



Classification of Sea Ice Types in Sentinel-1 SAR images

Jeong-Won Park^{1,3}, Anton A. Korosov¹, Mohamed Babiker¹, Joong-Sun Won², Morten W. Hansen¹, Hyun-Cheol Kim³

¹Ocean and Sea Ice Remote Sensing Group, Nansen Environmental and Remote Sensing Center, Bergen, 5006, Norway

5 ²Department of Earth System Sciences, Yonsei University, Seoul, 03722, South Korea

³Unit 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 ice charting. Previous studies mostly rely on manual work in selecting training and validation data. We show that the use of readily
10 available ice charts from an operational ice services allow to automate selection of large amount of training/testing data void of biased, subjective decisions. 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
15 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 3 generalized cover types (ice free, integrated first-year ice, old ice) with overall accuracies of 85% and 5 cover types (ice free, new ice, young ice, first-year ice, old ice) with accuracy of 58%. The errors are attributed
20 both to incorrect manual classification on the ice charts and to the 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
25 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, 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 Sentinel-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
5 TOPSAR image intensity is disturbed by the thermal noise particularly in cross-polarization channel. Although the European Space Agency (ESA) provides calibrated noise vectors for noise power subtraction, residual noise contribution is significant considering 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
10 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 of the previous works on sea ice classification, 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, and the number of samples were not enough to generalize the results because of the
15 laborious manual work. Therefore, increasing objectivity is crucial, and automating the classification process is encouraged. The use of public ice chart as validation reference data may help in solving the validation problem and enabling automation. 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
20 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 interest for developing and testing the proposed algorithm is the Fram Strait and Barents Sea including a part
25 of the Arctic Ocean (10°W-70°E, 75°N-85°N) as shown in Figure 1. Various sea ice types coexist 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 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
30 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 classifier and those from the second year (hereafter 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: 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
5 (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
10 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

15 To take the advantage of objective identification of the ice type and to develop an 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
20 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
25 image. Heinrichs et al. (2006) reported that the ice edge determined from the AMSR-E, which is a passive microwave radiometer, using the isoline of 15% concentration matches best the ice edge determined from RADARSAT-1, 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 in the SIGRID-3 format has precision of decimals. CP is also important to find the dominant ice type in the given polygons. SoD is so-called ice type.
30 It is challenging to differentiate ice types using SAR data only, thus we simplified the SoDs by merging into five classes: ice free, 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). In 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

5 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 and level sea ice, 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
10 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 Sentinel-1 images was denoised before further processes are applied.

2.2.3 Incidence angle correction

15 It is well known that there is a strong incidence angle dependency in the SAR backscattering intensity for ocean and sea ice surface. 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 several
20 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).

25 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. For HH polarization, the estimated slope was -0.200 dB/degree, which is slightly different from the estimation in Mäkynen and Karvonen (2017). 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.,
30 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.



2.2.4 Texture feature computation

Like many of the previously developed methods (Soh and Tsatsoulis, 1999; Deng and Clausi, 2005; 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 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, 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 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 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.

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

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

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 26 Haralick features, the coefficient of variation (CV) which is reported as useful feature for ice-water discrimination (Keller, 2017) is included. The CV is defined as follows:

$$CV = \sigma / \mu \quad (3)$$

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.

5 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 well for sparse dataset. 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. 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, 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} 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 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).

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 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 fitting the Richard's Curve (Richard, 1959). 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 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 3 days from the date of publication, the ice edges in some images match well with those in the ice chart. To automate image selection for training, a good ice/water classifier for SAR image is needed. Since even such a simple binary 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-”
10 automated for now.

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
15 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. We use confusion matrix for performance evaluation. The validation is done in the same way but using a completely independent 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: texture features and incidence angle, iii) FC3: texture features, incidence angle, and day of year.
20

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
30



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 5-class classifier consists of 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 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 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 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 in Table 2. The three cases with different feature configurations (FC1-FC3) were tested. The accuracies for ice free 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%). 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 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.

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 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.

5 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” first-year ice, and then training new classifiers. Table 4 and Table 5 show the confusion matrices for the 3-class classifiers. The dramatic increase in the accuracy for the integrated 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
10 similar level to the case of 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
15 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 the same time, it should be noted that training with ice chart might have included wrong 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, which used
20 manually classified ice maps as training and validation reference, are expected.

