swath-wise pixel averages featuring a thin-ice thickness (TIT) $\leq 0.2\,\mathrm{m}$ between OSCD and MOD/MYD29. Difference is calculated by subtracting MOD/MYD29 from OSCD; results with OSCD$\geq$MOD/MYD29 are shown in blue; results with OSCD$<$MOD/MYD29 in red. Orange vertical bars highlight days with S1-A/B swath coverage used for calibration/training of the OSCD algorithm. Green vertical bars show additional S1-A/B swaths used for validation between products (Figures 4/5). The top-left corner features a scatterplot of the daily polynya are with MOD/MYD29 against OSCD. Additional information about the S1-A/B swaths is provided in Table 2.

3.2 Daily-composite-based comparison

Based on median TIT of all available MODIS swaths per day, daily polynya area (PA) was ~~calculated~~computed (Paul et al.,
325 2015), and the difference between OSCD and MOD/MYD29 ~~estimated~~ calculated (i.e. OSCD minus MOD/MYD29; Fig. 6).

~~Larger estimates~~ Scattering of OSCD and MOD/MYD29 daily PA estimates against each other reveals a general tendency towards larger PA estimates in MOD/MYD29 ~~appear to dominate~~ data (Fig. 6; top-right Scatterplot inlet). However, there is also a strong seasonality in this MOD/MYD29 bias, which dominates from 1 April 2017 to mid May 2017, while OSCD estimates are ~~in general~~ predominately larger or equal to MOD/MYD29 between mid May and 30 September ~~2017.~~ 2017 (Fig. 6). For the year
330 2017, about ~~72.3~~64 %/~~56.0~~50.0 %/~~26.6~~27 % of the absolute daily median PA differences are below $1000\,\mathrm{km}^2/500\,\mathrm{km}^2/100\,\mathrm{km}^2$, respectively.

On average, OSCD estimates the daily polynya area (PA) between 1 April and 30 September 2017 to be ~~2.79~~1.88 $\times 10^3\ \mathrm{km}^2$ in contrast to ~~2.75~~2.69 $\times 10^3\ \mathrm{km}^2$ using MOD/MYD29 data (not shown). This corresponds to an average of ~~2~~about 44 % smaller daily mean PA for OSCD compared to MOD/MYD29. ~~From our comparisons based on swath data and the sometimes substantial differences we have~~
335 ~~found (Figs. 4&5), this appears to be a very small difference. An explanation could be that the composites counterbalance the individual shortcomings in the swath data from~~
~~misclassification of thin ice as cloud cover in the MOD/MYD29 product.~~
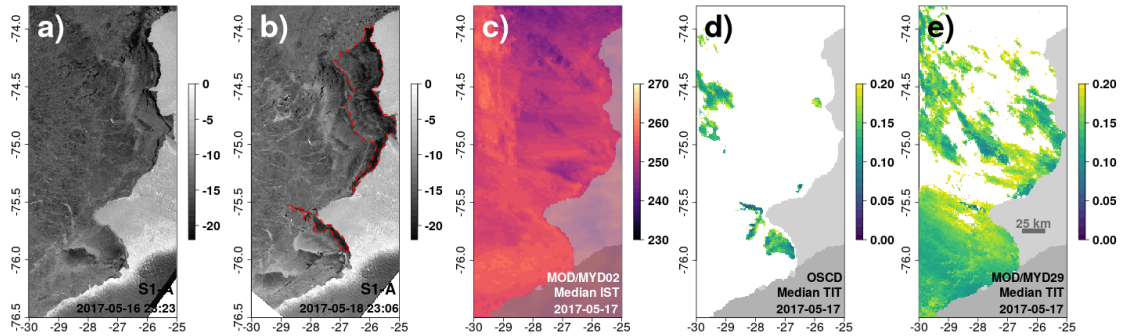
Figure 7. Compilation of S1-A swaths acquired at 16/18 May 2017 (a/b; calibrated backscatter in dB), the daily median ice-surface temperature (IST; in KK) composite of 17 May 2017 from all available MODIS swaths (c), as well as the resulting daily median thin-ice thickness (TIT; in mm) composites for the OSCD (d) and MOD/MYD29 (e) products for 17 May 2017, respectively. Red dashed line outlines the polynya on 18 May 2017 in S1-A.

