The Cryosphere

Classification of sea ice types in Sentinel-1 synthetic aperture radar images

Jeong-Won Park[1,2], Anton Andreevich Korosov[1], Mohamed Babiker[1], Joong-Sun Won[3], Morten Wergeland Hansen[1,4], and Hyun-Cheol Kim[2]

[1]Ocean and Sea Ice Remote Sensing Group, Nansen Environmental and Remote Sensing Center, 5006 Bergen, Norway
[2]Unit of Arctic Sea Ice Prediction, Korea Polar Research Institute, Incheon, 21990, South Korea
[3]Department of Earth System Sciences, Yonsei University, Seoul, 03722, South Korea
[4]Department of Remote Sensing and Data Management, Norwegian Meteorological Institute, 0371 Oslo, Norway

Correspondence: Jeong-Won Park (jeong-won.park@kopri.re.kr)

Abstract. A new Sentinel-1 image-based sea ice classification algorithm using a machine-learning-based model trained in a semi-automated manner is proposed to support daily ice charting. Previous studies mostly rely on manual work in selecting training and validation data. We show that the readily available ice charts from the operational ice services can reduce the amount of manual work in preparation of large amounts of training/testing data. Furthermore, they can feed highly reliable data to the trainer by indirectly exploiting the best ability of the sea ice experts working at the operational ice services. The proposed scheme has two phases: training and operational. Both phases start from the removal of thermal, scalloping, and textural noise from Sentinel-1 data and calculation of grey level co-occurrence matrix and Haralick texture features in a sliding window. In the training phase, the weekly ice charts are reprojected into the SAR image geometry. A random forest classifier is trained with the texture features on input and labels from the rasterized ice charts on output. Then, the trained classifier is directly applied to the texture features from Sentinel-1 images operationally. Test results from the two datasets spanning winter (January–March) and summer (June–August) seasons acquired over the Fram Strait and the Barents Sea showed that the classifier is capable of retrieving three generalized cover types (open water, mixed first-year ice, old ice) with overall accuracies of 87 % and 67 % in winter and summer seasons, respectively. For the summer season, the classifier failed in distinguishing mixed first-year ice from old ice with accuracy of only 12 %; however, it performed rather like an ice–water discriminator with

high accuracy of 98 % as the misclassification between the mixed first-year ice and old ice was between them. The accuracy for five cover types (open water, new ice, young ice, first-year ice, old ice) in the winter season was 60 %. The errors are attributed both to incorrect manual classification on the ice charts and to the semi-automated algorithm. Finally, we demonstrate the potential for near-real-time service of the ice map using daily mosaicked Sentinel-1 images.

1 Introduction

Wide swath SAR observation from several spaceborne SAR missions (RADARSAT-1, 1995–2013; Envisat ASAR, 2002–2012; ALOS-1 PALSAR, 2006–2011; RADARSAT-2, 2007-present; Sentinel-1, 2014–present) played an important role in studying global ocean and ice-covered polar regions. The Sentinel-1 constellation (1A and 1B) is producing dual-polarization observation data with the largest Arctic coverage and the highest temporal resolution ever. The cross-polarization is known to be more sensitive to the difference in scattering from sea ice and open water than the co-polarization (Scheuchl et al., 2004), and the combination of HH and HV polarizations has been widely used for ice edge detection and ice type classification (a nice overview is given in the paper by Zakhvatkina et al., 2019). However, most of the recent ice classification algorithms were developed using RADARSAT-2 ScanSAR (Leigh et al., 2014; Liu et al., 2015; Zakhvatkina et al., 2017), which has different sensor charac-
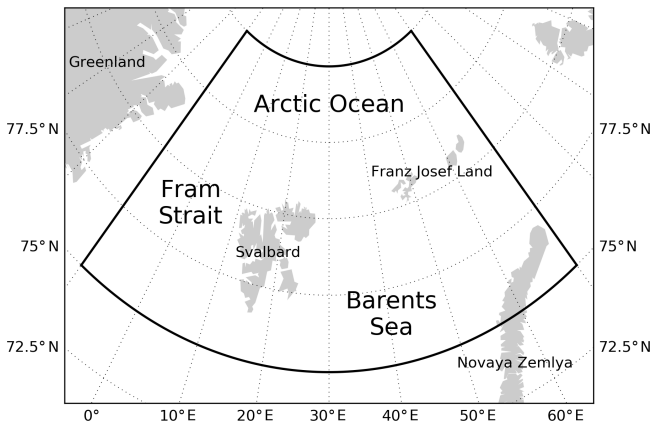
Figure 1. Study area.

Figure 2. Processing flow chart of the proposed algorithm. The grey colour shows the training phase

teristics from Sentinel-1 TOPSAR, and the use of Sentinel-1 for the same purpose is very limited in literature. The main drawback of applying existing algorithms to Sentinel-1 TOPSAR data is the relatively high level of thermal noise contamination and its propagation to image textures.

For proper use of dense time series of Earth observations using SAR sensors, radiometric properties must be well-calibrated. Thermal noise is often neglected in many cases but can seriously impact the utility of dual-polarization SAR data. Sentinel-1 TOPSAR image intensity is particularly disturbed by the thermal noise in the cross-polarization channel. Although the European Space Agency (ESA) provides calibrated noise vectors for noise power subtraction, residual noise contribution is still significant considering the relatively narrow backscattering distribution of the cross-polarization channel. In our previous study (Park et al. 2018), a new denoising method with azimuth de-scalloping, noise scaling, and inter-swath power balancing was developed and showed improved performance in various SAR intensity-based applications. Furthermore, when it came to texture-based image classification, we suggested a correction method for textural noise (Park et al., 2019) which distorts local statistics, and thus degrades texture information in the Sentinel-1 TOPSAR images.

In many of the previous studies on ice–water and/or sea ice classification (Soh and Tsatsoulis, 1999; Zakhvatkina et al., 2013; Leigh et al., 2014; Liu et al., 2015; Ressel et al., 2015; Zakhvatkina et al., 2017; Aldenhoff et al., 2018), the training and validation were done using manually produced ice maps. Although the authors claimed that the manual ice maps were drawn by ice experts, the selection of SAR scenes and interpretation could be inconsistent, and the number of samples might not be enough to generalize the results because of the laborious manual work. Furthermore, the results are hardly reproducible by others because the reference sources are not open to the public. Therefore, increasing objectivity is crucial, and automating the classification process is encouraged. The idea of training using SAR images and accompanying
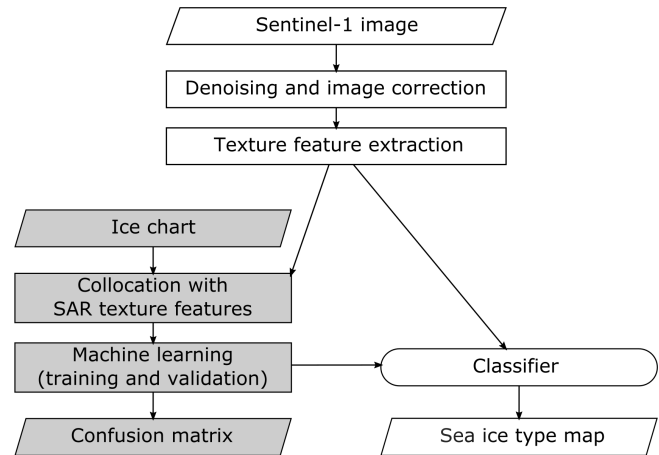
image analysis charts, which is a direct manual interpretation of SAR images by trained ice analysts working at operational ice services, was tested for sea ice concentration estimation by Wang et al. (2017); however, such image analysis charts are not accessible to the public.

The use of a public ice chart as training and validation reference data may help in solving the validation problem. The preparation of a public ice chart is also through manual inspection of various sources of satellite imagery and other sources of data (Partington et al., 2003; Johannessen et al., 2006); however, training using a large volume of these charts would reduce operator-to-operator bias, such as inconsistent decisions against similar ice conditions. The overall bias may exist since the public ice charts are produced in the interest of marine safety. Nevertheless, as the human interpretation available in the ice chart is currently considered the best available information on sea ice (Karvonen et al., 2015), the best practice to make a sea ice type classifier is to train with the public ice chart so that the best knowledge of ice analysts is mimicked.

In this work, we present a semi-automated Sentinel-1 image-based sea ice classification algorithm which takes advantage of our denoising method. The noise-corrected dual-polarization images are processed into image textures that capture sea ice features in various spatial scales, and they are used for supervised classification with a random forest classifier by relating with ice charts published by operational ice services. The use of ice charts has a dual purpose: semi-automatization of classifier training and minimization of human error.
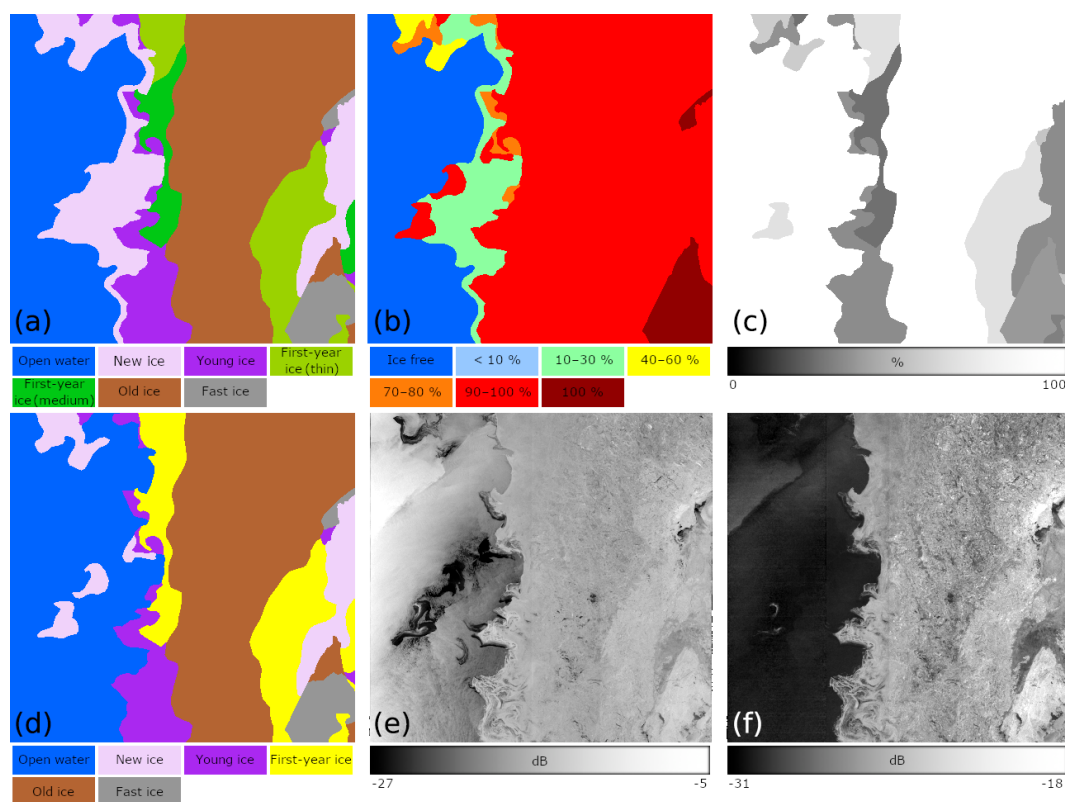
Figure 3. An example of the ice chart preprocessing. From the ice chart, stage of development (SoD; **a**), ice concentration (CT; **b**), and partial concentration of the dominant ice type (CP; **c**) maps are extracted. Then, some of the different SoDs are merged (e.g. thin and thick first-year ice is merged into a single label as first-year ice), and the area with low ice concentration is labelled as open water. The processed map of SoD (**d**) is related with textural features extracted from HH and HV polarization images (**e, f**). Note that the NIC ice chart which was published on 25 January 2018 and the Sentinel-1 product S1B_EW_GRDM_1SDH_20180122T075237_20180122T075337_009281_010A4D_65AA acquired over the Fram Strait were used in this example.