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
25 SAR-based sea ice type classifier. A conventional approach for selecting training/testing data by 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% for 3-class and 5-class ice type classifiers, respectively. These are slightly lower than the numbers in the
30 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. The proposed approach importantly showed that a daily ice type mapping from the Sentinel-1 data is feasible, and it can help capturing 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

- 5 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

- 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.
- 5
- 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.
- Cortes, C. and Vapnik, V.: Support-vector networks, *Mach. Learn.*, 20(3), 273-297, doi: 10.1007/BF00994018, 1995.
- 10
- 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.
- ESRI (Environmental Systems Research Institute, Inc.): ESRI Shapefile Technical Description, An ESRI White Paper, 1998.
- 15 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.
- 20 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.
- 25
- 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.
- 30 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.



- 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
5 format for sea ice georeferenced information and data, JCOMM Technical Report No. 23, Tech. rep., World Meteorological Organization, Geneva, Switzerland, 2014b.
- 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.
- 10
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.: Active/passive dual polarization sea ice detection, NISAR Applications Workshop: Sea-Ice, Maryland, USA, 23
15 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
20 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.
- 25 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.
- Mäkynen, M. and Karvonen, J.: Incidence angle dependence of first-year sea ice backscattering coefficient in Sentinel-1
30 SAR imagery over the Kara Sea, *IEEE T. Geosci. Remote*, 55(11), 6170-6181, doi:10.1109/TGRS.2017.2721981, 2017.
- 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.
- 5 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.
- 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.
- 10 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.
- 15 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.
- 20 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.
- 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.
- 25 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.
- 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, 30 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

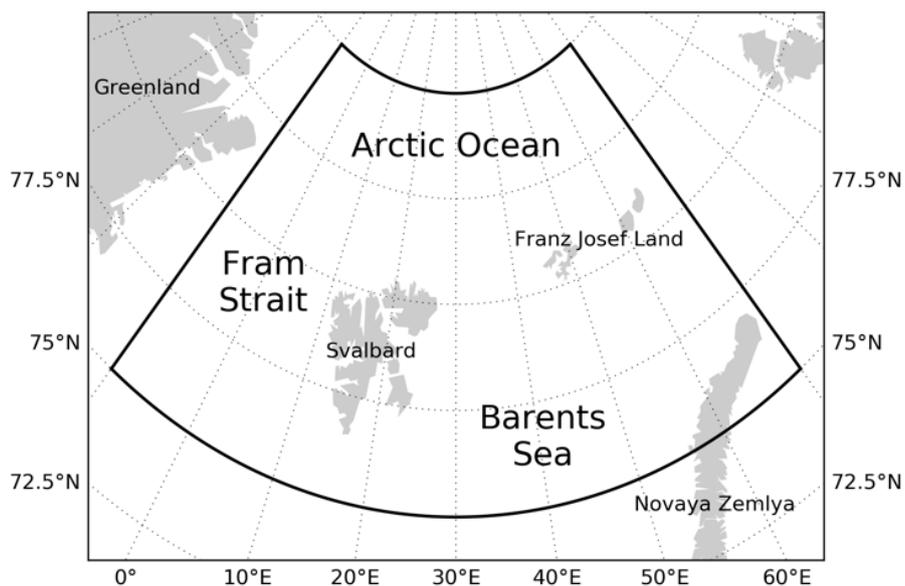


Figure 1: Study area.

