Classification of Winter Sea Ice Types in Sentinel-1 SAR images

Jeong-Won Park[1,3], Anton A. Korosov[1], Mohamed Babiker[1], Joong-Sun Won[2], Morten W. Hansen[1], Hyun-c~~C~~heol Kim[3]

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

Correspondence to: Jeong-Won Park (orepaku.andoid@gmail.com)

Abstract. A new Sentinel-1 image-based sea ice classification algorithm is proposed to support ~~automated~~ daily ice charting by using a machine learning-based model trained in a semi-automated manner. Previous studies mostly rely on manual work in selecting training and validation data. We show that the use of readily available ice charts from an operational ice services ~~allow to automate selection~~can reduce the amount of manual works in preparation of large amount of training/testing data. Furthermore, it reduces the ~~void of biased, subjective~~inconsistent decisions in the classification algorithm by indirectly exploiting the best ability of the sea ice experts working at operational ice services. The proposed scheme has two phases: training and operational. Both phases start from removal of thermal, scalloping and textural noise from Sentinel-1 data and calculation of gray 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 in operational manner. Test results from two winter season dataset acquired over the Fram Strait and Barents Sea area showed that the classifier is capable of retrieving three~~3~~ generalized cover types (~~ice free~~open water, ~~integrated~~ mixed first-year ice, old ice) with overall accurac~~ies~~y of 85% and five~~5~~ cover types (~~ice free~~open water, new ice, young ice, first-year ice, old ice) with accuracy of 58%. The errors are attributed both to incorrect manual classification on the ice charts and to the semi-automated algorithm. We demonstrate the potential for near-real time service of ice type classification through an example of ice maps made from daily mosaiced 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 region. 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.

However, most of the recent ice classification algorithms were developed using RADARSAT-2 ScanSAR images (Leigh et al., 2014; Liu et al., 2015; Zakhvatkina et al., 2017) which has different sensor characteristics 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 Sentiel-1 data is the relatively high level of thermal noise contamination and its propagation to image textures.

For a proper use of dense time-series of Earth observations, radiometric properties must be well calibrated. Thermal noise is often neglected in many applications but is impacting seriously the utility of dual-polarization SAR data. Sentinel-1 TOPSAR image intensity is disturbed by the thermal noise particularly in the cross-polarization channel. Although the European Space Agency (ESA) provides calibrated noise vectors for noise power subtraction, residual noise contribution is 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 thus degrades texture information in the Sentinel-1 TOPSAR images.

In ~~most~~ many of the previous works 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 can be ~~subjective~~inconsistent, and the number of samples were not enough to generalize the results because of the laborious manual work. 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 raw interpretation of SAR images by trained ice analysts working at operational ice services, were tested for sea ice concentration estimation by Wang et al. (2017); however, such image analysis charts are not accessible to public.

The use of public ice chart as training and validation reference data may help in solving the validation problem and enabling automation. The preparation of 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. 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 as the best available information of 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 certified ice analysts is mimicked.

In this work, we present a semi-automated Sentinel-1 image-based sea ice classification algorithm which takes an 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 random forest classifier by relating with ice charts published by operational ice services. The use of ice charts has dual purposes: semi-automatization of classifier training, and minimization of human error.

2 Data and methods

2.1 Study area and used data

The region of study ~~interest~~ for developing and testing the proposed algorithm is the Fram Strait and Barents Sea including a part of the Arctic Ocean (10ºW-70ºE, 75ºN-85ºN) as shown in Figure 1. Various sea ice types ~~coexist~~ are found in this area due to 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 images acquired in winter season for two~~2~~ years (Dec. 2017 – Mar. 2018 and Dec. 2018 – Mar. 2019) were collected from the Copernicus Open Access Hub (https://scihub.copernicus.eu). The number of daily image acquisitions covering the study area ranges from 6 to 10 depending on the orbits. The images from the first year (hereafter called 2018 data) is used to train the classifier and those from the second year (hearafter called 2019 data) is 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: U. S. National Ice Center (NIC) of the United States of America, and 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).

2.2 Methods

Figure 2 shows the flow of the semi-automated ice classification scheme that we propose. It is divided into two phases: training and operational. Both phases start from removal of thermal noise from Sentinel-1 data (Section 2.2.2), incidence angle calibration (Section 2.2.3) and calculating texture features (Section 2.2.4). The training phase (show by gray on Figure 2) continues with preprocessing and collocation of the ice charts with the Sentinel data (Section 2.2.1) and machine learning step (Section 2.2.5 and 2.2.6). The operational phase uses the classifier developed during the training phase for processing texture features computed from the input SAR data and for generating ice charts. Detailed explanations for each steps are given in the following subsections.

2.2.1 Ice chart preprocessing

To take ~~the~~ advantage of the objective identification of the ice type and to develop a semi~~n~~-~~-~~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 more than few meters, hence the reprojection of coordinates of ice chart polygons is done with Geospatial

Data Abstraction Library (GDAL; GDAL/OGR contributors, 2019) using a simple 3rd order polynomial fitted using the ground control points 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 seen like water in a SAR image. Heinrichs et al. (2006) reported that the ice edge determined from ~~the~~ AMSR-E passive microwave radiometer~~, which is a passive microwave radiometer,~~ data using the isoline of 15% concentration matches best the ice edge determined from RADARSAT-1 SAR data using visual inspection~~, which is a C-band HH-polarization SAR~~. After visual comparison of many SAR backscattering images and the corresponding reprojected ice charts, we set a threshold of 20% for CT to discard water-like pixels. Note that ice concentration label in the SIGRID-3 format is assigned in increments of 10%~~has precision of decimals~~. CP is also important to find the dominant ice type in the given polygons. SoD is so-called ice type. It is challenging to differentiate ice types using SAR data only, thus we simplified the SoDs by merging into five classes: ~~ice free~~open water, new ice, young ice, first-year ice, and old ice. ~~Bergy water is treated as open water since its ice concentration (by definition, less than 10%) is below the threshold that we set (i.e., 20%).~~

Figure 3 demonstrates an example of the ice chart preprocessing explained above with the colors following the WMO nomenclature (JCOMM, 2014b). Comparing the original SoD i~~I~~n the top left panel with the processed SoD in the bottom left panel, it is clear that the ice edge of the processed SoD match better with the SAR backscattering images.

2.2.2 Denoising of Sentinel-1 imagery

Sentinel-1 cross-polarization images suffer from strong noise originated 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 ~~ocean~~ open water and level sea ice without 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 preprocessing of Sentinel-1 TOPSAR dual polarization products. Denoising ensures beam-normalized texture properties for all subswaths, which helps seamless 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 Senitnel-1 images was denoised before further processes are applied.

2.2.3 Incidence angle correction

It is well known that there is a strong incidence angle dependency in the SAR backscattering intensity for ~~ocean~~ open water and sea ice surface (Mäkynen et al., 2002; Mäkynen and Karvonen, 2017). For 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 for typical first-year ice are reported as -0.24 and -0.16 dB/degree for HH- and HV-polarization, respectively (Mäkynen and Karvonen, 2017). To normalize the backscattering intensity for all swath range, these slopes are compensated for or used as input layer in ~~serveral~~several ice classification algorithms in 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 has updated the noise removal scheme in 2018 (Miranda, 2018).

Figure 4~~.~~ shows two-dimensional histograms of incidence angle versus sigma nought for sea ice pixels in HH and HV polarization channels from Sentinel-1 data collected over sea ice and open water in the study area in winter 2018. From the Sentinel-1 dataset described in Section 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 (OSISAF). For HH polarization, the estimated slope was -0.200 dB/degree, which is slightly different from the estimation of the first-year ice (-0.24 dB/degree) in Mäkynen and Karvonen (2017) and in between the estimations for first-year ice (-0.22 dB/degree) and multi-year ice (-0.16 dB/degree) in Mahmud et al. (2018). For HV polarization, the estimated slope was only -0.025 dB/degree, which is much lower than the estimation in Mäkynen and Karvonen (2017), however, it is in line with the estimations from RADARSAT-2 (Leigh et al., 2014; Liu et al., 2015). We compensate the incidence angle dependency using the estimated slopes, referencing to the nominal scene center angle of 34.5 degree. 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/degree for wind velocity of 5 m/s, 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~~; Karvonen, 2017; Zakhvatkina et al., 2013, 2017~~), the proposed approach starts from gray level co-occurrence matrices (GLCM) calculation. The GLCM is a four-dimensional matrix $P(i, j, d, a)$ calculated from the two grey tones of reference pixel i, and its neighbor j, with co-occurrence distance d, and ~~direction~~ orientation a. Haralick et al. (1973) has introduced a set of GLCM-based texture features called Haralick features, and the usefulness of it has been reported in several literatures. Since the 13 Haralick features can be calculated for each of the two-dimensional slices $P(i, j)$ for multiple d and a, the maximum number of texture features is to be as $2 \times 13 \times d \times a = 26da$, where 2 is for accounting dual polarization. It is common to take directional average for 0°, 45°, 90°, and 135° to reduce GLCM dimensionality. ~~Furthermore, f~~Further averaging for multiple distances (1 to $w/2$ where w is the size of subwindow for GLCM computation) is taken after computing normalized GLCM. The spatial resolution of the texture features is the pixel spacing of 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~~spatial resolution~~ is 1 km.

An important factor that influences the computed texture features is number of gray 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 the Figure 4 after incidence angle correction), the number of gray level 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.

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

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

Since L should be sufficiently large to take the full advantage of 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 the 13~~26~~ Haralick features, the coefficient of variation (CV) which is reported as useful feature for ice-water discrimination (Keller et al., 2017) is included. The CV is defined as follows:

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

where σ and μ are the standard deviation and mean of the samples in a given subwindow. Since CV can be computed for each polarization images, the number of texture features for Sentinel-1 dual polarization product is extended to 28. Other features can be added are incidence angle and day of year. 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 image textures, they can be used for classification as input features. Note that it is important to have spatially and temporally even distributions for each of the ice types if these two additional features are included, otherwise, the trained classifier will give 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. In Fernández-Delgado et al. (2014), the Random Forest (RF; Ho, 1998) was evaluated as the best classifier over various types of dataset, but the difference with the second best, Support Vector Machine (SVM; Cortes and Vapnik, 1995), was not statistically significant. In the literatures about sea ice classification, the SVM was used often because by nature it works relatively well when the~~for~~ ~~sparse~~ number of ~~dataset~~datasets are small. When training dataset is prepared by manual work (i.e., manual classification by human expert), the number of images is not large, usually less than 20 (e.g., 12 scenes in Zakhvatkina et al., 2013; 20 scenes in Leigh et al., 2013; 2 scenes in Liu et al., 2015; 4 scenes in Ressel et al., 2015). However, when the readily available ice charts are used as training reference, many more images become available with less effort, thus there is no need to rely on additional manual work prone to contamination by biased decisions. The RF has two practical advantages when processing large number of dataset. First, the RF is scale invariant. It can use the data as is, while the SVM requires preprocessing of scaling and normalization for the input features. Second, the computational complexity for the RF is lower than that of the SVM. For the SVM, the number of operations ~~they~~ are $O(n^2p + n^3)$ and $O(n_{sv}p)$ for training and prediction while for RF, $O(n^2pn_{tr})$ and $O(n_{tr}p)$, respectively, where n is the number of samples, p is the number of features, n_{sv} for the number of support vectors, and n_{tr} for 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 one-vs-all scheme (Anand et al., 1995). Although the standard RF algorithm can inherently deal with multiclass problem, the one-vs-all binarization to the RF results in better accuracy with smaller forest size than the standard RF (~~Ramírez et al., 2018~~Adnan and Islam, 2015).