However, especially during freeze-up (i.e. between 1 April 2017 and mid May 2017), the differences are oftentimes very large (16.614.9 $\times 10^3$ km^2 on 17 May 2017) and towards MOD/MYD29. To analyze this, we want to conduct a more detailed analysis of OSCD and MOD/MYD29 daily median TIT (Figs. 7&8).

340   Unfortunately, no S1-A/B swath was acquired over the BIS area for 17 May 2017. However, S1-A/B swaths were acquired the day before and after (Tab. 2and Fig. 6).

From the S1-A/B data (Fig. 7a/b), the existence of open water and/or thin ice very close to the ice-shelf edge around 27.4 °W and 76 °S for 18 May 2017 is evident.

The lack of any distinguishable clearly distinguishable positive temperature anomaly features in the MODIS daily median IST
345   composite (Fig. 7c) as well as the general texture of rather smooth temperature gradients patches are both signs for a persistently present cloud cover during 17 May 2017.

However, the relatively high temperatures of some of these potential clouds lead to an erroneous calculation of TIT and subsequent daily median TIT composite with an erroneously much larger polynya area (PA) for MOD/MYD29 compared to OSCD (Fig. 7d/e). Nonetheless, also OSCD features TIT estimates from cloud artifacts in the NW around 29.5 °W and
350   74-74.5 °S as well as in the area of the primary BIS polynya.

The individual swaths used for the computation of the both composites underline the absence of any pronounced positive temperature anomalies corresponding to open-water/thin-ice features (Fig. 8a-ha-g).