5

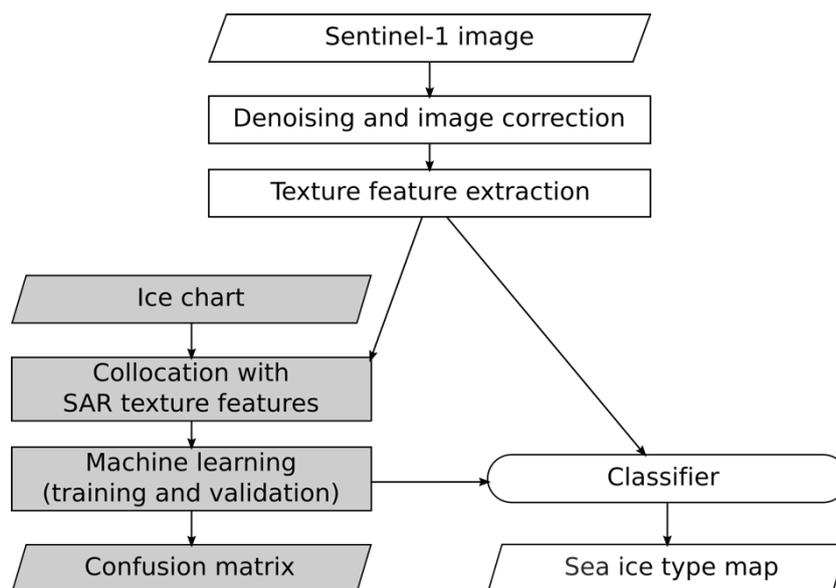
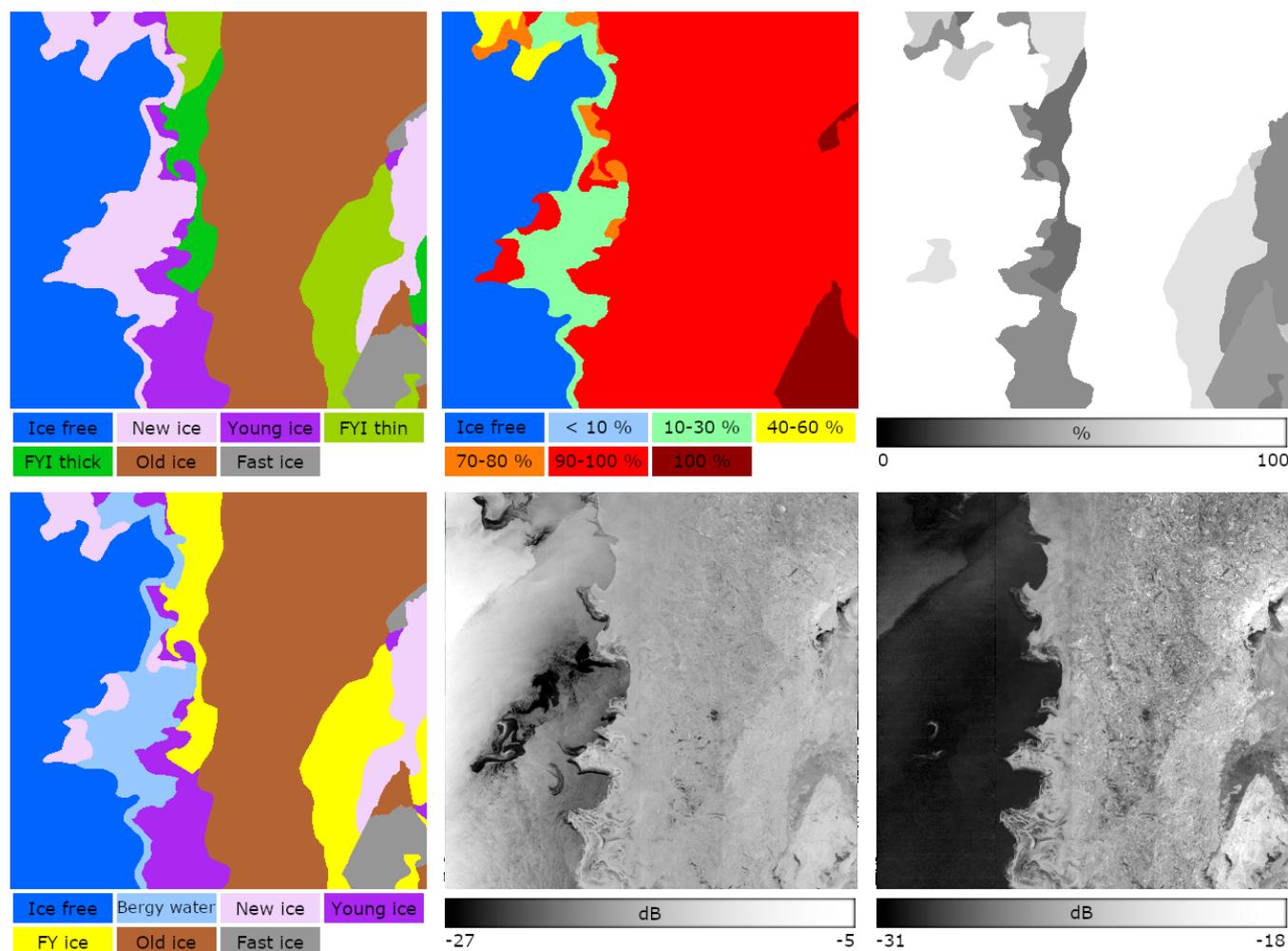


Figure 2: Processing flowchart of the proposed algorithm. Gray color shows training phase.