Three hyper parameters of the random forest 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 with five~~5~~-fold cross-validation (Kohavi, 1995) is used. The grid (all possible combinations of N_T, D, and N_F values) is set in logarithmic scale (see values in Table 1) because the performance change with hyperparameter is typically in 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. T~~t~~he Richard's Curve (Richard, 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 increment, difference between training and testing score, 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 ice chart including reprojection into the SAR image geometry, only the samples with spatially and temporally good match should be fed to the training phase. 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 three~~3~~ days from the date of publication, 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 needs to be identified first. ~~To automate image selection for training, a good ice/water classifier for SAR image is needed.~~ Since even ~~such~~ an ~~simple binary~~ 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 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 by selecting the pixel borders of open water class. Then the SAR backscattering image contrasts across the ice-water boundaries are examined both in HH- and HV-polarization because the image contrast between ice-water is larger in HV but smooth level ice is better recognizable in HH.

After the image selection, the samples in the selected images are split randomly into training and test dataset with a ratio of 7:3. For 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, etc.). In this study, only the data from pixels more than 3 km away from the polygon boundaries was 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, w~~W~~e use confusion matrix and Cohen's kappa coefficient κ (Cohen, 1960), which measures the agreement between two raters (in this study, they are the trained classifier and the reference ice chart) with taking account of the possibility of the agreement occurring by chance~~ for performance evaluation~~. The validation is done in the same way but using a completely independent dataset. The 2018 data was used to run the training phase. Among 958 images in total, we selected 57 images of which ice edges match well with the collocated ice chart. From the selected images, 6.4 million samples covering open water and sea ice were divided into training and test dataset.

3 Results and discussion

~~The 2018 data was used to run the training phase. Among 958 images in total, we manually selected 57 images of which ice edges match well with the collocated ice chart. From the selected images, 6.4 million pixels covering open water and sea ice were divided into training and test dataset.~~ We trained three RF classifiers with different feature configurations: i) FC1: ~~texture features from~~ Haralick texture features and CV, ii) FC2: Haralick texture features, CV, and ~~texture features and~~ incidence angle, iii) FC3: Haralick texture features, CV~~texture features~~, incidence angle, and day of year.

As expected, the classification score increases with the number of trees (crosses on Figure 5, upper panel) and the Richard's curve (dashed line) fits well to the observations (RMSE=2.3×10^{-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 on Figure 5, middle panel) but the Richard's curve (dashed line) doesn't fit so well (RMSE=3.6×10^{-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 (Figure 5, lower panel): small difference between the scores (for $D \leq 8$) indicate similar performance on training and testing datasets, while large difference (for $D > 8$) indicate that testing dataset is processed with worse results. The optimal D value is therefore selected where the score difference become higher than 0.03 and constitutes 8 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 of the hyper parameter estimation, 11, 8, and 10 were selected as the optimal values for the number of trees, the maximum tree depth, and the number of features, respectively.

The trained five~~5~~-class classifier consists of five~~5~~ binary sub-classifiers, each of them is used for discriminating one specific class from the others. For each sub-classifier, each of the texture features has different weight in decision making. The fraction of the samples that each of texture features contribute to 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 for the sub-classifiers is presented in Figure 6. Overall pattern shows that the features of HV polarization plays 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, variance- and entropy-related features were more important. The classifiers for ~~ice-free~~open water and old ice have more strong dependencies on HV polarization than others. This is understandable because the main radar scattering mechanisms for those two types are strongly characterized 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 distinctive pattern ~~so~~ 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 introduces high surface roughness and volume scattering (Isleifson et al., 2014), thus the new ice can appear either featureless dark or complex bright in SAR image (Dierking, 2010). The large range in backscatter values makes it hard to define characteristic texture in the new ice patch.

~~there is no characteristic texture in the new ice patch; typically, they look just dark in SAR image.~~

The confusion matrix for testing the trained classifier with the test dataset (2018 data) is shown is in Table 2. The three cases with different feature configurations (FC1-FC3) were tested. The accuracies for ~~ice-free~~open water and old ice were higher than 85%, however, those for young ice and first-year ice were around 60%. The mean difference between the results of FC1 and FC2 was only 1.6%, which indicates that residual angular dependency after the incidence angle correction was insignificant. However, when comparing the results of FC2 and FC3, there were notable accuracy improvements, especially for new ice (24.5%). κ for FC1, FC2, and FC3 were 0.70, 0.71, and 0.77, respectively. When the evaluation is carried out with the 2018 data, the training and test datasets~~Since the training and test datasets~~ were extracted from the same selected images, thus sharing the same overall ice conditions and spatial/temporal coverage, there might be correlation which is not preferable for proper evaluation. Table 3 shows the confusion matrix for validation results from 2019 data. Comparing to the results from 2018 data, the high accuracies for ~~ice-free~~open water and old ice were maintained in similar level, while those for new ice, young ice, and first-year ice were decreased considerably. The differences between the results of FC1 and FC2 were insignificant, while there were notable accuracy decreases from FC2 to FC3, which is the opposite to the result in Table 2. This means that the training with FC3 was overfitted, and the day of year may not ~~help ice type classification if the training does not cover the whole seasonal cycle.~~correspond to the temperature, air-sea fluxes, and weather regimes. κ for FC1, FC2, and FC3 were 0.67, 0.67, and 0.49, respectively.

To see how the denoising step in Section 2.2.2 led to improvements in the classification accuracies, the same training and evaluation was conducted for the same dataset without applying the textural noise correction, and Table 4 shows the results. In both FC1 and FC2, the improvements in accuracies for young ice (+8.2-9.8%) and first-year ice (+9.2-11.6%) were most pronounced compared to those for open water (+1.7%) and old ice (+1.2-1.7%). On the contrary, a small decrease was observed for new ice (-2.8-4.7%). Nevertheless, the improvement in kappa (+0.05) demonstrates clear improvement in the overall classification result.

Figure 7 shows a daily mosaic of Sentinel-1 SAR images over the study area and the classified ice map. For comparison, the NIC weekly ice chart is also displayed. The ice edges of ice chart match well with the SAR mosaic in most parts probably because the same SAR data was used when the ice chart was made, although the SAR images had been acquired three~~3~~ days before the ice chart was published. Overall, the discriminations between ice and non-ice, old ice and other ice types, and detection of new ice patch look reasonable. However, some young ice patches, for example the ice patch between the Svalbard archipelago, are misclassified as the first-year ice, and vice versa. Figure 8 shows another daily mosaic made by the images acquired on the same day of ice chart publication. Comparing the backscattering images in Figure 7 and Figure 8, there was notable ice drift. The SAR-based ice classification results in both figures look 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 date of weekly ice chart publication can be largely different from that in the ice chart. This example shows a clear potential of near-real time service of ice type classification.

To cope with the ambiguous classification for the ice types with low accuracy, we conducted a test with further simplification of ice types by combining the new ice, young ice, and first-year ice into the "~~integrated~~mixed" first-year ice, and then training new classifiers. Table 5~~4~~ and Table ~~5~~ 6 show the confusion matrices for the three~~3~~-class classifiers. κ for FC1, FC2, and FC3 were 0.84, 0.86, and 0.92 in 2018 data, and 0.80, 0.80, and 0.53 in 2019 data, respectively. The dramatic increase in the accuracy for the ~~integrated~~ mixed first-year ice class indicates that the misclassification for the new ice, young ice, and first-year ice were mostly among themselves. However, the accuracy decrease from 2018 data to 2019 data was in similar level to the case of the five~~5~~-class classification, and this could have been caused by ~~insufficient training of classifier and/or~~ inconsistent labeling in the reference ice chart.

Figure 9 shows an example of the inconsistent labeling in the reference ice chart. The SoDs from the NIC ice charts are superimposed on the Sentinel-1 backscattering images. The type of the same ice floe (red outline) is annotated differently in the two ice charts (old ice on the left panel and first-year ice on the right panel), while it looks almost the same in the SAR backscattering images. Considering the ice edges in ice charts match well with those in the SAR backscattering images, thus the ices in the inner parts are also expected to be ~~charted in~~mapped at the same time, it should be noted that training with ice chart might have included ~~wrong~~ mislabeled small features ~~samples~~ 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 the 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.

The proposed algorithm has several limitations as follows. First, the variations in radar backscattering and its corresponding image textures due to seasonal changes were not properly captured. Although the day of year was tested as a seasonality variable in the FC3 feature configuration, the result did not show any improvement. This is because day of year might not correspond to the same temperature, fluxes, and weather regimes. Second, the proposed method struggles with SAR image edges that the same sea ice on different side of an image edge was classified differently. This impose that the incidence angle dependence could not be normalized perfectly. An example of such a failure can be seen along the image boundaries at 80N, 35E and 82.5N, 60E, approximately. Third, some artifacts are observed under an extreme sea state. In the classified results in the bottom right panel of Figure 8, there is a misclassified FYI patch (yellow) in the open water area. According to the NOAA SAR wind image service, ANSWRS 2.0, the wind speed ranged from 17 to 21 m/s at the time of image acquisition, thus the water surface might be heavily roughened by strong wind.

4 Conclusion

A new semi-automated SAR-based sea ice type classification scheme was proposed in this study. For the first time several ice types can be successfully identified on Sentinel-1 SAR imagery. The main technological innovation is two-fold: i) minimized manual work in the preparation of training and validation reference data, ii) more objective evaluation of the SAR-based sea ice type classifier. A conventional approach for selecting training/testing data by annonymous~~a~~ human ice expert is undesirable not only because it is laborious, but also due to subjectivity and lack of standardization in assessment of the automated classifier. Therefore, the performance from different literature sources cannot be intercompared directly.