2 Data and methods

2.1 Study area and used data

The region of study for developing and testing the proposed algorithm is the Fram Strait and the Barents Sea including a part of the Arctic Ocean (75–85° N, 10° W–70° E) as shown in Fig. 1. Various sea ice types are found in this area due to the intensive export of multi-year ice through the Fram Strait (Smedsrud et al., 2017) and development of young and first-year ice between Svalbard and Franz Josef Land.

Sentinel-1 TOPSAR data in extended wide-swath (EW) mode acquired in summer (June–August in 2016–2018) and winter (January–March in 2017–2019) seasons were collected from the Copernicus Open Access Hub (https://scihub. copernicus.eu, last access: 18 August 2020). The number of daily image acquisitions covering the study area ranges from 6 to 10 depending on the orbits. The images from the first 2 years (hereafter called DS1) are used to train the classifier, and those from the third year (hereafter called DS2) are used for validation.

The ice charts covering the same periods were collected. There are two ice services that publish weekly ice charts with pan-Arctic coverage: the U.S. National Ice Center (NIC) of the United States of America and the Arctic and Antarctic Research Institute (AARI) of Russia. Although the accuracies are known to be comparable (Pastusiak, 2016) to each other, there is no partial ice concentration information in the AARI ice chart. In this study, we use the ice charts downloaded from the NIC website (https://www.natice.noaa.gov/ Main_Products.htm, last access: 18 August 2020). The NIC ice products are produced primarily using radar, microwave radiometer, scatterometer, visible, and infrared imagery from a variety of sources. In addition to imagery, drifting buoy data, ice model predictions, limited ship reports, meteorological and oceanographic observations, and ice information provided by other international centres are also used to make a comprehensive analysis of ice conditions (WMO, 2017).

2.2 Methods

Figure 2 shows the flow of the semi-automated ice classification scheme that we propose. It is divided into two phases:
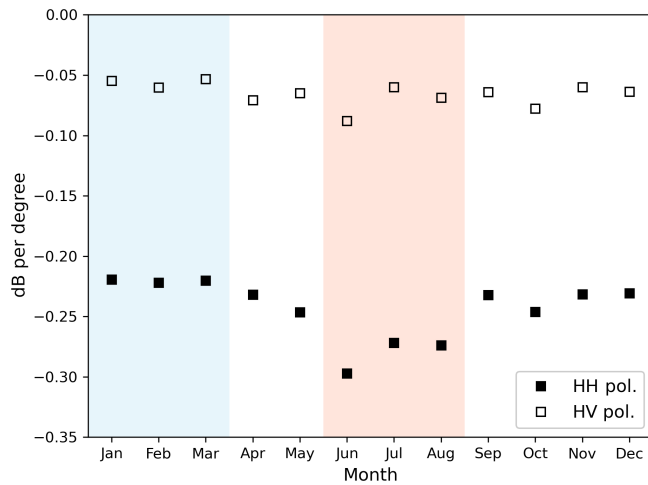
Figure 4. Incidence angle dependences of sigma naught in HH (closed squares) and HV (open squares) polarization channels. Pixels covering various types of sea ice were merged so that the averaged property can be estimated. The blue and red zones indicate winter and summer seasons, respectively.

Table 1. Hyperparameters used for grid search.

Parameters	Values						
N_T	1	2	4	8	16	32	64
D	1	2	4	8	16	32	64
N_F	1	2	4	8	16	28	

training and operational. Both phases start from the removal of thermal noise from Sentinel-1 data (Sect. 2.2.2), incidence angle calibration (Sect. 2.2.3), and calculation of texture features (Sect. 2.2.4). The training phase (shown in grey in Fig. 2) continues with preprocessing and collocation of the ice charts with the Sentinel-1 data (Sect. 2.2.1) and machine learning step (Sect. 2.2.5 and 2.2.6). The operational phase uses the classifier which developed during the training phase for processing texture features that were computed from the input SAR data and for generating ice charts. Detailed explanations for each step are given in the following subsections.

2.2.1 Ice chart preprocessing

To take advantage of the objective identification of the ice type from expert sources open to public and to develop a semi-automated processing scheme, the proposed algorithm uses electronic ice charts published by international ice chart services. The electronic ice chart follows SIGRID-3 format (JCOMM, 2014a), which is based on a vector format called shapefile (ESRI, 1998). The first step is to reproject the ice chart into the geometry of each SAR image. Although an accurate reprojection needs several pieces of information such as orbit, look angle, topographic height, etc., our interest is in the sea ice where the topographic difference does not exceed
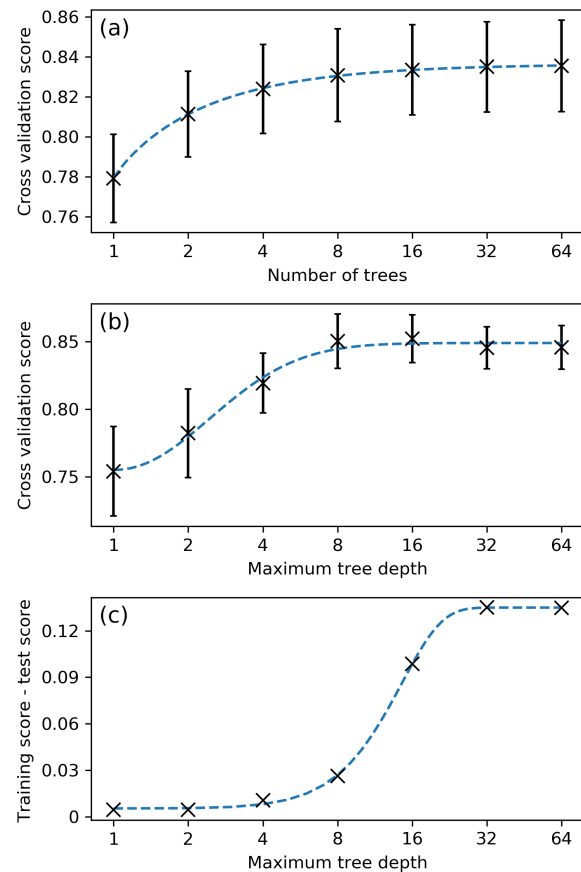

Figure 5. Hyperparameter optimization using grid search results (cross). Dashed lines represent the best-fit Richards curve. **(a)** The optimal values are extracted from the locations where the score increments per unit of each hyperparameter become lower than a threshold (e.g. 0.001). **(b)** If the curve does not fit the grid search results well, **(c)** the difference between training and test scores is used to find the locations where it does not exceed a threshold (e.g. 0.03) in order to avoid overfitting.

more than a few metres; hence the reprojection of coordinates of ice chart polygons is done with Geospatial Data Abstraction Library (GDAL; GDAL/OGR contributors, 2019) using a simple third-order polynomial fitted using the ground control point information from the Sentinel-1-product-included auxiliary data.

After the reprojection, the following three layers are extracted: total ice concentration (CT), partial ice concentration of each ice type (CP), and stage of development (SoD). CT is important because areas with low CT can be misinterpreted as open ocean in a SAR image. Heinrichs et al. (2006) reported that the ice edge determined from AMSR-E passive microwave radiometer data using the isoline of 15 % concentration best matches the ice edge determined from RADARSAT-1 SAR data using visual inspection. After the visual comparison of many SAR images and the corresponding reprojected ice charts, we set a threshold of 20 % for CT

Table 2. Distribution of the image acquisition dates prior to the publication of the reference ice chart.

	Training and test dataset (DS1)				Validation dataset (DS2)			
Days prior to the date of ice chart publication	3	2	1	0	3	2	1	0
Winter	124	168	77	50	78	75	67	61
Summer	119	125	112	65	87	67	48	30

to discard water-like pixels. Note that the ice concentration label in the SIGRID-3 format is assigned in an increment of 10 %. CP is also important in finding the dominant ice type in the given polygons. SoD is a so-called ice type. It is challenging to differentiate ice types using SAR data only; thus we merged the SoDs into five simple classes: open water, new ice, young ice, first-year ice, and old ice. For the summer season (June–August), there is almost no new ice and young ice annotated in the NIC ice chart, and the SoDs are further merged into three classes: open water, mixed first-year ice, and old ice. Figure 3 demonstrates an example of the ice chart preprocessing explained above with the colours following the WMO nomenclature (JCOMM, 2014b). Comparing the original SoD in panel (a) with the processed SoD in panel (d), it is clear that the ice edge of the processed SoD matches better with the SAR backscattering images.