5 **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 (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. 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 were used in this example.**

10

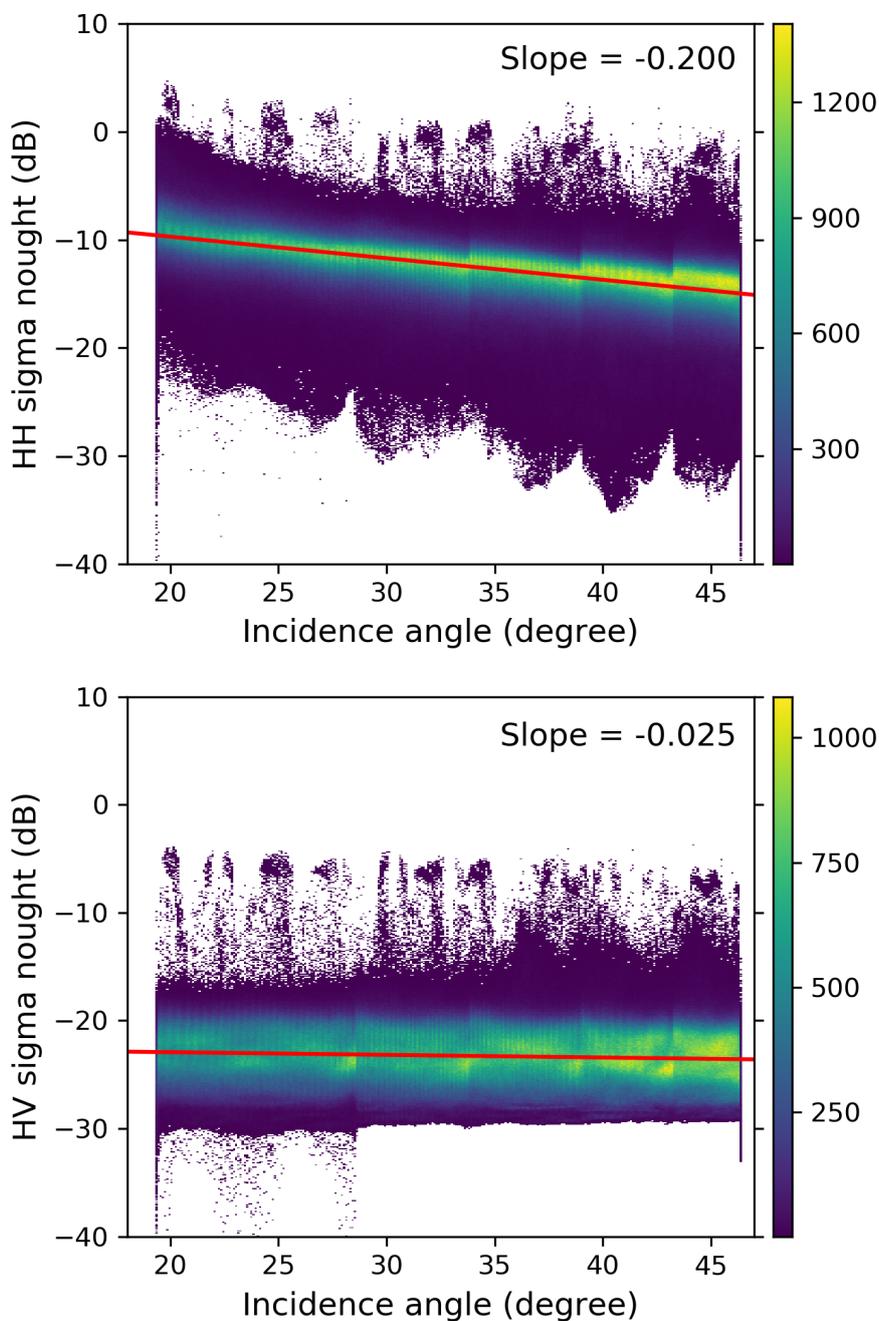


Figure 4: Two-dimensional histograms of incidence angle versus sigma nought for HH (top) and HV (bottom) polarization channels. Pixels covering various types of sea ice were merged so that the averaged property can be estimated. The best fit linear trends are shown with red lines.