Test results from two winter season dataset acquired over the Fram Strait and Barents Sea area showed overall accuracies of 85% and 58% and κ of 0.80 and 0.67 for the three~~3~~-class and five~~5~~-class ice type classifiers, respectively. These are slightly lower than the numbers in the previous studies, and the errors are attributed not only to the automated algorithm but also to the inconsistency of ice charts and the high level of their generalization. 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 EUMETSAT OSI-403-C (Aaboe et al., 2014) with higher spatial resolution. The proposed approach importantly showed that a daily ice type mapping from the Sentinel-1 data is feasible, and it can help ~~capturing~~ to capture ~~more~~ details in 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.

Acknowledgements

This work was supported by the French Service Hydrographique et Océanographique de la Marine (SHOM) under SHOM-ImpSIM Project, 111222, and the Research Council of Norway and the Russian Foundation for Basic Research under NORRUSS Project 243608, SONARC.

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.

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(76pt2), 112-123. doi: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(1), 117-124, doi: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(4), 385-395, 1991.

Cohen, J.: A coefficient of agreement for nominal scales, Educ. Psychol. Meas., 20(1), 37–46, doi:10.1177/001316446002000104, 1960.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20(3), 273-297, doi: 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(3), 528-538, doi: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(3), 1045-1058, doi:10.1109/TGRS.2009.2031806, 2010.

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

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, 2019. URL https://gdal.org

Haralick, R. M., Shanmugam, K., Dinstein, I.: Textural features for image classification, IEEE T. SYST. MAN. CY.-S., SMC-3(6), 610-621, doi: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, doi: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(11), 3070-3080, doi: 10.1109/TGRS.2006.880622, 2006.

Ho, T. K.: The random subspace method for constructing decision forests, IEEE T. Pattern Anal., 20(8), 832-844, doi: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(3), 1787-1798, doi: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, doi: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, doi: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(4), 1799-1807, 2015.

Karvonen, J.: Baltic sea ice concentration estimation using Sentinel-1 SAR and AMSR2 microwave radiometer data, IEEE T. Geosci. Remote, 55(5), 2871-2883, doi: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, 2017. Available at: https://patents.google.com/patent/US9652674~~Keller, M.: Active/passive dual polarization sea ice detection, NISAR Applications Workshop: Sea Ice, Maryland, USA, 23 Jun. 2017. Available at: https://www.star.nesdis.noaa.gov/sod/mecb/sar/NISAR_Sea_Ice_Workshop/Keller_et_al_23_June_2017_Active_Passive_Sea_Ice_Detection.pptx~~

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(9), 5529–5539, doi: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(4), 1601–1613, doi:10.1109/JSTARS.2014.2365215, 2015.

Louppe, G.: Understanding random forests: From theory to practice, PhD Thesis, U. of Liege, 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(11), 6686-6698, doi: 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(11), 6170-6181, doi: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(12), 2593-2605, doi: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

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(3), 1555–1565, doi: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(6), 4040-4049, doi: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(C11), doi: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(2), 229-248, doi: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, doi:10.1016/j.patcog.2011.04.006, 2011.

~~Ramírez, J., Górriz, J. M., Ortiz, A., Martínez-Murcia, F. J., Segovia, F., Salas-Gonzalez, D., Castillo-Barnes, D., Illán, I. A., and Puntonet, C. G.: Ensemble of random forests One vs. Rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares, J. Neurosci. Meth., 302, 47-57, doi: 10.1016/j.jneumeth.2017.12.005, 2018.~~

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(7), 3672-3680, doi:10.1109/JSTARS.2015.2436993, 2015.

Richards, F. J.: A flexible growth function for empirical use, J. Exp. Bot., 10(29), 290-300, doi: 10.1093/jxb/10.2.290, 1959.

Scheuchl, B., Flett, D., Caves, R., and Cumming, I.: Potential of RADARSAT-2 data for operational sea ice monitoring, Can. J. Remote Sens., 30(3), 448-471, 2004.

Shokr, M. E.: Evaluation of second-order texture parameters for sea ice classification from radar images, J. Geophys. Res.-Oceans, 96(C6), 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, doi: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(2), 780-795, doi:10.1109/36.752194, 1999.

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(5), doi:10.3390/rs9050408, 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(5), 2587-2600, doi: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, doi:10.5194/tc-11-33-2017, 2017.

Figures

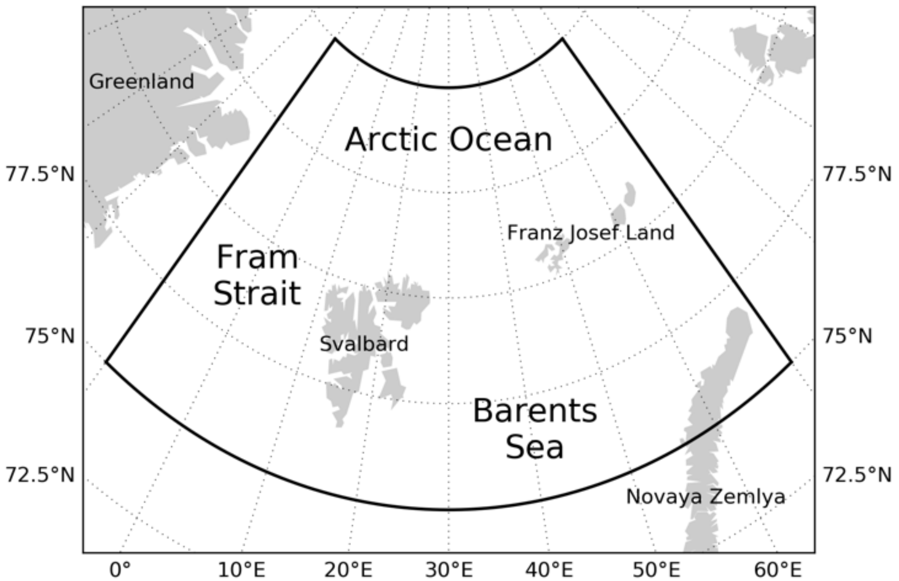


Figure 1: Study area.

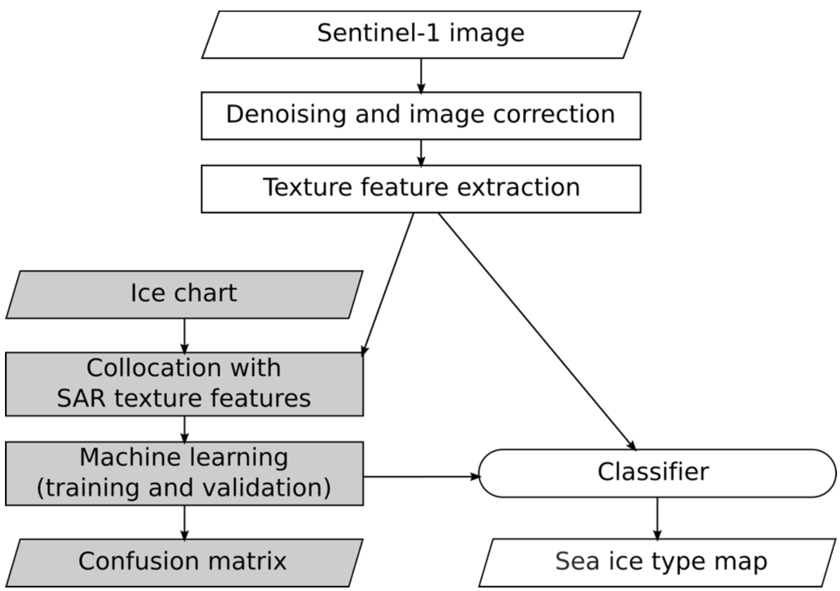
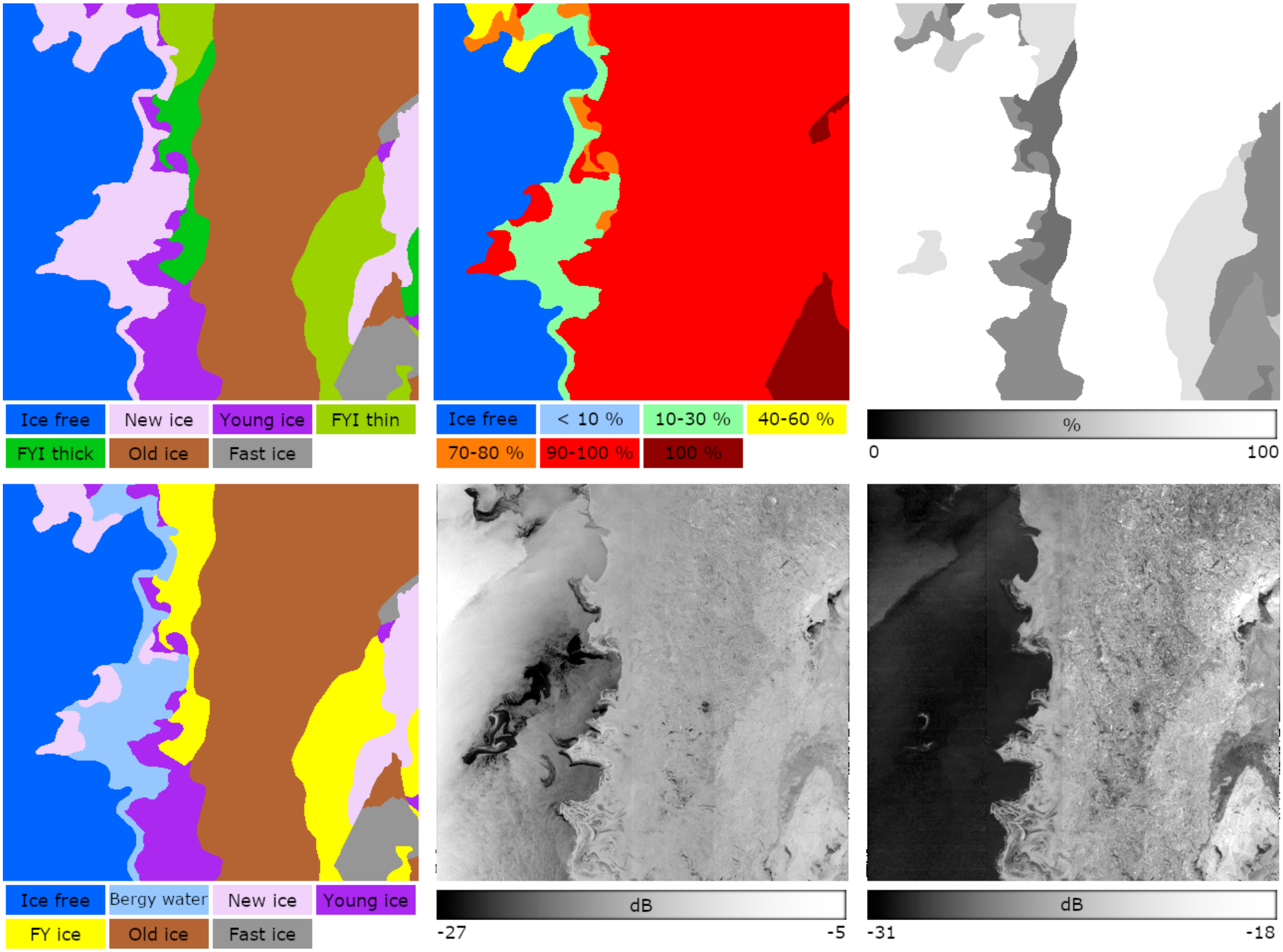


Figure 2: Processing flowchart of the proposed algorithm. Gray color shows training phase.