2.2.2 Denoising of Sentinel-1 imagery

Sentinel-1 cross-polarization images suffer from strong noise, some which originates from combined effects of the relatively low signal-to-noise ratio of the sensor system and insufficient noise vector information in the extra-wide-swath mode level 1 product (Park et al., 2018). For surfaces with low backscattering such as calm open water and level sea ice without the presence of frost flowers on top, the effects from thermal noise contamination are visible not only in the backscattering image but also in some of the texture images (Park et al., 2019). The authors have developed an efficient method for textural denoising which is essential for the pre-processing of Sentinel-1 TOPSAR dual-polarization products. Denoising ensures beam-normalized texture properties for all subswaths, which helps seamlessly mosaic of multi-pass images regardless of the satellite orbit and image acquisition geometry. By following the methods developed in Park et al. (2018, 2019), each of the Sentinel-1 images was denoised before further processes are applied. As the noise power subtraction yields negative intensity values where the backscattering power is close to the noise floor, more often in HV polarization, which has lower backscatter than in HH polarization, we added mean of the noise power back to the denoised result so that those pixels do not turn into NaN (not a number) by the sigma naught conversion of linear scale to log scale (decibel).

2.2.3 Incidence angle correction

It is well-known that there is a strong incidence angle dependence in the SAR backscattering intensity for open water and sea ice surface (Mäkynen et al., 2002; Mäkynen and Karvonen, 2017). For a wide-swath SAR system like Sentinel-1 TOPSAR, varying backscatter intensity confuses image interpretation. The quasi-linear slopes in the plane of incidence angle versus sigma nought in decibel scale are reported as -0.24 and -0.16 dB per degree for typical first-year ice (Mäkynen and Karvonen, 2017), -0.27 and -0.26 dB per degree for level first-year ice and -0.23 and -0.23 dB per degree for multi-year ice (Lohse et al., 2020) in HH and HV polarization, respectively. To normalize the backscattering intensity for all swath ranges, these slopes were compensated for or used as an input layer in several ice classification algorithms in the literature (Liu et al., 2015; Zakhvatkina et al., 2013, 2017; Karvonen, 2014, 2017; Aldenhoff et al., 2018). Although the angular dependency is not a system-dependent variable but is governed by physical characteristics of the backscattered surface, the numbers need to be reassessed because the estimations of Mäkynen and Karvonen (2017) might have been affected by the residual thermal noise which used to be very strong before the ESA updated the noise removal scheme in 2018 (Miranda, 2018).

Figure 4 shows incidence angle dependence in the SAR backscattering intensity for mixed sea ice types. From the Sentinel-1 dataset described in Sect. 2.1, sea ice pixels were extracted by using daily global sea ice edge products available from the EUMETSAT Ocean and Sea Ice Satellite Application Facilities (OSI SAF). For the midwinter season (January–March displayed as a blue background), the estimated mean slope in HH polarization was -0.21 dB per degree, which is slightly different from the estimation of the first-year ice (-0.24 dB per degree) in Mäkynen and Karvonen (2017) and in between the estimations for first-year ice (-0.22 dB per degree) and multi-year ice (-0.16 dB per degree) in Mahmud et al. (2018). For HV polarization, the estimated slope was only -0.06 dB per degree, which is much lower than -0.16 dB per degree for deformed first-year ice in Mäkynen and Karvonen (2017); however, it is in line with the estimations in Liu et al. (2015). Work by Leigh et al. (2014) stated that the HV polarization backscatter signatures are largely unaffected by incidence angle variation in their RADARSAT-2 dataset. For the summer season (June–
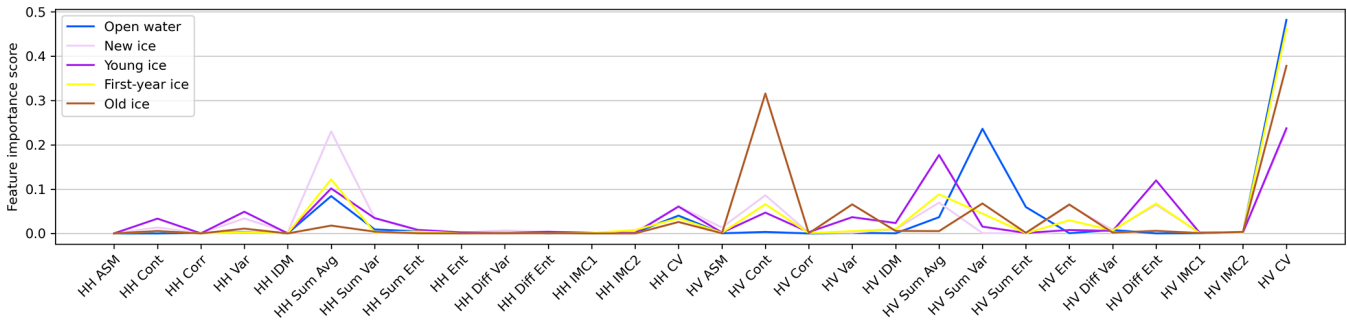
Figure 6. Feature importance of the binary sub-classifiers. ASM: angular second moment; Cont: contrast; Corr: correlation; Var: variance; IDM: inverse difference moment; Sum Avg: sum average; Sum Var: sum variance; Sum Ent: sum entropy; Ent: entropy; Diff Var: difference variance; Diff Ent: difference entropy; IMC: information measures of correlation; CV: coefficient of variation. For definitions of each parameter, please refer to Haralick et al. (1973).

August displayed by red background), the mean slopes increased to -0.28 and -0.08 dB per degree in HH and HV polarization, respectively. Scharien et al. (2014) reported significant slopes for ice adjacent to melt ponds in June, and Gill et al. (2015) also found slopes of -0.33 and -0.25 for smooth first-year ice in May in HH and HV polarization, respectively. The smaller slopes in our estimation are likely due to the mixed ice types and structures; the SAR backscattering of deformed ice has lower incidence angle dependency as shown in Mäkynen and Karvonen (2017).

We compensate for the incidence angle dependence using the estimated slopes with respect to the nominal scene centre angle of 34.5° as reference. Although the incidence angle dependence changes with ice type and radar frequency (Mahmud et al., 2018), the compensation is done for all pixels in the image using a single value of mean slope because the ice types are not identified in this stage. Open water areas of the image are also affected; however, the correction is also beneficial since the incidence angle dependence for open water is stronger (-0.65 dB per degree for wind velocity of $5\,\mathrm{m\,s^{-1}}$, computed from CMOD5 C-band geophysical model function in Hersbach et al., 2007); thus the corrected image has less incidence angle dependence.

2.2.4 Texture feature computation

Like many of the previously developed sea ice type classification methods (Shokr, 1991; Barber and LeDrew, 1991; Soh and Tsatsoulis, 1999; Deng and Clausi, 2005; Zakhvatkina et al., 2013; Leigh et al., 2014; Liu et al., 2015), the proposed approach starts from a grey level co-occurrence matrix (GLCM) calculation. The GLCM is a four-dimensional matrix $P(ijda)$ calculated from two grey tones of reference pixel i and its neighbour j, with co-occurrence distance d and orientation a. Haralick et al. (1973) introduced a set of GLCM-based texture features called Haralick features, and its practicality has been reported in numerous studies. Since 13 Haralick features can be calculated for each of the two-dimensional slices $P(ij)$ for multiple d and a

values, the maximum number of texture features is to be $2 \times 13 \times d \times a = 26\,da$, where 2 accounts for dual polarization. It is common to take the directional average for 0, 45, 90, and 135° to reduce GLCM dimensionality. Further averaging for multiple distances (1 to $w/2$, where w is the size of the subwindow for GLCM computation) is taken after computing the normalized GLCM. The spatial resolution of the texture features is the pixel spacing of the Sentinel-1 EW-mode GRDM image (40 m) multiplied by w. In this study, we set w as 25 so that the grid spacing of the result of texture analysis is 1 km.

An important factor that influences the computed texture features is the number of grey levels, L. Considering the radiometric stability of Sentinel-1 EW mode (0.32 dB; Miranda, 2018) and the range of sigma nought for various ice types (-31 to 0 dB for HH, -32 to -7 dB for HV; estimated from DS1 and DS2 after incidence angle correction), the number of grey levels should be sufficiently large enough to capture their actual differences in sigma nought values. The optimal quantization level can be calculated using the ratio of sigma nought range to radiometric resolution as follows:

$$\text{For HH,} \quad \frac{(0\,\mathrm{dB}) - (-31\,\mathrm{dB})}{0.32\,\mathrm{dB}} = 96.875, \tag{1}$$

$$\text{For HV,} \quad \frac{(-7\,\mathrm{dB}) - (-32\,\mathrm{dB})}{0.32\,\mathrm{dB}} = 78.125. \tag{2}$$

Since L should be sufficiently large to take full advantage of the system capability and yet the computation cost should not be too expensive, in this study, we set L as 64, which is the closest power of 2 to the resulting numbers from the equations above.

In addition to 13 Haralick features, the coefficient of variation (CV) which is reported as a useful feature for ice–water discrimination (Keller et al., 2017) is included. The CV is defined as follows:

$$\text{CV} = \sigma/\mu, \tag{3}$$

where σ and μ are the standard deviation and mean of the samples in a given subwindow. Since CV can be computed

Figure 7. The 1 d mosaics of Sentinel-1A/1B images (**a**: HH; **b**: HV) and the ice classification result (**d**) on 5 February 2019. The publication date of the reference weekly ice chart (**c**) is 8 February 2019.
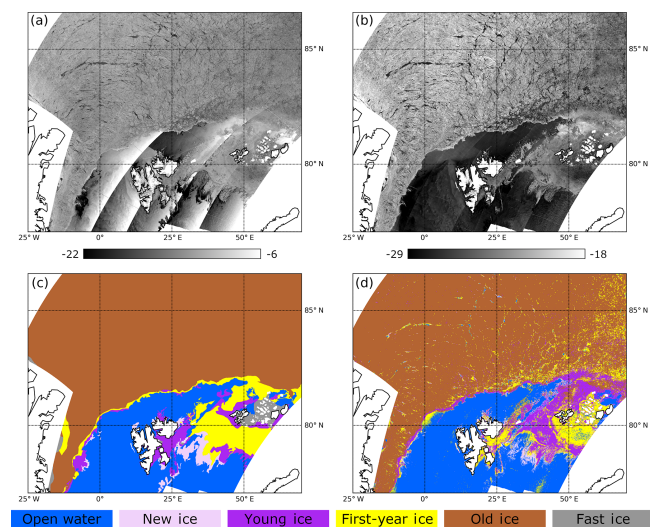

Figure 8. The 1 d mosaics of Sentinel-1A/1B images (**a**: HH; **b**: HV) and the ice classification result (**d**) on 8 February 2019. The publication date of the reference weekly ice chart (**c**) is 8 February 2019.

for each polarization image, the number of texture features for Sentinel-1 dual-polarization data is extended to 28. Incidence angle and day of the year can also be added. The former is adopted to account for possible residuals from the angular dependency correction while the latter is to account for seasonal variability. Although these two are not any type of textures, they can be used as input features for image classification. Note that it is important to have each ice type spatially and temporally evenly distributed if these two additional features are included; otherwise, the trained classifier will result in a biased prediction. The effects of including these extra features will be tested and discussed in later sections.