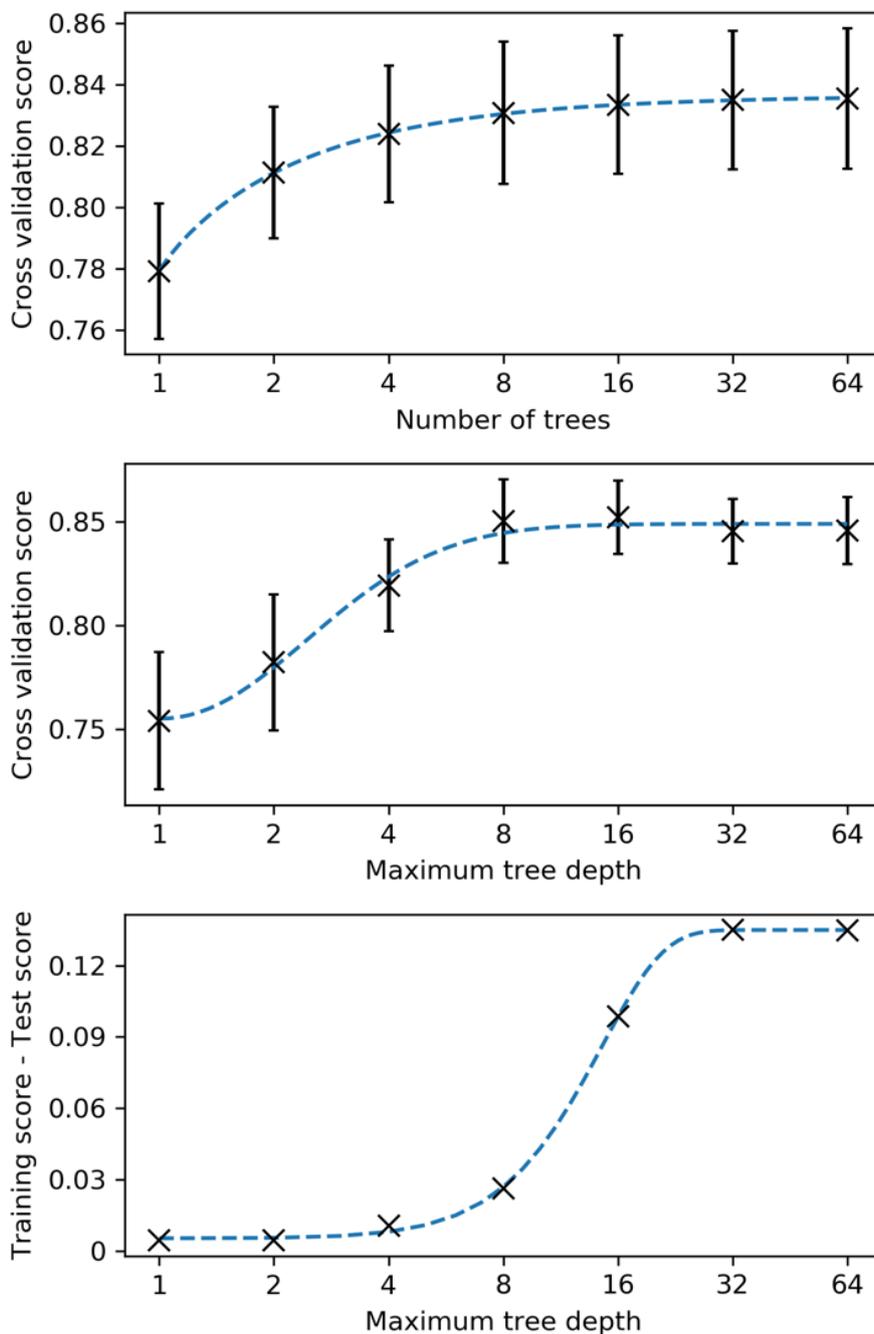


Figure 5: Hyperparameter optimization using grid search results (cross). Dashed lines represent the best-fit Richard's Curve. (Top panel) The optimal values are extracted from the locations where the score increments per unit of each hyperparameter become lower than a threshold (e.g., 0.001). (Center 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.