Ice free
New ice
Young ice
FYI thin
FYI thick
Old ice
Fast ice
Ice free
< 10 %
10-30 %
40-60 %
70-80 %
90-100 %
100 %
%
0
100
Ice free
Bergy water
New ice
Young ice
FY ice
Old ice
Fast ice
dB
-27
-5
dB
-31
-18

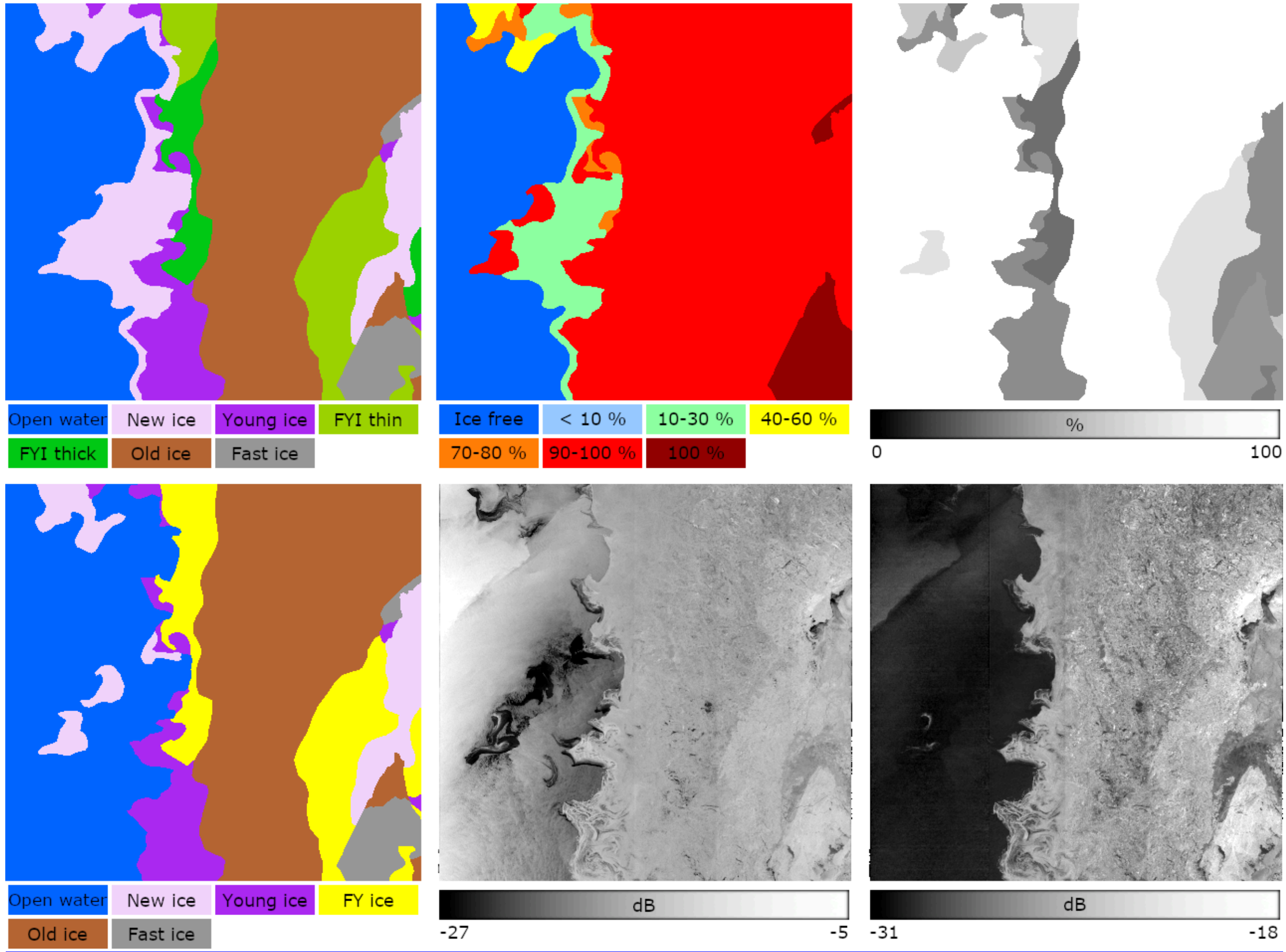
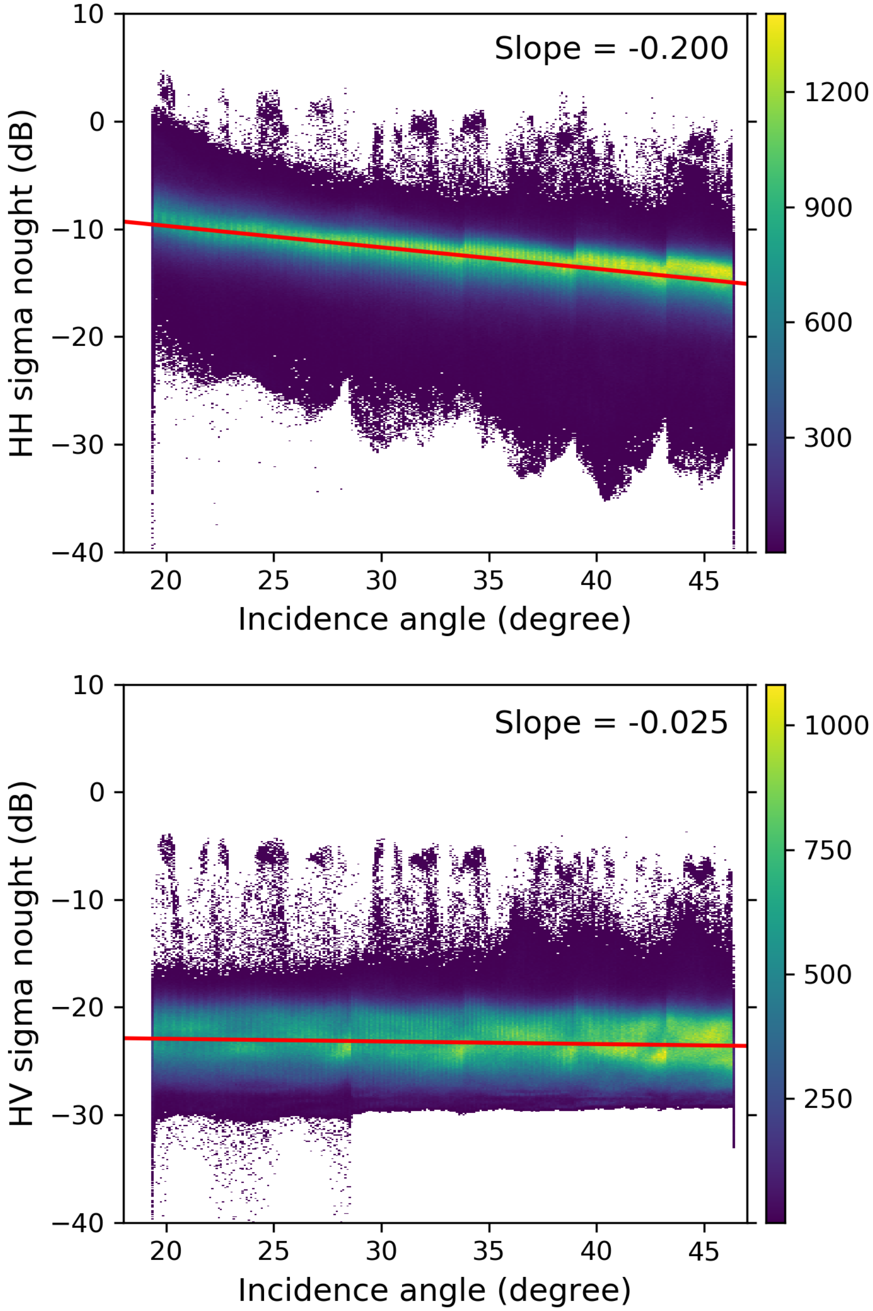


Figure 3: An example of ice chart preprocessing. From ice chart, stage of development (SoD; Top left), ice concentration (CT; Top center), and partial concentration of the dominant ice type (CP; Top right) maps are extracted. Then, some of different SoDs are merged (e.g., thin and thick first-year ices are merged into single label as first-year ice), and area with low ice concentration is labeled as ~~bergy ice~~open water. The processed map of SoD (Bottom left) is used to related with texture features extracted from HH and HV polarization images (Bottom center and bottom right). Note that the NIC ice chart published on January 25, 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.