The While cold clouds are reliably identified, the inability of the MODIS cloud mask to reliably identify these also reliably identify warm cloud patterns results in the computation of TIT in large patches West of BIS (Fig. 8q-xq-u). Conversely, these false
355   computations are not present or at least much reduced in the OSCD data (Fig. 8i-pj-n). However, while a small area West of the tip of the BIS around 28 °W and 75.5 °S corresponds well to the polynya signal in the S1-A data (Fig. 7b), the majority of TIT estimates appear to be cloud artifacts (Fig. 8pn).
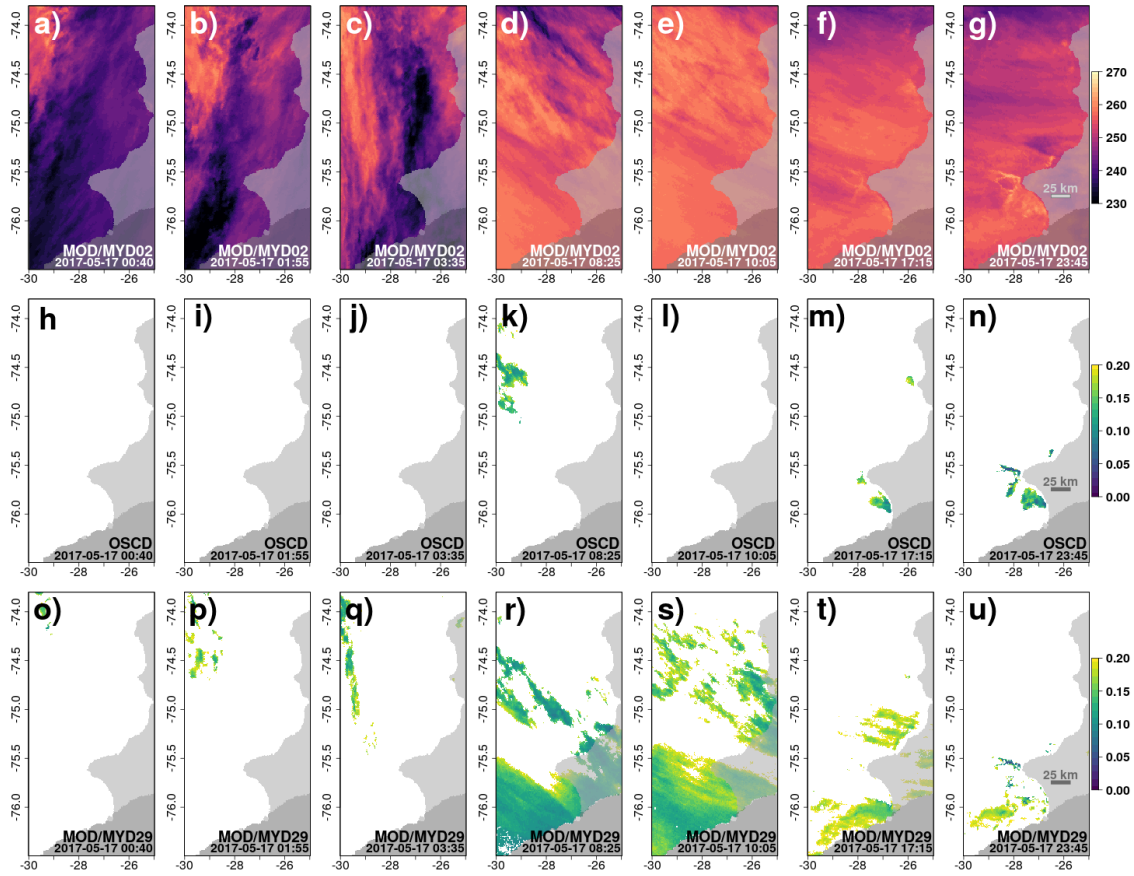
Figure 8. Compilation of MODIS swaths used for the computation of the data shown in Figure 7: swath-based ice-surface temperature (IST in ~~K~~K; ~~a-h~~a-g), resulting swath-wise thin-ice thickness (TIT in ~~m~~m) using OSCD (~~i-p~~h-n) and MOD/MYD29 (~~q-x~~o-u) data, respectively. Due to space limitations, we excludes a total of four MODIS swaths in this Figure that were used in the daily computations.

~~Two reasons can, therefore, explain the good overall agreement in daily median PA values~~ From our analysis of the swath-based and daily-composite comparisons, three major take home messages can be summarized:

360   1. Erroneous TIT estimates due to (especially) warm cloud-cover artifacts resulting from false-negative classifications in the MOD/MYD29 data increase the overall estimated PA ~~; and~~ substantially. These false-negative classifications are reduced in the OSCD data.

2. ~~Over the course of one day, the frequent MODIS coverage as well as the rapid growth rate of very thin ice potentially leads to rather sufficient coverage in the~~ False-positive clouds classifications over positive ice-surface temperature anomalies in the MOD/MYD29 ~~product~~ data re-

365   duce the products capability to estimate PA spatially and temporarily correct. These false-positive classifications are also reduced in the OSCD data.
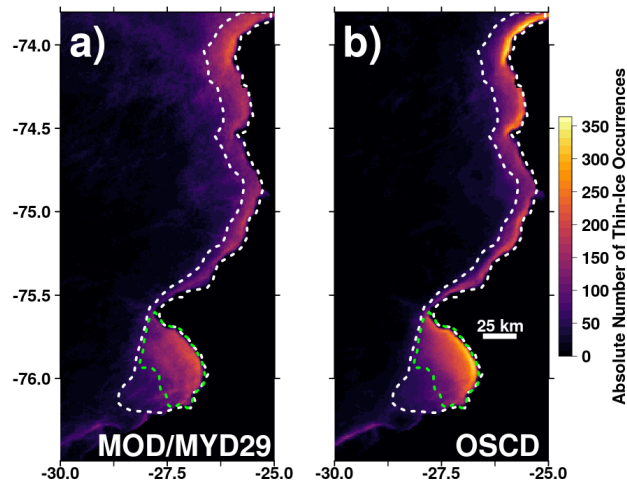
Figure 9. Comparison of per-pixel thin-ice occurrence based on all available swaths from 1 April to 30 September 2017 between the use of MOD/MYD29 (a) and OSCD data (b), respectively. White/Green dashed ~~line marks~~ lines mark the core ~~polynya~~ BIS region as well as the primary BIS polynya area, respectively, used for ~~analysis in the text~~further analyses.