5



5

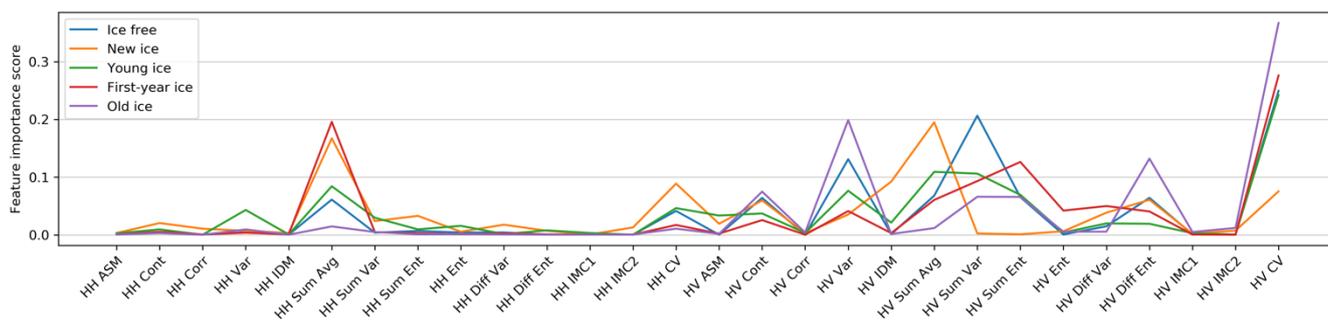


Figure 6: Feature importances of the binary sub-classifiers. ASM: angular second moment; Cont: contrast; Corr: correlation; Var: variance; IDM: inverse difference moment; Sum Avg: sum average; Sum Var: sum variance; Sum Ent: sum entropy; Ent: entropy; Diff Var: difference variance; Diff Ent: difference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

10

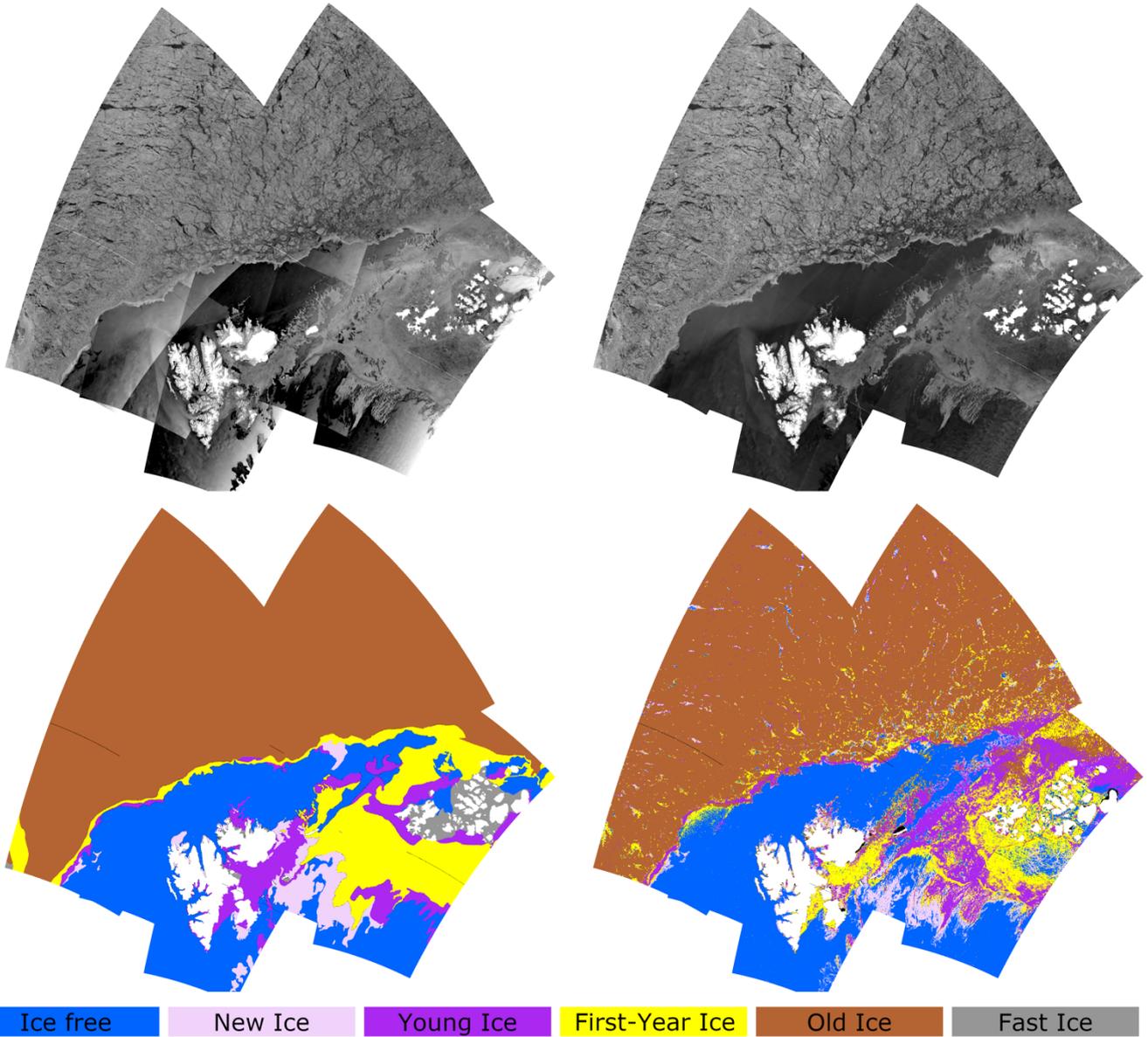


Figure 7: 1-day mosaics of Sentinel-1A/1B images (Top left: HH, Top right: HV) and the ice classification result (Bottom right) for 5 February 2019. The date of publication of the reference weekly ice chart (Bottom left) is 8 February 2019.