Slope = -0.200
HH sigma nought (dB)
Incidence angle (degree)
Slope = -0.025
HV sigma nought (dB)
Incidence angle (degree)

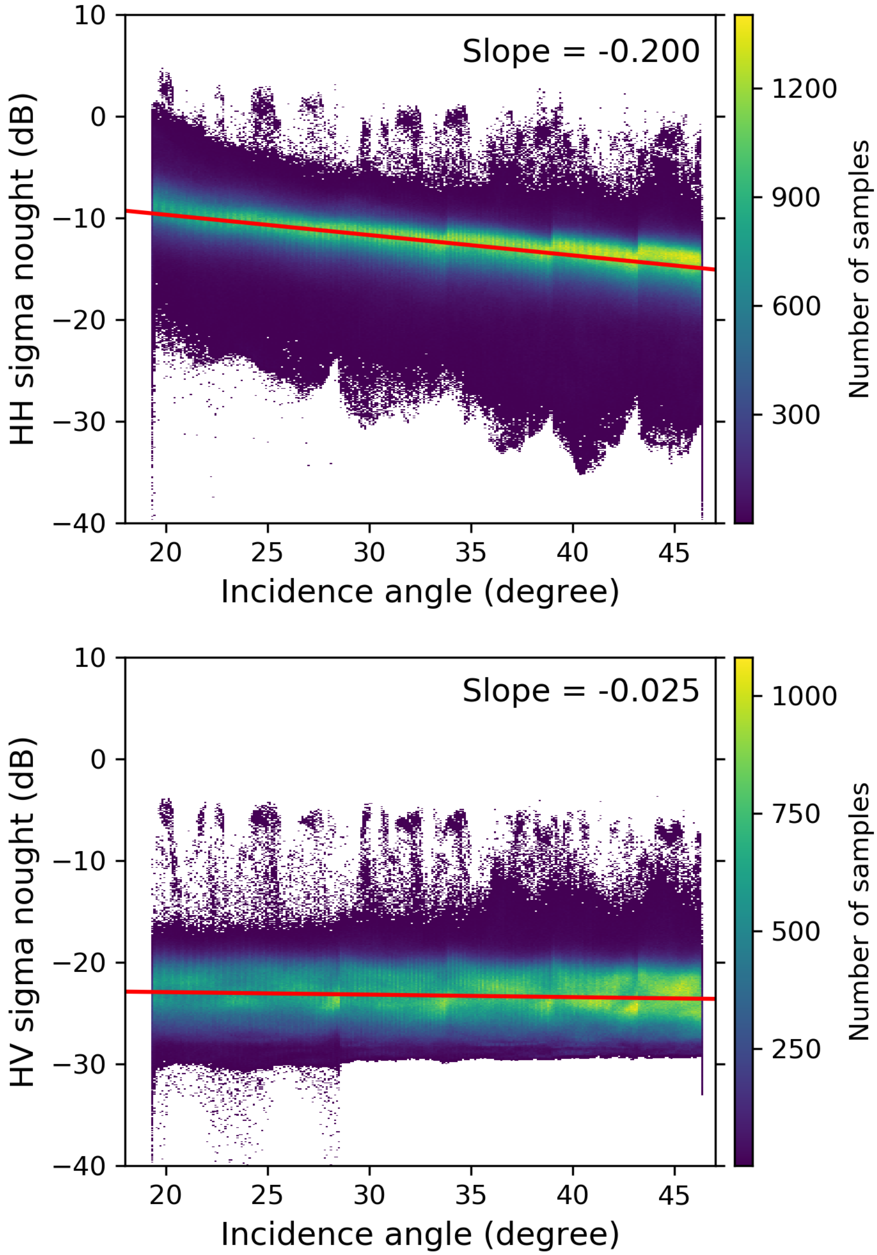


Figure 4: Two-dimensional histograms of incidence angle versus sigma nought for HH (top) and HV (bottom) polarization channels. Pixels covering various types of sea ice were merged so that the averaged property can be estimated. The best fit linear trends are shown with red lines.

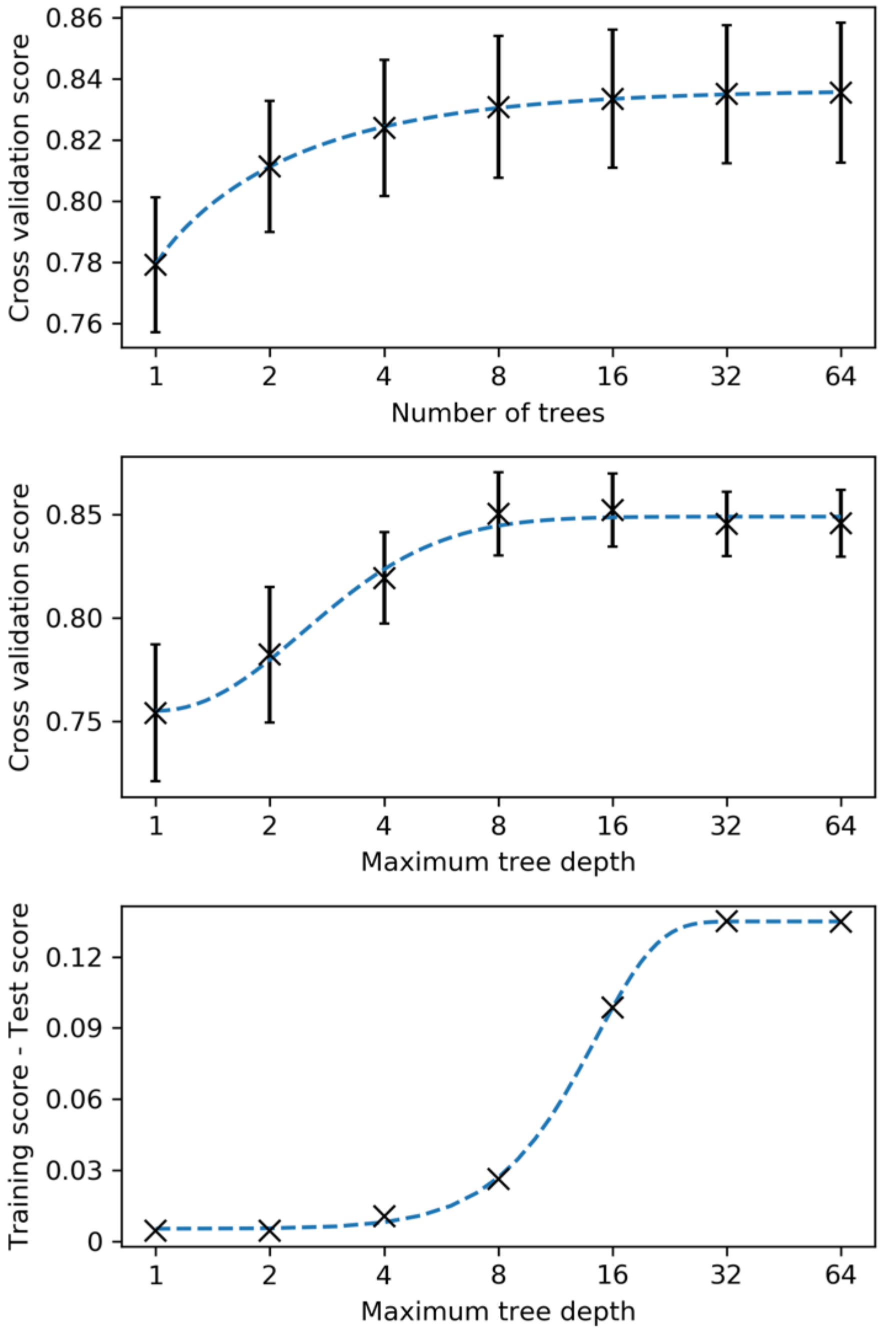


Figure 5: Hyperparameter optimization using grid search results (cross). Dashed lines represent the best-fit Richard's Curve. (Top panel) The optimal values are extracted from the locations where the score increments per unit of each hyperparameter become lower than a threshold (e.g., 0.001). (Center panel) If the curve does not fit the grid search results well, (Bottom panel) 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.

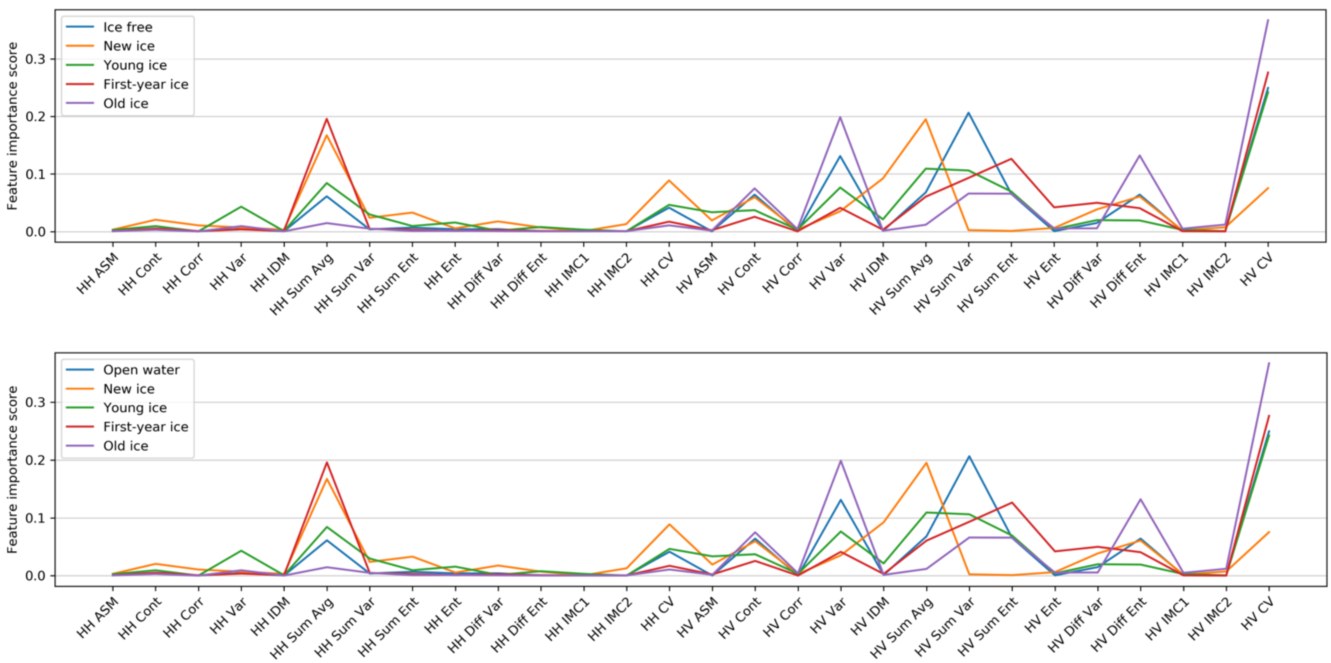
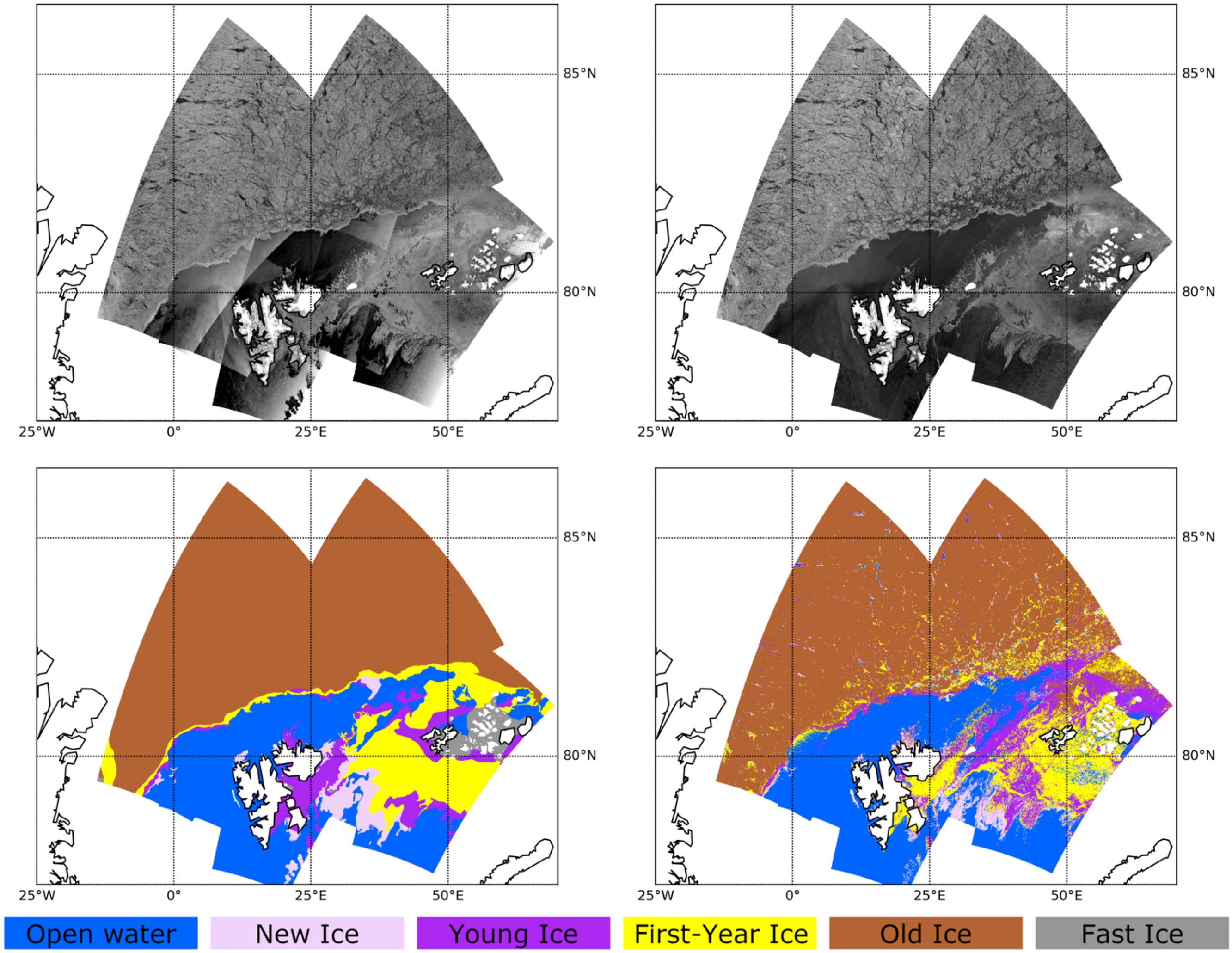