2.2.5 Machine learning classifier

Since there are hundreds of algorithms in the field of machine learning (ML) and each of the different algorithms has its own pros and cons, it is not easy to compare their performances and decide what to use. Fernández-Delgado et al. (2014) found that the random forest (RF; Ho, 1998) was the best classifier for various types of datasets with a slight difference from a support vector machine (SVM; Cortes and Vapnik, 1995). In the previous studies about sea ice classification (e.g. Leigh et al., 2014; Liu et al., 2015; Zakhvatkina et al., 2017), the SVM was used often because by nature it works relatively well even when the number of datasets is small. When the training dataset is prepared by manual work (i.e. manual classification by human expert), the number of images is not large, usually fewer than 20 (e.g. 12 scenes in Zakhvatkina et al., 2013; 20 scenes in Leigh et al., 2014; 1 scene in Liu et al., 2015; 4 scenes in Ressel et al., 2015). However, the number can increase with less effort when
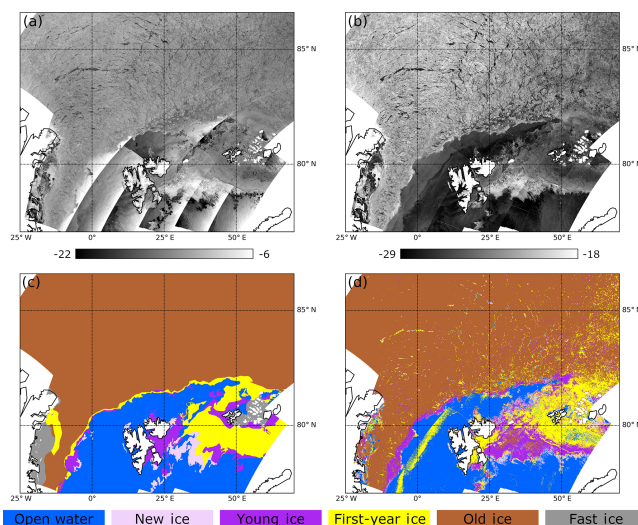
the readily available ice charts are used as training references. Besides, there is no need to rely on additional manual work prone to contamination by biased decisions. The RF has two practical advantages when processing a large number of datasets. First, the RF is scale-invariant and does not require preprocessing of the datasets, whereas the SVM requires scaling and normalization. Second, the computational complexity of the RF is lower than that of the SVM. For the SVM, the number of operations is $O(n^2 p + n^3)$ and $O(n_{sv} p)$ for training and prediction while for RF it is $O(n^2 p n_{tr})$ and $O(n_{tr} p)$, respectively, where n is the number of samples, p is the number of features, n_{sv} is the number of support vectors, and n_{tr} is the number of trees. Considering the practical requirements of fast processing for near-real-time ice charting services, the RF can be a reasonable solution. We use the RF with the Python Scikit-Learn implementation (Pedregosa et al., 2011).

We split the RF classifier into several binary classifiers using a one-vs.-all scheme (Anand et al., 1995). Although the standard RF algorithm can inherently deal with a multiclass problem, the one-vs.-all binarization to the RF results in better accuracy with smaller forest sizes than the standard RF (Adnan and Islam, 2015).

Three hyperparameters of the RF classifier were tuned: number of trees (N_T), maximum tree depth (D), and maximum number of features (N_F). Usually, with the higher N_T and D, the model better fits to the data. However, increasing forest size can slow down the training process considerably, and more importantly, it can cause overfitting. Therefore, it is important to tune these hyperparameters adequately so that the processing time and performance are in balance. To determine the best values of the hyperparameters, a grid search

Table 3. Confusion matrix of the five-class RF classifier which was trained with and applied to the DS1 winter dataset. The bold font indicates where the classification was done correctly.

		Predicted														
		OW (open water)			NI (new ice)			YI (young ice)			FYI (first-year ice)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**90.1**	**91.3**	**92.4**	1.3	1.1	1.2	1.6	1.7	1.6	6.9	5.9	4.8	0.0	0.0	0.0
	NI	30.0	28.0	26.1	**21.9**	**23.8**	**45.0**	27.1	26.7	13.3	16.7	17.7	11.8	4.2	3.9	3.8
	YI	3.7	3.7	3.9	5.1	4.8	5.2	**58.8**	**60.1**	**62.6**	24.3	24.0	21.1	8.1	7.4	7.2
	FYI	5.0	4.7	3.9	1.7	1.5	1.7	18.7	19.0	19.4	**64.4**	**65.3**	**65.9**	10.1	9.6	9.1
	OI	0.1	0.1	0.2	0.3	0.3	0.6	3.1	2.9	3.0	6.4	6.1	5.6	**90.1**	**90.6**	**90.6**

Table 4. Confusion matrix of the five-class RF classifier which was trained with the DS1 winter dataset and applied to the DS2 winter dataset. The bold font indicates where the classification was done correctly.

		Predicted														
		OW (open water)			NI (new ice)			YI (young ice)			FYI (first-year ice)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**89.1**	**90.2**	**90.3**	1.3	1.1	1.9	3.3	3.5	4.2	6.4	5.2	3.7	0.0	0.0	0.0
	NI	45.1	45.0	56.5	**31.9**	**30.6**	**17.6**	6.0	5.7	13.3	15.1	17.3	11.2	2.0	1.5	1.5
	YI	7.1	7.1	9.2	6.3	5.9	8.5	**47.6**	**48.0**	**55.0**	28.7	29.2	17.3	10.4	9.8	9.9
	FYI	5.6	5.0	5.8	3.8	3.5	3.2	32.8	33.0	35.0	**38.4**	**39.7**	**37.3**	19.3	18.8	18.7
	OI	0.3	0.3	0.7	0.5	0.4	0.7	1.9	1.8	1.9	4.6	4.8	4.5	**92.8**	**92.8**	**92.6**

with five-fold cross-validation (Kohavi, 1995) is used. The grid (all possible combinations of N_T, D, and N_F values) is set on a logarithmic scale (Table 1) because the performance change with hyperparameter is typically on a logarithmic scale. Classification scores with values ranging from 0 (worst performance) to 1 (best performance) are evaluated for each node of the grid and are interpolated between the nodes by curve fitting. The Richards curve (Richards, 1959) was used as the fit model because it allows easy estimation of the model's maximum value. The optimal values for N_T, D, and N_F are selected based on the saturation of score increments, difference between training and testing scores, and computational load considerations.

2.2.6 Training and validation

To train an ice type classifier, a set of collocated SAR images and ice charts is required. After the preprocessing of the ice chart including reprojection into the SAR image geometry, only the samples with spatially and temporally good matches should be fed to the training phase. The goodness of matching should be examined as the weekly ice chart is produced by merging information from many image sources acquired in different time instances; hence the ice locations and conditions are unlikely to match to those in every SAR image. As no explicit scene identifier or time information of the images used in ice charting is provided with the ice chart itself, the basic strategy in image selection is to find a pair
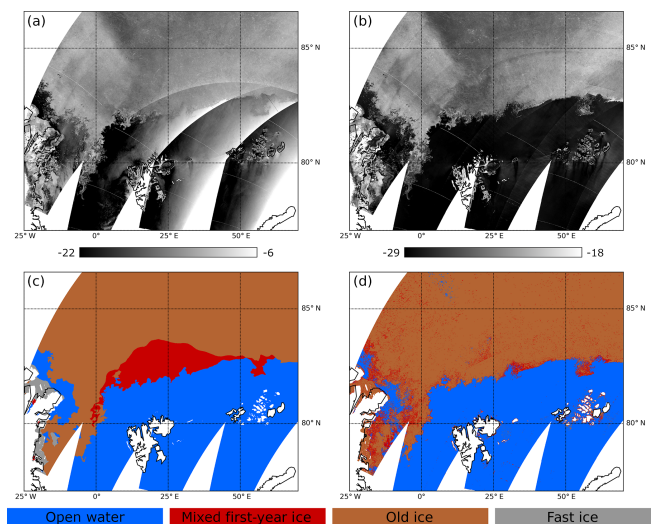

Figure 9. The 1 d mosaics of Sentinel-1A/1B images (**a**: HH; **b**: HV) and the ice classification result (**d**) on 13 August 2018. The publication date of the reference weekly ice chart (**c**) is 16 August 2018.

of a SAR image and an ice chart which match well visually. Such an image selection is trivial, but not easy to automate. Since the weekly ice chart is made partly based on the SAR images acquired in the past 3 d from the date of publication,

Table 5. Classification accuracies before and after applying textural denoising.

	Case								
	FC1			**FC2**			**FC3**		
Class	Thermal denoising only	Textural denoising applied	Difference	Thermal denoising only	Textural denoising applied	Difference	Thermal denoising only	Textural denoising applied	Difference
Open water	88.0	89.1	+1.1	88.7	90.2	+1.5	89.4	90.3	+1.9
New ice	32.2	31.9	−0.3	32.2	30.6	−1.6	20.0	17.6	−2.4
Young ice	45.2	47.6	+2.4	44.7	48.0	+3.3	48.8	55.0	+6.2
First-year ice	38.6	38.4	−0.2	39.4	39.7	+0.3	36.4	37.3	+0.9
Old ice	88.9	92.8	+3.9	88.5	92.8	+4.5	88.6	92.6	+4.0
Kappa	0.59	0.67	+0.08	0.59	0.67	+0.08	0.59	0.67	+0.08

Table 6. Confusion matrix of the three-class RF classifier which was trained with and applied to the DS1 summer dataset. The bold font indicates where the classification was done correctly.