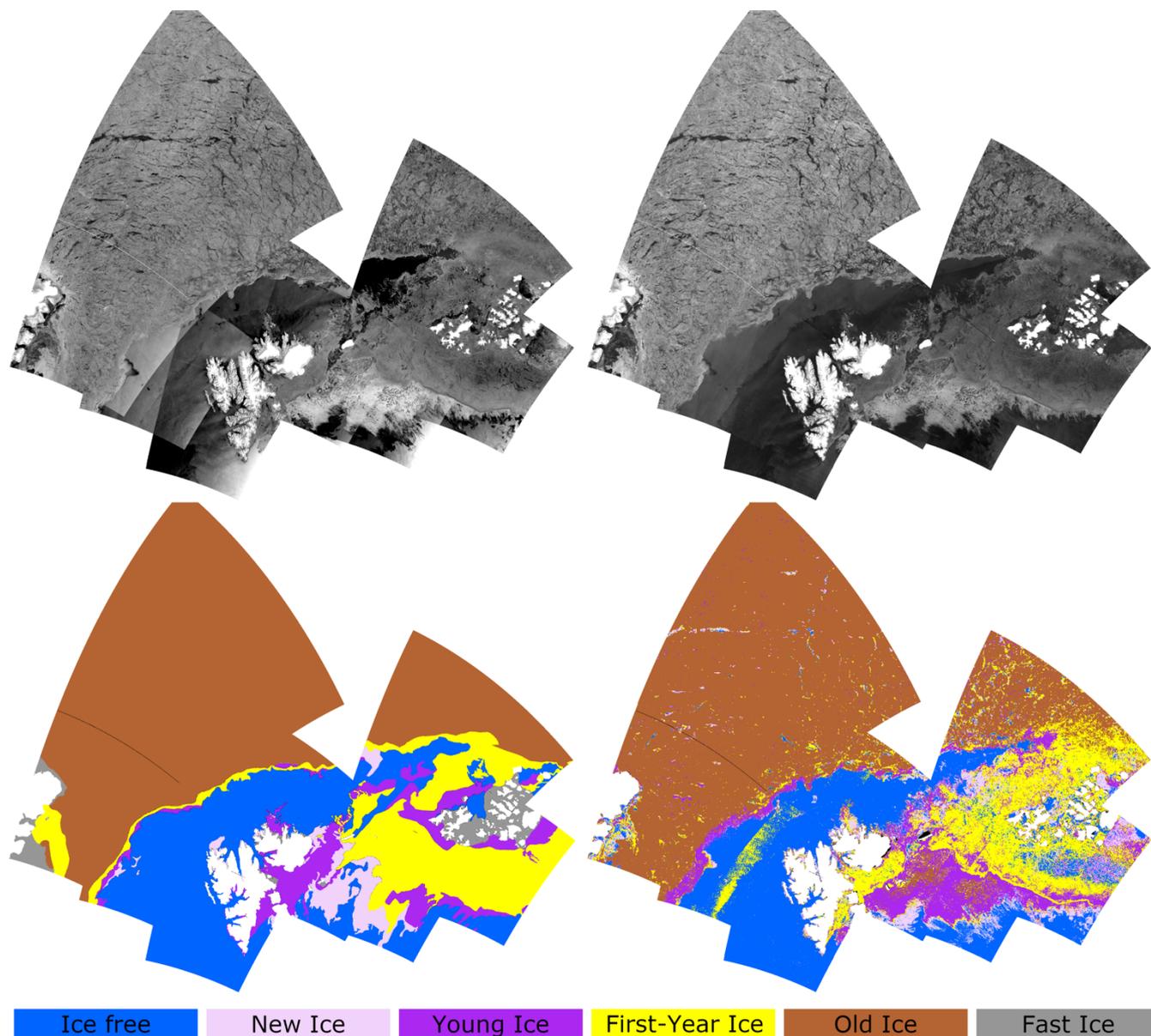


Figure 8: 1-day mosaics of Sentinel-1A/1B images (Top left: HH, Top right: HV) and the ice classification result (Bottom right) for 8 February 2019. The date of publication of the reference weekly ice chart (Bottom left) is 8 February 2019.