3. Eliminating the thinnest sea ice fraction of the thin-ice spectrum due to false-positive classifications potentially leads to a 'thick' thin-ice bias during the daily composite procedure.

~~This~~ The combined effect leads to spatially misplaced TIT estimates, likely not resolving the correct shape and (sub-)daily thickness distribution of the open-water/thin-ice areas. Studies such as Paul et al. (2015) and Preußer et al. (2019), therefore, try to mitigate the effect of points 1&2 by introducing pre-defined masks.

3.3 Coverage comparison

~~While the annual average daily PA appears to be consistent between OSCD and MOD/MYD29 data,~~ In order to pick up on the last point, we would like to analyze the per-swath coverage ~~of thin ice~~ in more detail as this also influences the sub-daily TIT distribution and, therefore, the thickness distribution of the resulting daily composite. It appears that the per-swath thin-ice occurrence frequency is much higher in the OSCD data compared to the MOD/MYD29 data (Fig. 9).

~~Especially in the very active part of the polynya with frequent thin-ice occurrences around 27~~Quantifying the differences in the outlined sub regions (Fig 9; white/green dashed outlines), results in a 10 ~~and 76~~%/20 ~~, OSCD produces a much~~ % (BIS area/primary BIS polynya) higher detection-rate of thin-ice pixels over all MODIS swaths between 1 April 2017 and 30 September 2017 in the OSCD data (Fig 9b).

~~For the core polynya region (Green dashed outline ;~~ This improved coverage likely leads to a higher quality daily composite, as the impact from outliers is reduced. It admittedly sounds counter-intuitively at first to have improved coverage (Fig. 9) with at the same time substantially less average PA (Fig. 6). This effect can be explained from the difference between swath and daily-composite data. Here, the increase in coverage mainly focuses around the primary polynya at BIS (green outline in
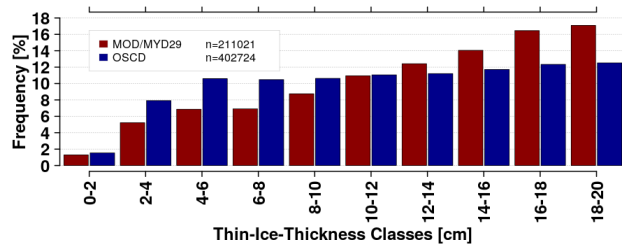
Figure 10. Comparison of binned thin-ice-thickness classes with a bin size of 2 cm based on all available swaths from 1 April to 30 September 2017 between the use of MOD/MYD29 (red) and OSCD (blue) data (b) for the primary BIS polynya (blue dashed outline in Fig. 9).

385 Fig. 9), use of OSCD results in a substantial 31. However, the substantial decrease in daily PA results from reducing the false-negative classifications of warm clouds as sea ice, primarily off the BIS edge to the West. These misclassification-related TIT estimates push the resulting average PA for the MOD/MYD29 data.