		Predicted								
		OW (open water)			mFYI (mixed first-year ice)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual OW	OW	**98.1**	**97.9**	**98.7**	0.7	0.7	0.6	1.2	1.4	0.6
mFYI	mFYI	4.1	3.9	2.4	**14.9**	**15.5**	**41.8**	81.1	80.6	55.8
OI	OI	1.5	1.4	1.0	5.5	5.7	4.9	**93.0**	**92.9**	**94.1**

the ice edges in some images match well with those in the ice chart.

In order to automate image selection, the ice edges in SAR images need to be identified first. Since even an ice–water classifier has not been well-developed yet for Sentinel-1, the image selection procedure has to be done manually in the beginning. However, once a classifier is generated with high accuracy, it can be used to automate the procedure; then the whole process in the proposed scheme will be fully automated. This is why the proposed algorithm is named "semi-" automated for now. Nevertheless, the manual selection to guarantee a "good match" is done by visual inspection of ice–water boundaries overlaid on SAR images. The ice–water boundary can be extracted easily from the reprojected ice chart. Then the SAR backscattering image contrasts across the ice–water boundaries are examined in both HH and HV polarization because the image contrast between ice–water is larger in HV while smooth level ice is more easily identified in HH.

After the image selection, the samples in the selected images are split randomly into training and test datasets with a ratio of 7 : 3. For the training dataset, further data selection is made by excluding the samples residing close to the polygon boundaries. This is to account for possible mismatch due to various reasons (e.g. ice drift, vector mapping error, image geocoding error). In this study, only the data from pixels more than 3 km away from the polygon boundaries were

fed into the training process. Once the hyperparameter optimization is done, the RF classifier is trained for the training dataset. The trained classifier is then applied to the test dataset. For performance evaluation, we use confusion matrix and Cohen's kappa coefficient κ (Cohen, 1960), which measures the agreement between two rasters (in this study, they are the output from the trained classifier and the reference ice chart), taking account of the possibility of the agreement occurring by chance. The validation is done in the same way but using a completely independent dataset. The DS1 was used to run the training phase. Among 4485 images in total, we selected 840 images (419 for the winter season and 421 for the summer season) of which ice edges match well with the collocated ice chart. From the selected images, 120 million samples covering open water and sea ice were divided into training and test datasets. The DS2 was used to evaluate the performance of the trained classifier using a temporally independent dataset of 513 images (281 for the winter season and 232 for the summer season). The distribution of the image acquisition dates prior to the publication of the reference ice chart is shown in Table 2.

It might not be enough to assess the quality of the classifier output when it is trained with, and evaluated against, only NIC ice charts. The accuracy could be indirectly investigated by comparing the output from our classifier against another data source, such as the OSI SAF sea ice type product (OSI-403-c). The ice classes of OSI-403-c are assigned from at-

Table 7. Confusion matrix of the three-class RF classifier which was trained with the DS1 summer dataset and applied to the DS2 summer dataset. The bold font indicates where the classification was done correctly.

	Case	Predicted								
		OW (open water)			mFYI (mixed first-year ice)			OI (old ice)		
		FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**99.5**	**99.4**	**96.2**	0.2	0.2	3.2	0.3	0.4	0.6
	mFYI	5.4	5.1	3.0	**12.0**	**11.2**	**25.8**	82.5	83.7	71.2
	OI	2.9	2.7	2.2	5.8	5.8	13.4	**91.2**	**91.4**	**84.4**

Table 8. Confusion matrix of the three-class RF classifier which was trained with the DS1 summer dataset and applied to the DS1 summer dataset of each month. The bold font indicates where the classification was done correctly.

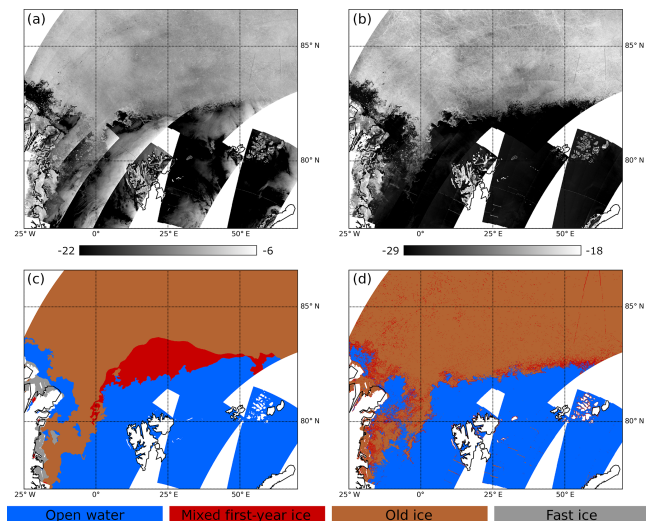| | Case | Predicted (FC1) | | | | | | | | |
		OW (open water)			mFYI (mixed first-year ice)			OI (old ice)		
		Jun	Jul	Aug	Jun	Jul	Aug	Jun	Jul	Aug
Actual	OW	**99.0**	**99.0**	**98.1**	0.9	0.3	0.3	0.1	0.7	1.6
	mFYI	4.7	2.1	3.9	**60.6**	**32.6**	**26.5**	34.7	65.3	69.6
	OI	0.6	0.8	2.1	7.7	7.7	10.1	**91.7**	**91.5**	**87.8**

Figure 10. The 1 d mosaics of Sentinel-1A/1B images (**a**: HH; **b**: HV) and the ice classification result (**d**) on 16 August 2018. The publication date of the reference weekly ice chart (**c**) is 16 August 2018.

mospherically corrected brightness temperatures of passive microwave radiometers (SSMIS and AMSR2) and backscatter values of radar scatterometer (ASCAT), using a Bayesian approach (Aaboe et al., 2018).
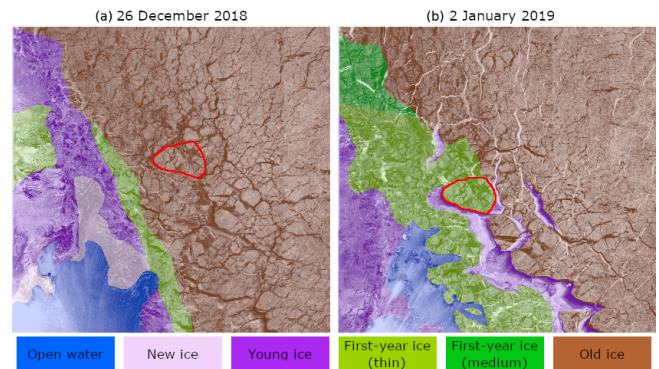

Figure 11. An example of the inconsistency of the ice charts. Note that the SoD labels and colours are of an original NIC ice chart, while those in Figs. 7 and 8 are of a simplified version as described in Sect. 2.2.1. The SoDs from the NIC ice charts on different dates (26 December 2018 and 2 January 2019) are superimposed on the Sentinel-1 backscattering image of the corresponding dates. The same ice floe (red outline) is classified differently in each ice chart (old ice in **a** and first-year ice in **b**) despite the similarity in the SAR backscattering images. Source credits: U.S. National Ice Center (colours) and European Space Agency (background).

3 Results and discussion

We trained three RF classifiers with different feature configurations: (i) FC1: Haralick texture features and CV; (ii) FC2: Haralick texture features, CV, and incidence angle; and (iii) FC3: Haralick texture features, CV, incidence angle, and day of the year.

Table 9. Confusion matrix of the three-class RF classifier which was trained with the DS1 summer dataset and applied to the DS2 summer dataset of each month. The bold font indicates where the classification was done correctly.

| | | Predicted (FC1) | | | | | | | | |
| | | OW (open water) | | | mFYI (mixed first-year ice) | | | OI (old ice) | | |
	Case	Jun	Jul	Aug	Jun	Jul	Aug	Jun	Jul	Aug
Actual	OW	**90.6**	**99.5**	**99.7**	8.1	0.2	0.1	1.3	0.3	0.2
	mFYI	3.4	4.0	4.2	**55.6**	**41.4**	**11.0**	41.0	54.9	84.9
	OI	1.4	3.4	3.4	11.2	10.9	6.3	**87.5**	**85.6**	**90.3**

As expected, the classification score increases with the number of trees (crosses in Fig. 5a), and Richards curve (dashed line) fits well to the observations (RMSE $= 2.3 \times 10^{-4}$). The optimal N_T value is selected where the score increment per tree (i.e. local slope) becomes less than 0.001 (i.e. accuracy increase of 0.1 %) and constitutes 11 trees, thus keeping the forest size small. The scores also increase with the maximum tree depth (crosses in Fig. 5b), but Richards' curve (dashed line) does not fit so well (RMSE $= 3.6 \times 10^{-3}$) and cannot be used for finding the optimal D value. This can be explained by overfitting of the classifier and illustrated by the difference between training and testing scores (Fig. 5c): a small difference between the scores (for $D \leq 8$) indicates similar performance on training and testing datasets, while a large difference (for $D > 8$) indicates that the testing dataset is processed with worse results. The optimal D value is therefore selected where the score difference becomes higher than 0.03 and constitutes eight levels. The optimal value of the number of features (N_F) was selected using the same criterion as for N_T, and the value constitutes 10 features. As a result, the optimal hyperparameters of the number of trees, the maximum tree depth, and the number of features were 11, 8, and 10, respectively.

The trained five-class classifier consists of five binary sub-classifiers; each of them is used for discriminating one specific class from the others. For each sub-classifier, each texture feature has a different weight in decision making. The fraction of the samples that each texture feature contributes can be used to compute the relative importance of the features, and the averaged estimates of them over several randomized trees serve as an indicator of feature importance (Louppe, 2014). The feature importance of the sub-classifiers is presented in Fig. 6. The overall pattern shows that the features of HV polarization play a more important role than those of HH polarization. For HH polarization, the sum average, which is equal to the mean backscattering intensity in each subwindow, was the prominent feature. For HV polarization, however, contrast and variance- and entropy-related features were more important. The classifiers for open water and old ice have stronger dependencies on HV polarization than others. This is understandable because the main radar scattering mechanisms for those two types are strongly char-

acterized by the portion of volume scattering: low for calm water and high for dry ice with low salinity (old ice). The classifier for new ice has a distinctive pattern that the sum averages in both polarizations are much more important than other features. This might be because the new ice has different types of recently formed ice including nilas, which is smooth but rafting can make rough features, and frost flowers, which introduce high surface roughness and volume scattering (Isleifson et al., 2014). Thus the new ice can appear either featureless and dark or complex and bright in a SAR image (Dierking, 2010). The large range in backscatter values makes it hard to define characteristic texture in the new ice patch.