Figure 6: Feature importances of the binary sub-classifiers. ASM: angular second moment; Cont: contrast; Corr: correlation; Var: variance; IDM: inverse difference moment; Sum Avg: sum average; Sum Var: sum variance; Sum Ent: sum entropy; Ent: entropy; Diff Var: difference variance; Diff Ent: difference entropy; IMC: information measures of correlation; CV: coefficient of variation. For definitions of each parameters, please refer to Haralick et al., 1973.


85°N
80°N
25°W
0°
25°E
50°E
Open water
New Ice
Young Ice
First-Year Ice
Old Ice
Fast Ice

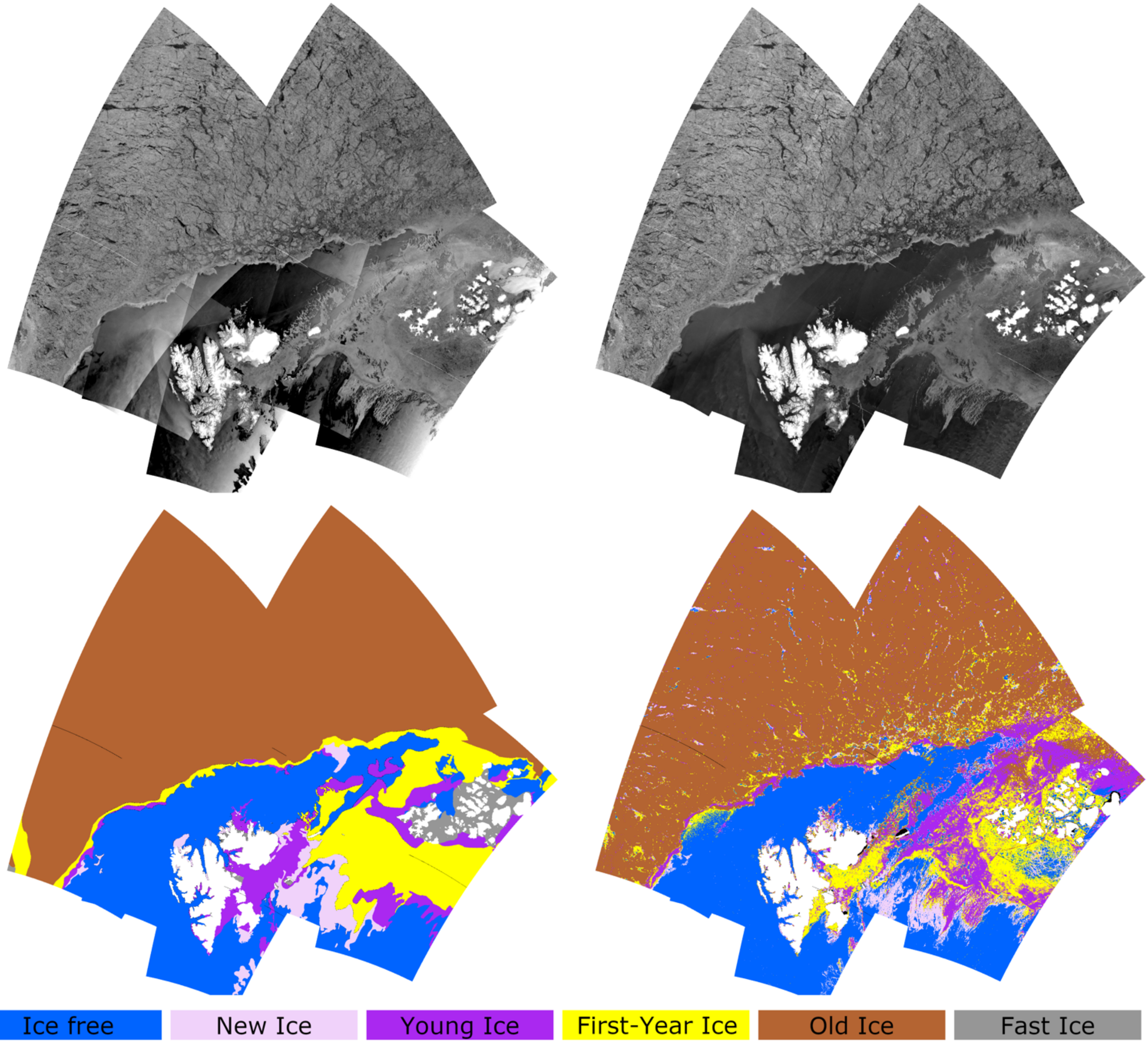
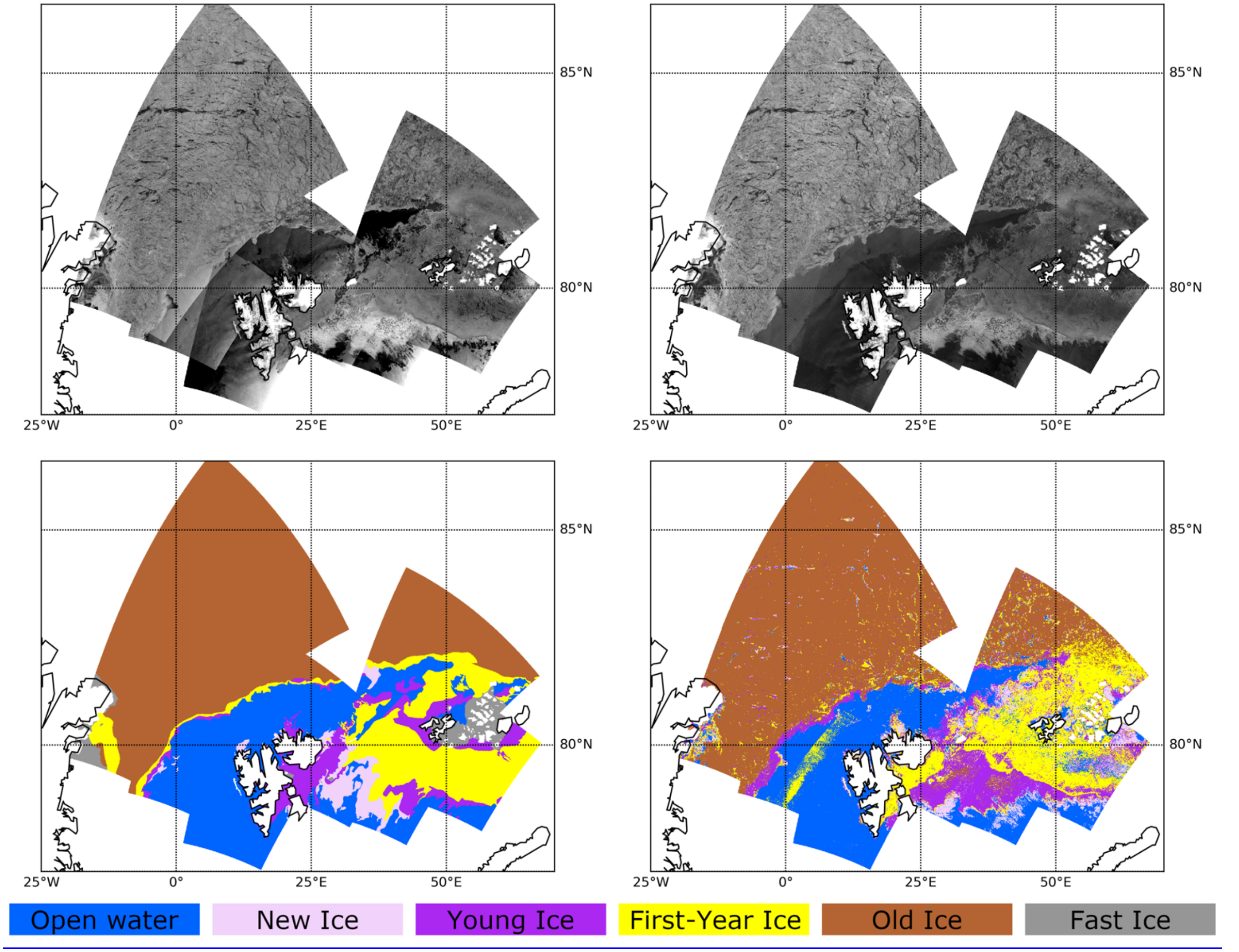


Figure 7: One~~1~~-day mosaics of Sentinel-1A/1B images (Top left: HH, Top right: HV) and the ice classification result (Bottom right) for 5 February 2019. The date of publication of the reference weekly ice chart (Bottom left) is 8 February 2019.