Based on our analysis in Subsections 3.1 and 3.2, we drew assumption that these additional thin-ice occurrences likely feature very thin ice, therefore, reducing the a potential 'thick' thin sea-ice bias in the MOD/MYD29 data. This is evident
390 from Figure 10. Here, the TIT occurrence frequency based on 2 increase in coverage cm bins for all available swaths between 1 April 2017 and 30 September 2017. The improved coverage likely leads to a higher quality daily composite,as the impact from outliers is reduced. This will likely play 2017 are shown for MOD/MYD29 and OSCD data. Thickness classes between 0 cm and 20 cm are much more frequent in OSCD data (402,724; Fig. 10) compared to the MOD/MYD29 standard product (211,021; Fig. 10). The largest difference between both products, however, is the overall higher occurrence frequency of the thinnest ice fractions (between 0 cm
395 and 10 cm) in the OSCD data compared to MOD/MYD29. As assumed before, there is a 'thick' thin-ice bias present in the MOD/MYD29 data, which potentially plays an important role especially in the estimation of sea-ice production based on daily composites.

Despite great care during the manual categorization, uncertainty remains due to the lack of measured ground-truth data for the training-data generation.
400 However, the underlying statistical basis from the unsupervised FCM clustering in combination with a second stage of fully classifying all co-located MODIS swaths using RF/NN NN before generating the calibration/validation swath-split final training data for the OSCD algorithm appears to provide a realistic representation of the present sea-ice conditions in the BIS area.

4 Summary and Outlook

In this study, we present a novel approach to improve the detection of wintertime cloud-cover over Antarctic sea ice and their its
405 discrimination from sea-ice cover and open-water/thin-ice areas in MODIS thermal-infrared data using a deep neural network.

We established a labeled training data set using dimension reduction, unsupervised clustering, and supervised learning techniques in combination with manual visual screening and categorization. Through this effort, we generated a total of ~~17.0×~~13.1 $\times 10^6$ data points for ~~45~~33 different predictors.

With this data set, we trained a deep neural network and used it to discriminate between open-water/thin-ice, sea-ice and
410 cloud-covered areas in the Brunt Ice Shelf region for the freezing period of 2017 (1 April to 30 September). Here, we computed the thin-ice thickness up to $0.2\,\mathrm{m}$ of open-water/thin-ice areas and evaluated the difference in daily polynya area and daily swath coverage to results using the standard NSIDC MOD/MYD29 sea-ice product.

Based on our approach, we obtain a ~~2~~44 % ~~higher~~ lower average polynya area but ~~31~~20 % higher swath coverage rate ~~. However~~compared to the standard MOD/MYD29 product. On the one hand, the polynya area in MOD/MYD29 is likely dominated
415 through frequent ~~misclassifications~~ false-negative classifications of warm clouds as thin ice, that lead to unrealistically large open-water/thin-ice areas, especially during freeze-up. ~~The much higher~~ On the other hand, the much lower coverage rate likely ~~increases~~ decreases the quality and accuracy of TIT estimates in the daily median TIT composites when using MOD/MYD29 data. Both factors are reduced in our OSCD data. This also reduces the impact of single outliers on the daily median TIT composites and, therefore, also increases the quality of derived information such as sea-ice production.

420 In the future, we plan to create an open-access comprehensive OSCD-based IST/TIT product covering all major Antarctic coastal polynyas, as well as providing higher-level parameters such as polynya area, sea-ice production, and associated ocean salt flux. We expect this data set to be of great use to ocean/sea-ice/ice-shelf model community as well as for potential biological applications.

Data availability. The generated training data set will be made available through PANGAEA. Sources of all used data sets are referenced in
425 the text.

Author contributions. SP designed the study/methodology, conducted the analysis, and drafted the original manuscript. MH assisted in the study design and the adaptation of the machine learning algorithms as well as with writing the manuscript.

Competing interests. The authors declare no conflict of interests.

Acknowledgements. The authors want to thank the LAADS DAAC and the ASF DAAC for the provision of the here used MOD/MYD02
430 and S1-A/B data. The corresponding author appreciates the help of his family in enabling him the time to finally write this manuscript during SARS-CoV-2 induced home office. The comments by two anonymous reviewers as well as editor Claude Duguay helped to substantially improve the quality of this manuscript.