The confusion matrix for testing the trained classifier for winter season with the test dataset (DS1) is shown in Table 3. Three cases with different feature configurations (FC1–FC3) were tested. The accuracies for open water and old ice were higher than 90 %; however, those for young ice and first-year ice were around 60 %. The mean difference between the results of FC1 and FC2 was only 1.2 %, indicating that residual angular dependency was negligible after the incidence angle correction. However, the accuracy significantly improved from FC2 to FC3, especially with new ice (21.2 %). The Cohen kappa coefficients κ for FC1, FC2, and FC3 were 0.70, 0.71, and 0.77, respectively. It should be noted that the evaluation of the DS1 was carried out with the input dataset that was used for training. Thus, the test and training data share the same ice conditions as well as spatio-temporal coverage. As a result, the κ might contain correlation which is not preferable for proper evaluation. Table 4 shows the confusion matrix for validation results from the DS2 of which the accuracy of open water and old ice was at a similar level, compared to the DS1. Meanwhile, the accuracy of young ice and first-year ice decreased considerably. The differences between the results of FC1, FC2, and FC3 were insignificant. This result is opposite to the DS1 inferring that the training with FC3 was overfitted and the day of the year may not correspond to the temperature, air–sea fluxes, or weather regimes. The κ values for FC1, FC2, and FC3 with the DS2 were 0.67, 0.67, and 0.67, respectively.

To see how the denoising step in Sect. 2.2.2 led to improvements in the classification accuracies, the same training
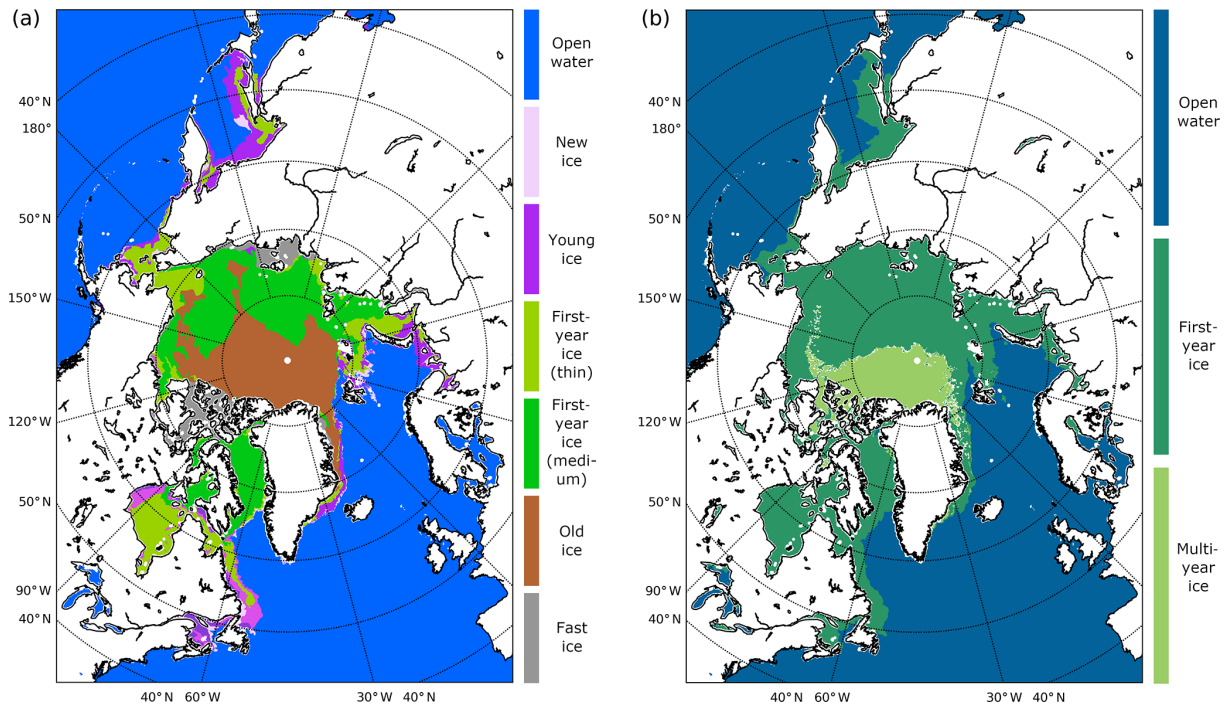
Figure 12. The ice types in NIC ice chart **(a)** and OSI SAF sea ice type product **(b)** for the same date (8 February 2019). Note that the SoD labels and colours follow those defined in each ice chart format. Source credits: U.S. National Ice Center **(a)** and EUMETSAT Ocean and Sea Ice Satellite Application Facilities **(b)**.

Table 10. Confusion matrix of the three-class RF classifier which was trained and applied to the DS1 winter dataset. The bold font indicates where the classification was done correctly.

		Predicted								
		OW (open water)			mFYI (mixed first-year ice)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**92.2**	**92.5**	**93.6**	7.8	7.4	6.3	0.0	0.0	0.0
	mFYI	6.4	5.5	5.5	**83.8**	**85.3**	**85.5**	9.8	9.2	9.0
	OI	0.2	0.2	0.2	8.9	8.5	8.7	**90.9**	**91.3**	**91.1**

and evaluation were conducted for the same dataset without applying the textural noise correction (Table 5). In all configurations (FC1–FC3), the accuracies improved for young ice (+2.4 % to +6.2 %) and old ice (+3.9 % to +4.5 %), which were most pronounced compared to those for open water (+1.1 % to +1.9 %) and first-year ice (−0.2 % to +0.9 %). On the contrary, a small accuracy decrease was observed for new ice (= 2.4 % to −0.3 %). Nevertheless, the improvement in κ (+0.08) demonstrates a clear improvement in the overall classification result.

The confusion matrix for testing the trained classifier for the summer season with the test dataset (DS1) is shown in Table 6. As described in Sect. 2.2.1, the further simplified three-class classification is applied. The accuracies for open water and old ice were higher than 92 %; however, the accuracies for mixed first-year ice were only around 15 % in

both FC1 and FC2 and 42 % in FC3. The large difference between the results of FC1-FC2 and FC3 indicates that the mixed first-year ice likely changes at short timescales. The misclassifications for mixed first-year ice were mostly into old ice. This might be because of the surface melting and the corresponding image textures which make the discrimination between the mixed first-year ice and old ice difficult. The same patterns were observed from the confusion matrix (Table 7) for validation results from the DS2, except that the accuracy decrease from Tables 6 to 7 was particularly large for FC3, meaning overfitting for DS1. However, it is unclear if the low accuracy for mixed first-year ice is due to the classifier itself or the data. To unravel this, the data were divided into three groups of 1 month each (June, July, August), and then separate classifiers were trained and tested. Tables 8 and 9 show the results with FC1 configuration for

Table 11. Confusion matrix of the three-class RF classifier which was trained with the DS1 winter dataset and applied to the DS2 winter dataset. The bold font indicates where the classification was done correctly.

		Predicted								
		OW (open water)			mFYI (mixed first-year ice)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**91.4**	**91.7**	**91.7**	8.6	8.5	8.5	0.0	0.0	0.0
	mFYI	9.4	8.5	9.9	**75.0**	**76.5**	**74.6**	15.6	15.2	15.5
	OI	0.3	0.3	0.3	6.3	6.5	6.6	**93.3**	**93.2**	**93.1**

DS1 and DS2, respectively. There was a rapid accuracy decrease for mixed first-year ice from June to August in both results (from 60.6 % to 26.5 % for DS1 and from 55.6 % to 11.0 % for DS2). As there is no particularly large difference between the numbers in Tables 8 and 9, meaning the classifiers were not overfitting, the very low accuracies for mixed first-year ice in Table 6 (14.9 %) and Table 7 (12.0 %) seem to be due to a too large temporal extent for a single classifier. In other words, the classifier for the summer season needs to be split into multiple groups with shorter time spans, and/or a feature that can effectively account for surface melting needs to be introduced into the algorithm for further development. Based on these results so far, the trained classifiers in the summer season for 3 months in bulk failed in distinguishing mixed first-year ice from old ice; thus they are close to ice–water discriminators rather than ice type classifiers.

Figure 7 shows a daily mosaic of Sentinel-1 SAR images over the study area and the classified ice map in the winter season. For comparison, the NIC weekly ice chart is also displayed. Despite the SAR images being acquired 3 d before the ice chart was published, the ice edges of the ice chart match well with the SAR mosaic in most parts because the same SAR data were used. Overall, the discriminations between ice and non-ice, old ice and other ice types, and detection of new ice patches look reasonable. However, some young ice patches, for example the ice patches between the Svalbard archipelago, are misclassified as first-year ice. Figure 8 shows another daily mosaic made by the images acquired on the same day of the ice chart publication. Considering notable ice drift in the backscattering images in Figs. 7 and 8, the SAR-based ice classification results in both figures looking consistent, well in line with the ice drift. Although the weekly ice chart is supposed to represent the averaged ice status in the past few days, the actual ice distribution on the actual date of the publication can be largely different. This example shows a clear potential of near-real-time service of ice type classification.

Figures 9 and 10 show the same mosaics for the case in the summer season. As shown in Tables 6 and 7, the misclassifications for the mixed first-year ice into old ice are pronounced in the large ice patches north to Svalbard, while the ice edge positions of the ice chart and the classification result are in good agreement with each other.

To cope with the ambiguous classification for the winter season ice types with low accuracy, we conducted a test with the three-class classification, and Tables 10 and 11 show the resulting confusion matrices. The κ values for FC1, FC2, and FC3 were 0.83, 0.84, and 0.84 in DS1 and 0.75, 0.75, and 0.74 in DS2, respectively. The dramatic increase in the accuracy of the mixed first-year ice indicates that the misclassification for the new ice, young ice, and first-year ice was mostly among themselves. However, the accuracy decrease from DS1 to DS2 was at a similar level to the case of the five-class classification. This could have been caused by inconsistent labelling in the reference ice chart.