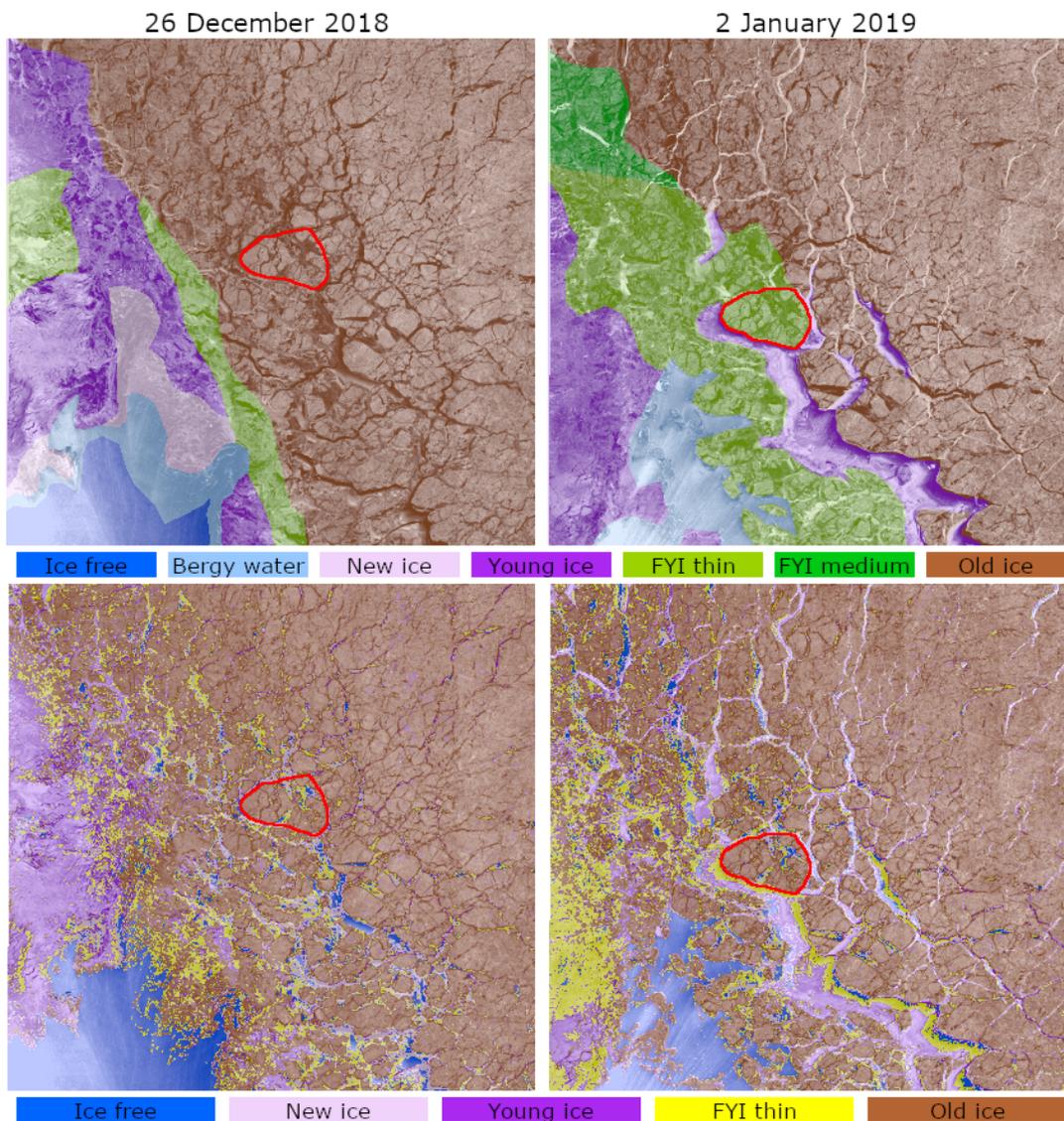


Figure 9: An example of inconsistency in ice types in the ice charts. The SoDs from the NIC ice charts are superimposed on the Sentinel-1 backscattering images. The type of the same ice floe (red outline) is annotated differently in the two ice charts (old ice on the top left panel and first-year ice on the top right panel) while it looks almost the same in the SAR backscattering images. Such inconsistency was not observed in our 5-class classification results (bottom panels).

5



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	

5

Table 2: Confusion matrix for the 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

10 **Table 3: Confusion matrix for the 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



Table 4: Confusion matrix for the 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

5

Table 5: Confusion matrix for the 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