References

Ackerman et al., S.: MODIS Atmosphere L2 Cloud Mask Product. NASA MODIS Adaptive Processing System., Goddard Space Flight Center, USA, https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD35_L2.006, 2015.

Adams, S., Willmes, S., Schröder, D., Heinemann, G., Bauer, M., and Krumpen, T.: Improvement and Sensitivity Analysis of Thermal Thin-Ice Thickness Retrievals, Geoscience and Remote Sensing, IEEE Transactions on, 51, 3306–3318, 2013.

Allaire, J. and Chollet, F.: keras: R Interface to 'Keras', https://CRAN.R-project.org/package=keras, r package version 2.3.0.0, 2020.

Atkinson, P. M. and Tatnall, A. R. L.: Introduction Neural networks in remote sensing, International Journal of Remote Sensing, 18, 699–709, https://doi.org/10.1080/014311697218700, https://doi.org/10.1080/014311697218700, 1997.

Aulicino, G., Sansiviero, M., Paul, S., Cesarano, C., Fusco, G., Wadhams, P., and Budillon, G.: A New Approach for Monitoring the Terra Nova Bay Polynya through MODIS Ice Surface Temperature Imagery and Its Validation during 2010 and 2011 Winter Seasons, Remote Sensing, 10, 366, https://doi.org/10.3390/rs10030366, http://dx.doi.org/10.3390/rs10030366, 2018.

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences, 10, 191 – 203, https://doi.org/https://doi.org/10.1016/0098-3004(84)90020-7, http://www.sciencedirect.com/science/article/pii/0098300484900207, 1984.

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

Cao, W., Wang, X., Ming, Z., and Gao, J.: A Review on Neural Networks with Random Weights, Neurocomput., 275, 278–287, https://doi.org/10.1016/j.neucom.2017.08.040, https://doi.org/10.1016/j.neucom.2017.08.040, 2018.

Charbonnier, P., Blanc-Feraud, L., Aubert, G., and Barlaud, M.: Deterministic edge-preserving regularization in computed imaging, IEEE Transactions on Image Processing, 6, 298–311, 1997.

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, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dong, G., Liao, G., Liu, H., and Kuang, G.: A Review of the Autoencoder and Its Variants: A Comparative Perspective from Target Recognition in Synthetic-Aperture Radar Images, IEEE Geoscience and Remote Sensing Magazine, 6, 44–68, 2018.

Drucker, R., Martin, S., and Moritz, R.: Observations of ice thickness and frazil ice in the St. Lawrence Island polynya from satellite imagery, upward looking sonar, and salinity/temperature moorings, J. Geophys. Res., 108, 3149, https://doi.org/10.1029/2001JC001213, 2003.

Dunn, J. C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, Journal of Cybernetics, 3, 32–57, https://doi.org/10.1080/01969727308546046, https://doi.org/10.1080/01969727308546046, 1973.

Fraser, A. D., Massom, R. A., and Michael, K. J.: A Method for Compositing Polar MODIS Satellite Images to Remove Cloud Cover for Landfast Sea-Ice Detection, IEEE Transactions on Geoscience and Remote Sensing, 47, 3272–3282, https://doi.org/10.1109/TGRS.2009.2019726, 2009.

Fraser, A. D., Massom, R. A., and Michael, K. J.: Generation of high-resolution East Antarctic landfast sea-ice maps from cloud-free MODIS satellite composite imagery, Remote Sensing of Environment, 114, 2888–2896, http://www.sciencedirect.com/science/article/pii/S0034425710002221, 2010.