Figure 11 shows an example of the inconsistent labelling in the reference ice chart. The SoDs from the NIC ice charts are superimposed on the Sentinel-1 backscattering images. The same ice floe (red outline) is classified differently in two different ice charts (old ice on the left panel and first-year ice on the right panel), although it looks almost the same in the corresponding SAR backscattering images. It should be noted that training with the ice chart might have included mislabelled small features even if the image selection based on ice edge matching was successful. Furthermore, the boundaries between different ice types in the ice chart are normally not as precise as those in the SAR image-based classification results. Therefore, the lower classification accuracies compared to those in previous studies (80 % in Zakhvatkina et al., 2013; 91.7 % in Liu et al., 2015; 87.2 % in Aldenhoff et al., 2018), which used manually classified ice maps as training and validation reference, are expected.

Unfortunately, we could not find an official report regarding the accuracy of the NIC ice chart information.

Table 12 shows the confusion matrices for our three-class classifiers when their prediction results are compared with the OSI-403-c product as reference. For one-to-one comparison, it was assumed that the ideal characteristics of the mixed first-year ice and the old ice in our three-class classification are equivalent to those of the first-year ice and the multi-year ice in OSI-403-c. Comparing with the results in Table 11, the accuracies for open water decreased by 6 %; however, this is mainly because the ice concentration threshold for ice–

Table 12. Confusion matrix of the three-class RF classifier which was trained with the DS1 winter dataset and applied to the DS2 winter dataset with reference to the OSI SAF sea ice type product (OSI-403-c). Bold indicates where the classification was done correctly.

| | | | Predicted (classifier was trained with NIC ice chart) | | | | | | | | |
| | | | Open water | | | First-year ice | | | Multi-year ice | | |
		Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Reference (OSI SAF)		Open water	**85.9**	**86.1**	**86.2**	12.6	12.4	12.1	15.7	15.3	16.2
		First-year ice	1.9	1.6	2.0	**26.0**	**26.8**	**26.9**	72.1	71.6	71.2
		Multi-year ice	0.1	0.1	0.1	1.5	1.4	1.4	**98.4**	**98.5**	**98.5**

Table 13. Averaged percent agreement of NIC weekly ice chart and OSI SAF daily sea ice type product (OSI-403-c) for the same publication dates (12 different days) in the studied domain during January–March 2019. Bold indicates where the classification was done correctly.

| | | NIC | | |
		Open water	First-year ice	Multi-year ice
OSI SAF	Open water	**90.0**	10.0	0.1
	First-year ice	0.9	**58.8**	40.3
	Multi-year ice	0.0	1.0	**99.0**

water discrimination in OSI-403-c is 35 %, which is higher than the 20 % that we set in our preprocessing of the NIC ice chart (Sect. 2.2.1). Thus areas with low ice concentration in marginal ice zone are most likely annotated as open water in OSI-403-c. The accuracies for open water at points in the NIC charts with ice concentration between 20 % and 40 % only were considerably lower, with 67.4 %, 67.8 %, and 70.1 % for FC1, FC2, and FC3, respectively (not presented in Table 12). For first-year ice, large portions (72 %) are misclassified as old ice. This might be partly explained in Fig. 12, which shows the ice classes in the NIC ice chart and OSI-403-c for the same publication date. A large extent of old ice in the NIC ice chart is annotated as multi-year ice in OSI-403-c. As our classifiers were trained with the NIC ice chart, it is natural to result in more multi-year ice for the area where the ice type is classified as first-year ice in OSI-403-c. For multi-year ice, the accuracy was the highest, 98 %.

The inconsistency in ice types between the NIC ice chart and OSI-403-c seems persistent at least for the time coverage of DS2 (January–March in 2019). Table 13 shows averaged percent agreement of the two sea ice type products for the same publication dates over 12 weeks (12 one-to-one comparisons as the NIC ice chart is a weekly product). To make a fair comparison, the ice-covered areas with ice concentrations lower than 35 %, which is the threshold for ice–water discrimination in OSI-403-c, were excluded. The percent agreement for first-year ice (58.8 %) was much lower

than that of open water (90.0 %) and multi-year ice (99.0 %), which is in line with the results in Table 12. Finding the reason for the clear discrepancy of the extent of first-year ice between the NIC ice chart and OSI-403-c is beyond the scope of this study; however, it should be noted that elaborate future work for cross-calibrating ice types in different ice charts is necessary.

The proposed algorithm has several limitations. First of all, the variations in radar backscattering and its corresponding image textures due to seasonal changes were not properly captured. Although day of the year was tested as a seasonality variable in the FC3 feature configuration, the result did not show any improvement. This is because SAR image features, which partially reflect temperature fluxes and weather regimes, might not correspond to day of the year. Second, the proposed method struggles when the same type of sea ice is located on different edges of the range swath of SAR images because the incidence angle dependence could not be normalized perfectly. An example of such a failure can be seen along the image boundaries at 79.5° N, 45° E in Fig. 7 and 79° N, 50° E in Fig. 8, approximately. Third, some artefacts were observed under large ocean swells. In the classified results in Fig. 8d, there is a misclassified first-year ice patch (yellow) in the open water area. According to the high-resolution sea surface wind data from SAR on the Sentinel-1 satellites (https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:SAR-WINDS-S1, last access: 18 August 2020), the wind speed ranged from 17 to 21 m s^{-1} at the time of image acquisition, heavily roughing the water surface. Although we have included images with both high and low wind conditions in our training data, the image textures of wind-roughened water surface and ice were confused in some cases, and the same happened in the image textures of calm water surface and smooth level ice.

4 Conclusions

A new semi-automated SAR-based sea ice type classification scheme was proposed in this study. For the first time several ice types were successfully identified on Sentinel-1 SAR imagery in the winter season, while only an ice–water discrim-

ination was feasible in the summer season. The main technological innovation is two-fold: (i) reduced manual work in the preparation of a large amount of training and validation reference data using readily available public ice charts and (ii) more objective evaluation of the SAR-based sea ice type classifier compared to the previous studies conducted with a small number of images and customized ice type references from sources not open to the public. A conventional approach for selecting training and testing data by anonymous human ice experts is undesirable not only because it is laborious, but also due to subjectivity and lack of standardization in the assessment of the automated classifier. Therefore, the performance from different literature sources cannot be intercompared directly.

Test results from the datasets of the winter season acquired over the Fram Strait and the Barents Sea area showed overall accuracies of 87 % and 60 % and the Cohen kappa coefficients of 0.75 and 0.67 for the three-class and five-class ice type classifiers, respectively. These are slightly lower than the numbers in previous studies, and the errors are attributed not only to the automated algorithm but also to the inconsistency of the ice charts and the high level of their generalization. Test results from the datasets of summer seasons showed overall accuracy of 67 % and the Cohen kappa coefficient of 0.78 for the three-class classifiers. Considering the misclassifications in different ice types were among themselves, the three-class classifiers are not really a sea ice type classifier, but they performed well at least as an ice–water discriminator with accuracy of 98 %.

Based on the results, we envisage that three-class ice type classification from SAR imagery would be useful for making a global sea ice type product like OSI SAF OSI-403-c with higher spatial resolution. The proposed approach importantly showed that a daily ice type mapping from the Sentinel-1 data is feasible and can help capture details of short-term changes in the stage of sea ice development. Based on the achieved results, we believe that the proposed approach may be efficiently used for operational ice charting services for supporting navigation in the Arctic.

Author contributions. JWP and AAK formulated the research plan. JWP and AAK developed the algorithm. JWP implemented the algorithm and performed the data processing. JWP, AAK, MB, JSW, MWH, and HCK carried out the analyses, and JWP wrote the paper.

References

Aaboe, S., Breivik, L.-A., and Eastwood, S.: Improvement of OSI SAF product of sea ice edge and sea ice type, EUMETSAT Meteorological Satellite Conference, Geneva (Switzerland), 22–26 September 2014.

Aaboe, S., Breivik, L.-A., Sørensen, A., Eastwood, S., and Lavergne, T.: Global sea ice edge and type product user's manual OSI-402-c & OSI-403-c, Ocean & Sea Ice Satellite Application Facilities (OSI SAF), version 2.3, available at: http://osisaf.met.no/docs/osisaf_cdop3_ss2_pum_sea-ice-edge-type_v2p3.pdf (last access: 18 August 2020), 2018.

Adnan, M. N. and Islam, M. Z.: One-Vs-All Binarization Technique in the Context of Random Forest, Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges (Belgium), 22–24 April 2015.

Aldenhoff, W., Heuzé, C., and Eriksson, L.: Comparison of ice/water classification in Fram Strait from C- and L-band SAR imagery, Ann. Glaciol., 59, 112–123, https://doi.org/10.1017/aog.2018.7, 2018.

Anand, R., Mehrotra, K., Mohan, C. K., and Ranka, S.: Efficient classification for multiclass problems using modular neural networks, IEEE T. Neural Networ., 6, 117–124, https://doi.org/10.1109/72.363444, 1995.

Barber, D. G. and LeDrew, E. F.: SAR sea ice discrimination using texture statistics: A multivariate approach, Photogramm. E. Rem. S., 57, 385–395, 1991.

Cohen, J.: A coefficient of agreement for nominal scales, Educ. Psychol. Meas., 20, 37–46, https://doi.org/10.1177/001316446002000104, 1960.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Deng, H. and Clausi, D. A.: Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model, IEEE T. Geosci. Remote, 43, 528–538, https://doi.org/10.1109/TGRS.2004.839589, 2005.

Dierking, W.: Mapping of different sea ice regimes using images from Sentinel-1 and ALOS synthetic aperture radar, IEEE T. Geosci. Remote, 48, 1045–1058, https://doi.org/10.1109/TGRS.2009.2031806, 2010.

ESRI (Environmental Systems Research Institute, Inc.): ESRI Shapefile Technical Description, An ESRI White Paper, available at: http://downloads.esri.com/support/whitepapers/mo_/shapefile.pdf (last access: 18 August 2020), 1998.