85°N
80°N
25°W
0°
25°E
50°E
Open water
New Ice
Young Ice
First-Year Ice
Old Ice
Fast Ice

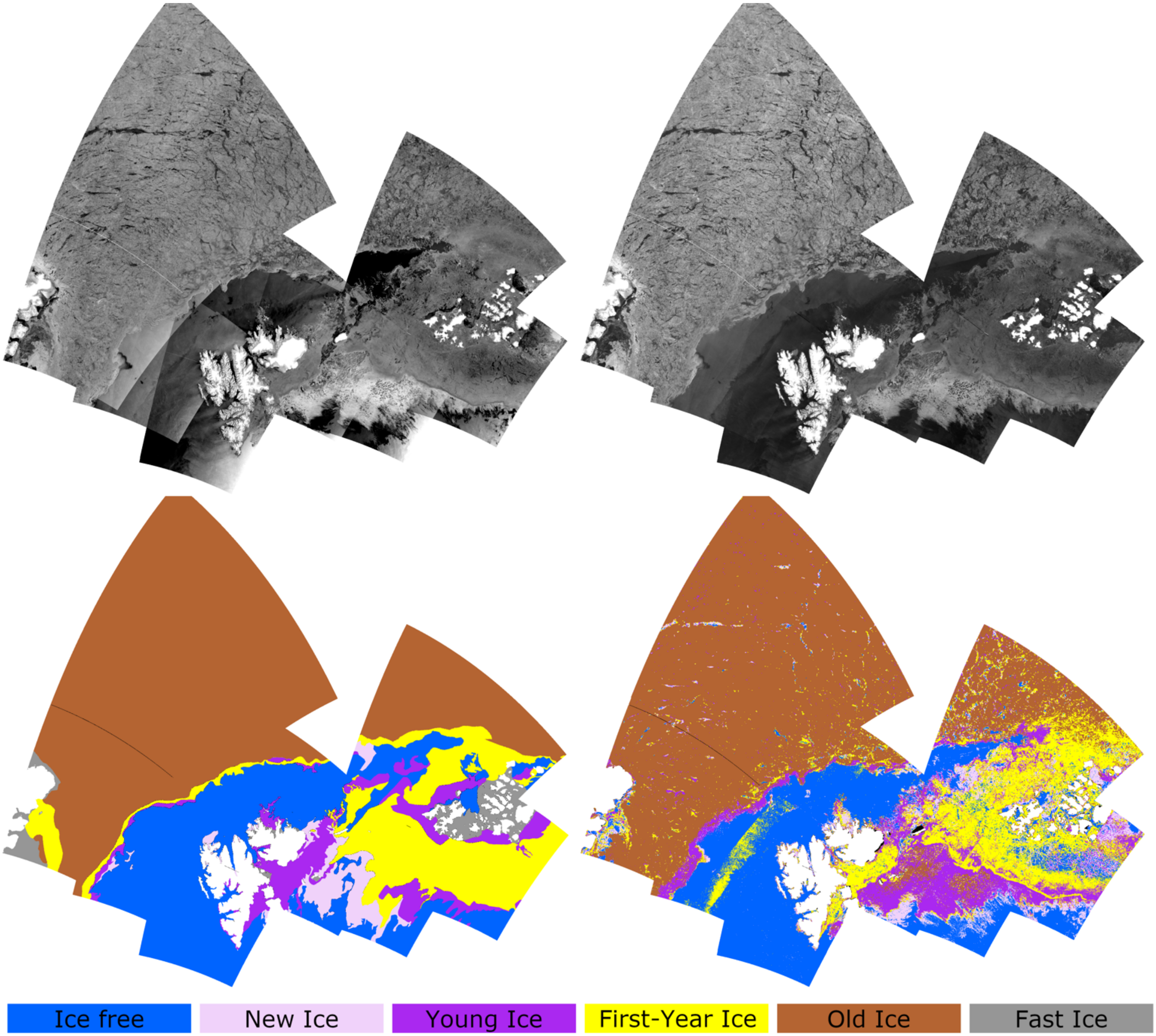
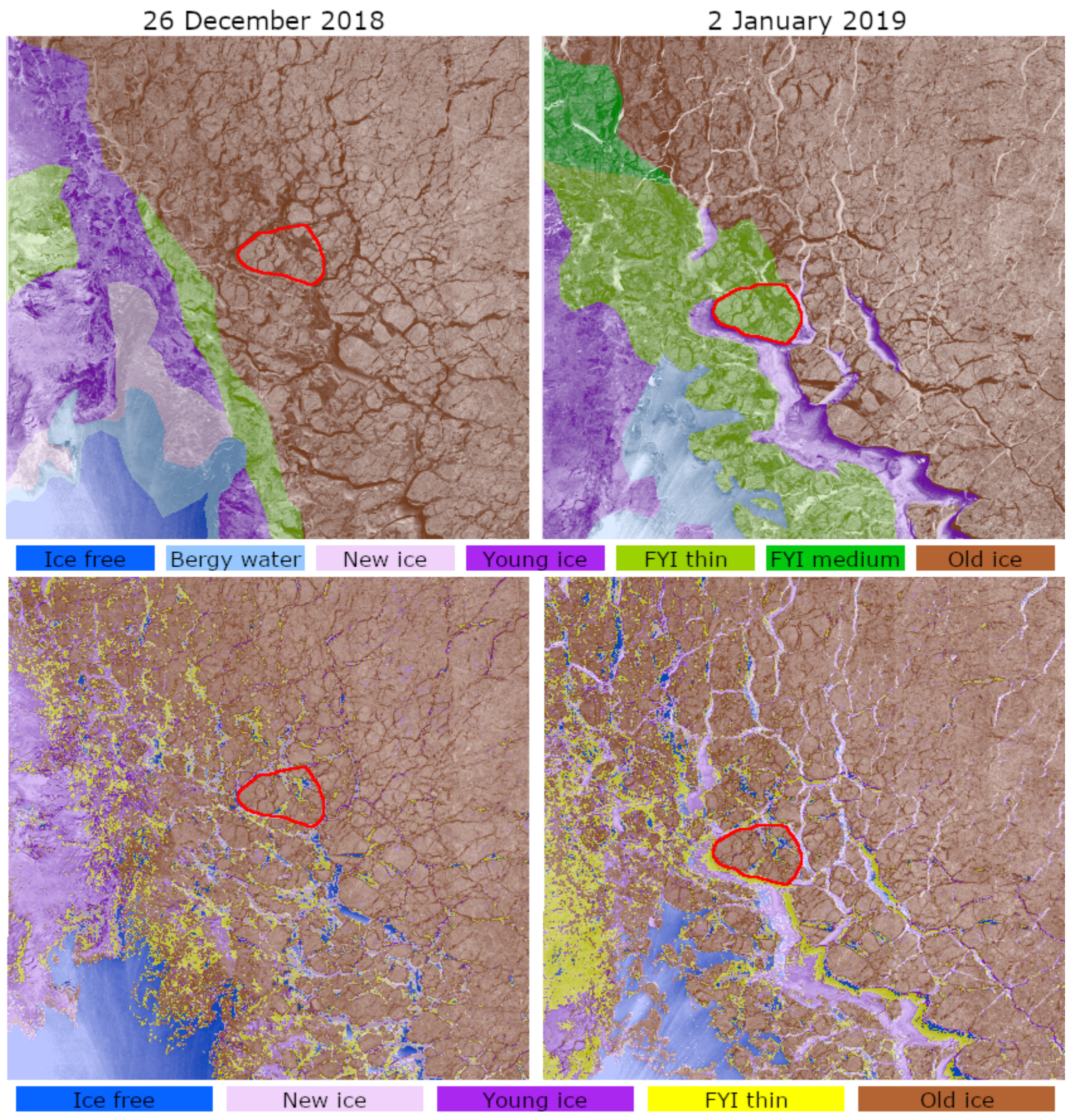


Figure 8: One~~1~~-day mosaics of Sentinel-1A/1B images (Top left: HH, Top right: HV) and the ice classification result (Bottom right) for 8 February 2019. The date of publication of the reference weekly ice chart (Bottom left) is 8 February 2019.


26 December 2018
2 January 2019
Ice free
Bergy water
New ice
Young ice
FYI thin
FYI medium
Old ice
Ice free
New ice
Young ice
FYI thin
Old ice

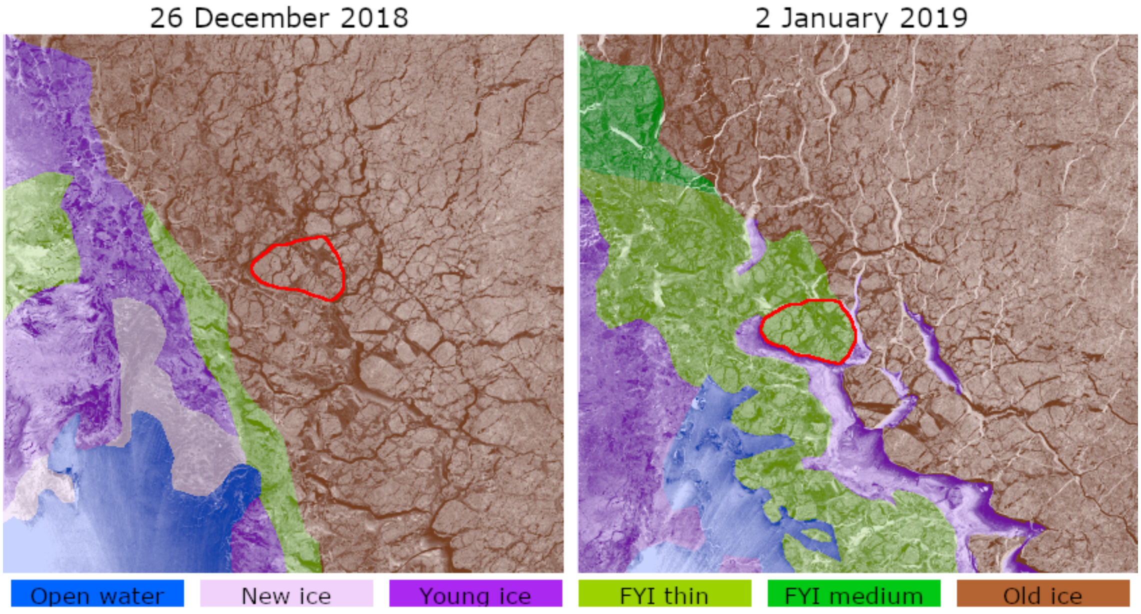


Figure 9: An example of inconsistency in ice types in the ice charts. The SoDs from the NIC ice charts are superimposed on the Sentinel-1 backscattering images. The type of the same ice floe (red outline) is annotated differently in the two ice charts (old ice on the left panel and first-year ice on the right panel) while it looks almost the same in the SAR backscattering images.