470 Fraser, A. D., Massom, R. A., Ohshima, K. I., Willmes, S., Kappes, P. J., Cartwright, J., and Porter-Smith, R.: High-resolution mapping of circum-Antarctic landfast sea ice distribution, 2000–2018, Earth System Science Data, 12, 2987–2999, https://doi.org/10.5194/essd-12-2987-2020, https://essd.copernicus.org/articles/12/2987/2020/, 2020.

Frey, R. A., Ackerman, S. A., Liu, Y., Strabala, K. I., Zhang, H., Key, J. R., and Wang, X.: Cloud Detection with MODIS. Part I: Improvements in the MODIS Cloud Mask for Collection 5, J. Atmos. Oceanic Technol., 25, 1057–1072, https://doi.org/10.1175/2008JTECHA1052.1, 475 2008.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Gultepe, I., Isaac, G. A., Williams, A., Marcotte, D., and Strawbridge, K. B.: Turbulent heat fluxes over leads and polynyas, and their effects on arctic clouds during FIRE.ACE: Aircraft observations for April 1998, Atmosphere-Ocean, 41, 15–34, https://doi.org/10.3137/ao.410102, https://doi.org/10.3137/ao.410102, 2003.

480 Hall, D., Key, J., Casey, K., Riggs, G., and Cavalieri, D.: Sea ice surface temperature product from MODIS, Geoscience and Remote Sensing, IEEE Transactions on, 42, 1076 – 1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Hall, D. K. and Riggs, G. A.: MODIS/Terra Sea Ice Extent 5-min L2 Swath 1km, Version 6, 2017, https://doi.org/10.5067/MODIS/MOD29.006, 2015a.

Hall, D. K. and Riggs, G. A.: MODIS/Aqua Sea Ice Extent 5-min L2 Swath 1km, Version 6, 2017, 485 https://doi.org/10.5067/MODIS/MYD29.006, 2015b.

Hall, D. K., Nghiem, S. V., Rigor, I. G., and Miller, J. A.: Uncertainties of Temperature Measurements on Snow-Covered Land and Sea Ice from In Situ and MODIS Data during BROMEX, J. Appl. Meteor. Climatol., 54, 966–978, https://doi.org/10.1175/JAMC-D-14-0175.1, http://dx.doi.org/10.1175/JAMC-D-14-0175.1, 2015.

Hall-Beyer, M.: Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spa-490 tial scales, International Journal of Remote Sensing, 38, 1312–1338, https://doi.org/10.1080/01431161.2016.1278314, https://doi.org/10.1080/01431161.2016.1278314, 2017.

Haralick, R. M.: Statistical and structural approaches to texture, Proceedings of the IEEE, 67, 786–804, 1979.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, 610–621, 1973.

495 Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, 28, 100–108, http://www.jstor.org/stable/2346830, 1979.

Holz, R. E., Ackerman, S. A., Nagle, F. W., Frey, R., Dutcher, S., Kuehn, R. E., Vaughan, M. A., and Baum, B.: Global Moderate Resolution Imaging Spectroradiometer MODIS cloud detection and height evaluation using CALIOP, J. Geophys. Res., 113, D00A19–, http://dx.doi.org/10.1029/2008JD009837, 2008.

500 Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kohonen, T.: An introduction to neural computing, Neural Networks, 1, 3 – 16, https://doi.org/https://doi.org/10.1016/0893-6080(88)90020-2, http://www.sciencedirect.com/science/article/pii/0893608088900202, 1988.

Lammers, B.: ANN2: Artificial Neural Networks for Anomaly Detection, https://CRAN.R-project.org/package=ANN2, r package version 2.3.2, 2019.

505 LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

Lee, J., Weger, R. C., Sengupta, S. K., and Welch, R. M.: A neural network approach to cloud classification, IEEE Transactions on Geoscience and Remote Sensing, 28, 846–855, 1990.

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2, 18–22, https://CRAN.R-project.org/doc/Rnews/, 2002.