European Space Agency: available at: at https://scihub.copernicus.eu/, last access: 19 August 2020.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D.: Do we need hundreds of classifiers to solve real world classification problems?, J. Mach. Learn. Res., 15, 3133–3181, 2014.

GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library, Open Source Geospatial Foundation, available at: https://gdal.org (last access: 18 August 2020), 2019.

Gill, J. P.S., Yackel, J. J., Geldsetzer, T., and Fuller, M. C.: Sensitivity of C-band synthetic aperture radar polarimetric parameters to snow thickness over landfast smooth first-year sea ice, Remote Sens. Environ., 166, 34–49, https://doi.org/10.1016/j.rse.2015.06.005, 2015.

Haralick, R. M., Shanmugam, K., and Dinstein, I.: Textural features for image classification, IEEE T. SYST. MAN. CY.-S., SMC-3, 610–621, https://doi.org/10.1109/TSMC.1973.4309314, 1973.

Hersbach, H., Stoffelen, A., and de Haan, S.: An improved C-band scatterometer ocean geophysical model function: CMOD5, J. Geophys. Res., 112, C03006, https://doi.org/10.1029/2006JC003743, 2007.

Heinrichs, J. F., Cavalieri, D. J., and Markus, T.: Assessment of the AMSR-E sea ice concentration product at the ice edge using RADARSAT-1 and MODIS imagery, IEEE T. Geosci. Remote, 44, 3070–3080, https://doi.org/10.1109/TGRS.2006.880622, 2006.

Ho, T. K.: The random subspace method for constructing decision forests, IEEE T. Pattern Anal., 20, 832–844, https://doi.org/10.1109/34.7096011998, 1998.

Isleifson, D., Galley, R. J., Barber, D. G., Landy, J. C., Komarov, A. S., and Shafai, L.: A study on the C-band polarimetric scattering and physical characteristics of frost flowers on experimental sea ice, IEEE T. Geosci. Remote, 52, 1787–1798, https://doi.org/10.1109/TGRS.2013.2255060, 2014.

JCOMM (Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology): Ice chart colour code standard, JCOMM Technical Report No. 24, Tech. Rep., World Meteorological Organization, Geneva, Switzerland, 2014a.

JCOMM (Joint WMO-IOC Technical Commission for Oceanography and Marine Meteorology): SIGRID-3: a vector archive format for sea ice georeferenced information and data, JCOMM Technical Report No. 23, Tech. Rep., World Meteorological Organization, Geneva, Switzerland, 2014b.

Johannessen, O. M., Alexandrov, V., Frolov, I. Y., Sandven, S., Pettersson, L. H., Bobylev, L. P., Kloster, K., Smirnov, V. G., Mironov, Y. U., and Babich, N. G.: Remote sensing of sea ice in the Northern Sea route: Studies and applications, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-48840-8, 2006.

Karvonen, J.: A sea ice concentration estimation algorithm utilizing radiometer and SAR data, The Cryosphere, 8, 1639–1650, https://doi.org/10.5194/tc-8-1639-2014, 2014.

Karvonen, J., Vainio, J., Marnela, M., Eriksson, P., and Niskanen, T.: A comparison between high-resolution EO-based and ice analyst-assigned sea ice concentrations, IEEE J. Sel. Top. Appl., 8, 1799–1807, 2015.

Karvonen, J.: Baltic sea ice concentration estimation using Sentinel-1 SAR and AMSR2 microwave radiometer data, IEEE T. Geosci. Remote, 55, 2871–2883, https://doi.org/10.1109/TGRS.2017.2655567, 2017.

Keller, M. R., Gifford, C. M., Walton, W. C., and Winstead, N. S.: Ice analysis based on active and passive radar images, U.S. Patent 9652674 B2, May 16, available at: https://patents.google.com/patent/US9652674 (last access: 18 August 2020), 2017.

Kohavi, R.: A Study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14th international joint conference on Artificial intelligence, Montreal, Canada, 20–25 August 1995, 2, 1137–1143, 1995.

Leigh, S., Wang, Z., and Clausi, D. A.: Automated ice-water classification using dual polarization SAR satellite imagery, IEEE T. Geosci. Remote, 52, 5529–5539, https://doi.org/10.1109/TGRS.2013.2290231, 2014.

Liu, H., Guo, H., and Zhang, L.: SVM-based sea ice classification using textural features and concentration from RADARSAT-2 dual-pol ScanSAR data, IEEE J. Sel. Top. Appl., 8, 1601–1613, https://doi.org/10.1109/JSTARS.2014.2365215, 2015.

Lohse, J., Doulgeris, A. P., and Dierking, W.: Mapping sea ice types from Sentinel-1 considering the surface-type dependent effect of incidence angle, Ann. Glaciol., https://doi.org/10.1017/aog.2020.45, 2020.

Louppe, G.: Understanding random forests: From theory to practice, PhD Thesis, U. of Liege, University of Liège, 123–144, 2014.

Mahmud, M. S., Geldsetzer, T., Howell, S. E. L., Yackel, J. J., Nandan, V., and Scharien, R. K.: Incidence angle dependence of HH-polarized C- and L-band wintertime backscatter over Arctic sea ice, IEEE T. Geosci. Remote, 56, 6686–6698, https://doi.org/10.1109/TGRS.2018.2841343, 2018.

Mäkynen, M. and Karvonen, J.: Incidence angle dependence of first-year sea ice backscattering coefficient in Sentinel-1 SAR imagery over the Kara Sea, IEEE T. Geosci. Remote, 55, 6170–6181, https://doi.org/10.1109/TGRS.2017.2721981, 2017.

Mäkynen, M. P., Manninen, A. T., Simila, M. H., Karvonen, J. A., and Hallikainen, M. T.: Incidence angle dependence of the statistical properties of C-band HH-polarization backscattering signatures of the Baltic Sea ice, IEEE T. Geosci. Remote, 40, 2593–2605, https://doi.org/10.1109/TGRS.2002.806991, 2002.

Miranda, N.: S-1 constellation product performance status, SeaSAR 2018, Frascati, Italy, 7–10 May 2018, available at: http://seasar2018.esa.int/files/presentation216.pdf (last access: 18 August 2020), 2018.

Ocean and Sea Ice Satellite Application Facilities: available at: http://osisaf.met.no/, last access: 19 August 2020.

Park, J.-W., Korosov, A. A., Babiker, M., Sandven, S., and Won, J.-S.: Efficient thermal noise removal for Sentinel-1 TOPSAR cross-polarization channel, IEEE T. Geosci. Remote, 56, 1555–1565, https://doi.org/10.1109/TGRS.2017.2765248, 2018.

Park, J.-W., Won, J.-S., Korosov, A. A., Babiker, M., and Miranda, N.: Textural noise correction for Sentinel-1 TOPSAR cross-polarization channel images, IEEE T. Geosci. Remote, 57, 4040–4049, https://doi.org/10.1109/TGRS.2018.2889381, 2019.

Partington, K., Flynn, T., Lamb, D., Bertoia, C., and Dedrick, K: Late twentieth century Northern Hemisphere sea-ice record from the U.S. National Ice Center ice charts, J. Geophys. Res., 108, 3343, https://doi.org/10.1029/2002JC001623, 2003.

Pastusiak, T.: Accuracy of sea ice data from remote sensing methods, its impact on safe speed determination and planning of voyage in ice-covered areas, International Journal on Marine Navigation and Safety of Sea Transportation, 10, 229–248, https://doi.org/10.12716/1001.10.02.06, 2016.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, https://doi.org/10.1016/j.patcog.2011.04.006, 2011.

Ressel, R., Frost, A., and Lehner, S.: A neural network-based classification for sea ice types on X-band SAR images, IEEE J. Sel. Top. Appl., 8, 3672–3680, https://doi.org/10.1109/JSTARS.2015.2436993, 2015.

Richards, F. J.: A flexible growth function for empirical use, J. Exp. Bot., 10, 290–300, https://doi.org/10.1093/jxb/10.2.290, 1959.

Scharien, R. K., Landy, J., and Barber, D. G.: First-year sea ice melt pond fraction estimation from dual-polarisation C-band SAR – Part 1: In situ observations, The Cryosphere, 8, 2147–2162, https://doi.org/10.5194/tc-8-2147-2014, 2014

Scheuchl, B., Flett, D., Caves, R., and Cumming, I.: Potential of RADARSAT-2 data for operational sea ice monitoring, Can. J. Remote Sens., 30, 448–471, 2004.

Shokr, M. E.: Evaluation of second-order texture parameters for sea ice classification from radar images, J. Geophys. Res.-Ocean., 96, 10625–10640, 1991.

Smedsrud, L. H., Halvorsen, M. H., Stroeve, J. C., Zhang, R., and Kloster, K.: Fram Strait sea ice export variability and September Arctic sea ice extent over the last 80 years, The Cryosphere, 11, 65–79, https://doi.org/10.5194/tc-11-65-2017, 2017.

Soh, L.-K. and Tsatsoulis, C.: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, IEEE T. Geosci. Remote, 37, 780–795, https://doi.org/10.1109/36.752194, 1999.

U.S. National Ice Center: available at: https://www.natice.noaa.gov/, last access: 19 August 2020.

Wang, L., Scott, K. A., and Clausi, D. A.: Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network, Remote Sens.-Basel, 9, 408, https://doi.org/10.3390/rs9050408, 2017.

WMO (World Meteorological Organization): Sea-ice information services in the world, WMO-No. 574, World Meteorological Organization, Geneva, Switzerland, 2017.

Zakhvatkina, N. Y., Alexandrov, V. Y., Johannessen, O. M., Sandven, S., and Frolov, I. Y.: Classification of sea ice types in ENVISAT synthetic aperture radar images, IEEE T. Geosci. Remote, 51, 2587–2600, https://doi.org/10.1109/TGRS.2012.2212445, 2013.

Zakhvatkina, N., Korosov, A., Muckenhuber, S., Sandven, S., and Babiker, M.: Operational algorithm for ice-water classification on dual-polarized RADARSAT-2 images, The Cryosphere, 11, 33–46, https://doi.org/10.5194/tc-11-33-2017, 2017.

Zakhvatkina, N., Smirnov, V., and Bychkova, I.: Satellite SAR data-based sea ice classification: An overview, Geosci. J., 9, 152, https://doi.org/10.3390/geosciences9040152, 2019.