Tables

Table 1: Values of hyper parameters 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	

Table 2: Confusion matrix for the five~~5~~-class RF classifier trained for 2018 data and applied to 2018 data

		~~Predicted~~														
		~~IF (ice free)~~			~~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~~	~~IF~~	**~~93.7~~**	**~~94.6~~**	**~~95.6~~**	~~1.8~~	~~1.4~~	~~0.9~~	~~0.4~~	~~0.4~~	~~0.4~~	~~4.1~~	~~3.7~~	~~3.1~~	~~0.0~~	~~0.0~~	~~0.0~~
	~~NI~~	~~20.4~~	~~19.1~~	~~18.7~~	**~~32.5~~**	**~~33.8~~**	**~~58.3~~**	~~31.4~~	~~31.4~~	~~14.6~~	~~13.3~~	~~12.8~~	~~5.9~~	~~2.5~~	~~2.9~~	~~2.6~~
	~~YI~~	~~2.0~~	~~2.0~~	~~1.9~~	~~4.5~~	~~3.9~~	~~6.9~~	**~~60.5~~**	**~~59.1~~**	**~~61.3~~**	~~26.5~~	~~29.4~~	~~25.2~~	~~6.5~~	~~5.7~~	~~4.6~~
	~~FYI~~	~~4.4~~	~~4.2~~	~~3.4~~	~~3.1~~	~~2.8~~	~~2.9~~	~~22.3~~	~~19.8~~	~~17.8~~	**~~56.8~~**	**~~60.7~~**	**~~64.5~~**	~~13.3~~	~~12.5~~	~~11.5~~
	~~OI~~	~~0.3~~	~~0.3~~	~~0.4~~	~~0.9~~	~~0.9~~	~~1.7~~	~~5.8~~	~~5.3~~	~~3.6~~	~~7.9~~	~~7.6~~	~~6.2~~	**~~85.1~~**	**~~85.9~~**	**~~88.1~~**
		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	**94.5**	**95.2**	**96.7**	1.4	1.1	0.6	0.3	0.3	0.3	3.7	3.4	2.4	0.0	0.0	0.0
	NI	19.3	17.3	14.8	**33.1**	**38.9**	**68.7**	33.3	31.1	9.3	12.1	10.6	5.5	2.2	2.1	1.7
	YI	1.9	1.8	1.6	3.8	3.6	6.7	**62.3**	**62.8**	**64.5**	26.1	27.5	23.8	5.9	4.3	3.4
	FYI	4.2	3.6	2.4	2.6	2.6	2.1	21.7	20.6	15.6	**58.1**	**61.1**	**69.8**	13.4	12.1	10.1
	OI	0.3	0.3	0.4	0.6	0.8	1.4	5.8	5.0	3.0	7.3	7.3	4.1	**86.0**	**86.7**	**91.2**

Table 3: Confusion matrix for the five~~5~~-class RF classifier trained for 2018 data and applied to 2019 data

		~~Predicted~~

		~~IF (ice free)~~			~~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~~	~~IF~~	~~**89.6**~~	~~**90.0**~~	~~**87.3**~~	~~3.5~~	~~3.5~~	~~5.1~~	~~1.1~~	~~1.1~~	~~1.7~~	~~5.7~~	~~5.4~~	~~5.9~~	~~0.0~~	~~0.0~~	~~0.0~~
	~~NI~~	~~19.9~~	~~22.9~~	~~31.2~~	~~**29.1**~~	~~**25.3**~~	~~**20.2**~~	~~40.3~~	~~40.7~~	~~41.1~~	~~8.0~~	~~8.5~~	~~5.7~~	~~2.8~~	~~2.7~~	~~1.8~~
	~~YI~~	~~7.3~~	~~7.2~~	~~6.9~~	~~4.1~~	~~3.6~~	~~2.3~~	~~**42.7**~~	~~**41.4**~~	~~**50.4**~~	~~36.2~~	~~38.9~~	~~33.6~~	~~9.7~~	~~8.8~~	~~6.8~~
	~~FYI~~	~~6.2~~	~~5.8~~	~~5.9~~	~~4.3~~	~~3.9~~	~~2.0~~	~~24.4~~	~~23.3~~	~~25.1~~	~~**39.2**~~	~~**41.9**~~	~~**47.6**~~	~~25.9~~	~~25.0~~	~~19.4~~
	~~OI~~	~~0.6~~	~~0.6~~	~~0.6~~	~~1.4~~	~~1.3~~	~~0.5~~	~~2.4~~	~~2.7~~	~~7.1~~	~~3.1~~	~~3.5~~	~~21.1~~	~~**92.5**~~	~~**92.0**~~	~~**70.8**~~
		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**	**90.6**	**85.4**	3.1	2.7	5.7	1.0	1.1	2.0	5.8	5.7	6.9	0.0	0.0	0.0
	NI	20.1	24.5	28.3	**28.0**	**23.0**	**23.9**	42.0	42.4	40.9	7.6	7.9	5.5	2.4	2.1	1.4
	YI	6.7	6.1	6.3	3.3	3.4	3.1	**44.7**	**44.6**	**51.5**	36.0	38.2	33.5	9.3	7.7	5.7
	FYI	5.4	4.4	4.9	3.6	3.8	2.7	25.8	25.3	27.5	**38.9**	**42.0**	**46.0**	26.3	24.5	18.9
	OI	0.5	0.5	0.5	1.3	1.2	0.7	2.7	3.0	7.7	2.8	3.6	24.9	**92.7**	**91.7**	**66.3**

Table 4: Changes in classification accuracies before and after applying textural denoising

class	case								
	FC1			FC2			FC3		
	Thermal denoising only	Textural denoising applied	difference	Thermal denoising only	Textural denoising applied	difference	Thermal denoising only	Textural denoising applied	difference
OW	88.4	90.1	+1.7	88.9	90.6	+1.7	88.0	85.4	-2.6
NI	30.2	28.0	-2.8	27.7	23.0	-4.7	31.8	23.9	-7.9
YI	34.9	44.7	+9.8	36.2	44.6	+8.2	43.4	51.5	+8.1
FYI	29.3	38.9	+9.6	30.4	42.0	+11.6	38.0	47.0	+9.0
OI	91.5	92.7	+1.2	90.3	91.7	+1.4	75.2	66.3	-8.9
kappa	0.62	0.67	+0.05	0.62	0.67	+0.05	0.54	0.49	-0.05

Table 5~~4~~: Confusion matrix for the three~~3~~-class RF classifier trained for 2018 data and applied to 2018 data

		~~Predicted~~								
		~~IF (ice free)~~			~~FYI (first-year ice)~~			~~OI (old ice)~~		
	~~Case~~	~~FC1~~	~~FC2~~	~~FC3~~	~~FC1~~	~~FC2~~	~~FC3~~	~~FC1~~	~~FC2~~	~~FC3~~
~~Actual~~	~~IF~~	~~**96.5**~~	~~**96.7**~~	~~**96.9**~~	~~3.5~~	~~3.3~~	~~3.1~~	~~0.0~~	~~0.0~~	~~0.0~~
	~~MFYI~~	~~5.8~~	~~5.7~~	~~4.8~~	~~**84.5**~~	~~**85.8**~~	~~**87.2**~~	~~9.8~~	~~8.6~~	~~7.9~~
	~~OI~~	~~0.5~~	~~0.5~~	~~0.5~~	~~14.4~~	~~13.9~~	~~12.4~~	~~**85.1**~~	~~**85.6**~~	~~**87.1**~~
		Predicted								
		OW (open water)			mFYI (mixed FYI)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**96.7**	**97.3**	**99.1**	3.3	2.6	0.9	0.0	0.0	0.0
	mFYI	5.2	4.7	2.5	**85.8**	**87.6**	**92.3**	9.0	7.7	5.2
	OI	0.4	0.4	0.2	13.2	12.4	6.0	**86.4**	**87.2**	**93.8**

Table ~~5~~6: Confusion matrix for the three~~3~~-class RF classifier trained for 2018 data and applied to 2019 data

		~~Predicted~~								
		~~IF (ice free)~~			~~FYI (first-year ice)~~			~~OI (old ice)~~		
	~~Case~~	~~FC1~~	~~FC2~~	~~FC3~~	~~FC1~~	~~FC2~~	~~FC3~~	~~FC1~~	~~FC2~~	~~FC3~~
~~Actual~~	~~IF~~	~~**93.4**~~	~~**93.4**~~	~~**91.9**~~	~~6.5~~	~~6.6~~	~~8.1~~	~~0.0~~	~~0.0~~	~~0.0~~
	~~MFYI~~	~~9.8~~	~~9.2~~	~~8.9~~	~~**71.0**~~	~~**72.9**~~	~~**75.3**~~	~~19.2~~	~~17.9~~	~~15.8~~
	~~OI~~	~~0.7~~	~~0.7~~	~~0.6~~	~~6.8~~	~~7.7~~	~~18.5~~	~~**92.5**~~	~~**91.6**~~	~~**81.0**~~
		Predicted								
		OW (open water)			mFYI (mixed FYI)			OI (old ice)		
	Case	FC1	FC2	FC3	FC1	FC2	FC3	FC1	FC2	FC3
Actual	OW	**93.6**	**93.6**	**86.3**	6.4	6.4	13.6	0.0	0.0	0.0
	mFYI	8.8	7.5	7.1	**72.4**	**75.3**	**81.6**	18.8	17.2	11.3
	OI	0.6	0.6	0.4	6.7	8.1	39.8	**92.7**	**91.4**	**59.7**