Liu, Y. and Key, J. R.: Less winter cloud aids summer 2013 Arctic sea ice return from 2012 minimum, Environmental Research Letters, 9, 044 002–, https://doi.org/10.1088/1748-9326/9/4/044002, 2014.

Liu, Y., Key, J. R., Frey, R. A., Ackerman, S. A., and Menzel, W.: Nighttime polar cloud detection with MODIS, Remote Sensing of Environment, 92, 181–194, https://doi.org/10.1016/j.rse.2004.06.004, 2004.

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, https://tc.copernicus.org/articles/13/2051/2019/, 2019.

MacQueen, J.: Some methods for classification and analysis of multivariate observations, pp. –, 1967.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F.: e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, https://CRAN.R-project.org/package=e1071, r package version 1.7-2, 2019.

MODIS Characterization Support Team (MCST): MODIS 1km Calibrated Radiances Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, https://doi.org/http://dx.doi.org/10.5067/MODIS/MYD021KM.06, 2017a.

MODIS Characterization Support Team (MCST): MODIS 1km Calibrated Radiances Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center, USA, https://doi.org/http://dx.doi.org/10.5067/MODIS/MYD021KM.06, 2017b.

Paul, S., Willmes, S., and Heinemann, G.: Long-term coastal-polynya dynamics in the southern Weddell Sea from MODIS thermal-infrared imagery, The Cryosphere, 9, 2027–2041, https://doi.org/10.5194/tc-9-2027-2015, 2015.

Preußer, A., Ohshima, K. I., Iwamoto, K., Willmes, S., and Heinemann, G.: Retrieval of Wintertime Sea Ice Production in Arctic Polynyas Using Thermal Infrared and Passive Microwave Remote Sensing Data, Journal of Geophysical Research: Oceans, 124, 5503–5528, https://doi.org/10.1029/2019JC014976, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JC014976, 2019.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

Reiser, F., Willmes, S., and Heinemann, G.: A New Algorithm for Daily Sea Ice Lead Identification in the Arctic and Antarctic Winter from Thermal-Infrared Satellite Imagery, Remote Sensing, 12, 1957, https://doi.org/10.3390/rs12121957, http://dx.doi.org/10.3390/rs12121957, 2020.

Riggs, G. and Hall, D.: MODIS Sea Ice Products User Guide to Collection 6, National Snow and Ice Data Center, University of Colorado, Boulder, CO 80309-0449 USA, 2015.

Schaffer, J., Timmermann, R., Arndt, J. E., Kristensen, S. S., Mayer, C., Morlighem, M., and Steinhage, D.: A global, high-resolution data set of ice sheet topography, cavity geometry, and ocean bathymetry, Earth System Science Data, 8, 543–557, https://doi.org/10.5194/essd-8-543-2016, https://www.earth-syst-sci-data.net/8/543/2016/, 2016.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85 – 117, https://doi.org/https://doi.org/10.1016/j.neunet.2014.09.003, http://www.sciencedirect.com/science/article/pii/S0893608014002135, 2015.

Toller, G., Xu, G., Kuyper, J., Isaacman, A., and Xiong, J.: MODIS Level 1B Product User's Guide, NASA/Goddard Space Flight Center, 2009.

Welch, R. M., Sengupta, S. K., Goroch, A. K., Rabindra, P., Rangaraj, N., and Navar, M. S.: Polar Cloud and Surface Classification Using AVHRR Imagery: An Intercomparison of Methods, Journal of Applied Meteorology (1988-2005), 31, 405–420, www.jstor.org/stable/26186465, 1992.

Yu, Y. and Rothrock, D. A.: Thin ice thickness from satellite thermal imagery, J. Geophys. Res., 101, 25 753–25 766, https://doi.org/10.1029/96JC02242, 1996.

Zvoleff, A.: glcm: Calculate Textures from Grey-Level Co-Occurrence Matrices (GLCMs), https://CRAN.R-project.org/package=glcm, r package version 1.6.4, 2019.