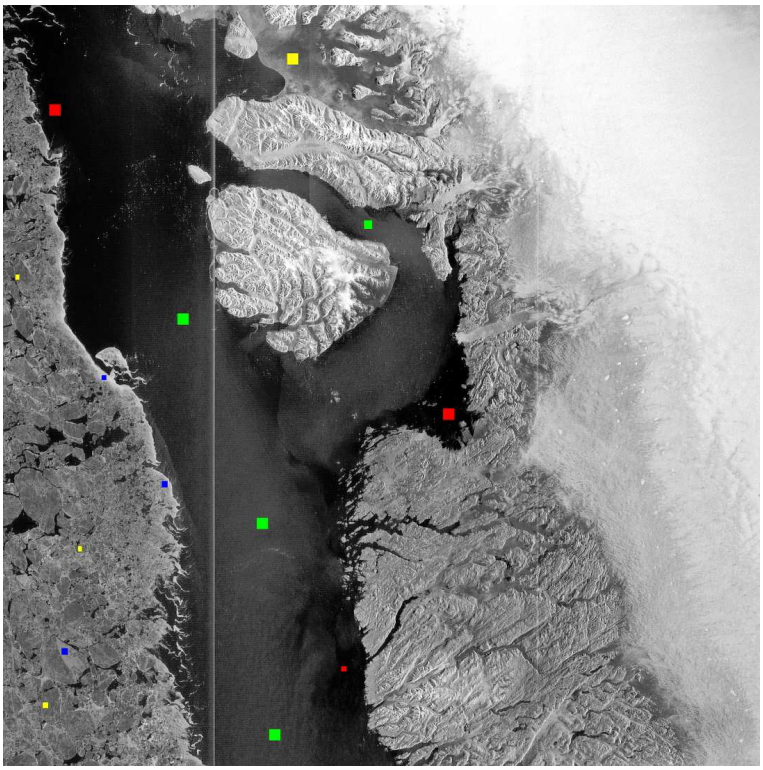
Danish Meteorological Institute
Ministry of Transport and Energy

Scientific Report 05-04

Sea Ice Concentration from Single-Polarized SAR data using Second-Order Grey Level Statistics and Learning Vector Quantization

Thomas Bøvith and Søren Andersen



Colophone

Serial title:
Scientific Report 05-04

Title:
Sea Ice Concentration from Single-Polarized SAR data using Second-Order Grey Level Statistics and Learning Vector Quantization

Authors:
Thomas Bøvith and Søren Andersen

Responsible Institution:
Danish Meteorological Institute

Language:
English

Keywords:
sea ice, concentration, synthetic aperture radar, classification

Url:
http://www.dmi.dk/dmi/sr05-04

ISSN:
1399-1949 (online)

ISBN:
87-7478-523-0

Website:
http://www.dmi.dk/

Abstract

A classification system for obtaining the sea ice concentration from satellite based single-polarized SAR data is presented. The method is based on a supervised neural network classification of second-order grey level statistics features. Learning Vector Quantization is used for describing the boundaries between the surface classes. Five RADARSAT ScanSAR Wide images of the sea ice off of the East and West coast of Greenland were classified with resulting classification accuracies from 80 to 98 percent. The algorithm is robust to the selection of training data which was measured by classifying the same dataset twice using multiple experts for selecting training data. Thus, the precision by which sea ice and open water was assessed to be approximately two percent. The classification accuracy is dependent on the sea state, especially the degree to which the water is roughened by wind. The current classification system is not able to distinguish between high backscatter sea ice and high backscatter ocean areas and therefore areas of wind roughened sea water are masked out after classification.

Contents

1. Introduction

The presented work is part of the IOMASA project[1](IOMASA, 2002a) in which a main goal is to achieve improved remote sensing of sea ice with more accurate determination of sea ice concentration of the Arctic Seas. In the IOMASA project, passive microwave ice concentration algorithms are employed, and for the validation of these algorithms, high resolution ice concentration measures are necessary (IOMASA, 2002b). This report describes a new method based on texture measures and neural networks for obtaining high resolution sea ice concentration from single-polarized wide swath SAR data from the RADARSAT and ENVISAT satellites.

Currently, ice concentration is available for selected areas of the Arctic Sea on an operational basis from navigational ice charts. However, the ice concentration from ice charts are only accurate within an interval of 10-20% and are based on subjective judgements and generalizing of the ice concentration. Optical sensors in the visible or infrared range of the electromagnetic spectrum are of little use in the Arctic due to poor or little sunlight in the winter months and due to the extensive cloud cover of the region. To validate the global coverage of today's passive microwave radiometers and scatterometers, the only feasible source of data is swath SAR data (typical swath widths of 500 km).

In this report an accurate and objective method for obtaining a high resolution and independent ice concentration measure is presented. The method is semi-automatic and requires user interaction in the selection of training areas and fine-tuning of the classification result.

Firstly, the SAR data is pre-processed and manual selection of training data is carried out by ice charting experts. Then the SAR image is classified using the *Learning Vector Quantization* (LVQ) algorithm as described by Kohonen (2001) and Hastie et al. (2001). The algorithm is a fast and flexible neural network-like algorithm. Second-order grey level statistics (calculated using Grey-Level Cooccurence Matrices) are used as features and the LVQ algorithm uses the training data to classify the SAR image into a number of relevant classes (e.g. calm water, rough water, high and low backscatter sea ice). The classification framework is based on extensions to software by Lars Kaleschke of the University of Bremen, who used LVQ for sea ice classification using ERS-2 SAR images (Kern et al., 2003). Finally, post-processing is carried out to produce a classification mask to indicate areas of successful classification.

2. Data

For the development and testing of the classification system five representative RADARSAT ScanSAR Wide images are classified. The images display varying conditions of sea ice and meteorological conditions; from very simple to more difficult to interpret. The ScanSAR Wide images (listed in table 3.1) have an original pixel size of 50x50 m and a swath width of 500 km.

The following section will give a short description of each data set. The figures showing the images are shown in Appendix A.

Scene 1 (Figure A.1): RADARSAT image from West Coast of Greenland. The image is easy to interpret due to the good weather conditions (little wind). There are few different surface types: First year ice, fast ice and open water (calm and turbulent). From visual inspection the surface types

[1]IOMASA (Integrated Observation and Modeling of Arctic Sea ice and Atmosphere), a 'Research and Technological Development'-project of the European Commission under contract EVK3-CT-2002-00067

RADARSAT ScanSAR Wide C_{HH} data			
no	date	orbit	area
1	2000-05-24	23765	Disko Bay, West Coast of Greenland
2	2000-07-22	24607	Scoresby Sund, East Coast of Greenland
3	2001-12-16	31932	Disko Bay, West Coast of Greenland
4	2001-12-22	32018	Disko Bay, West Coast of Greenland
5	2001-12-26	32075	Disko Bay, West Coast of Greenland

Table 3.1: Image data

appears to be well separated.

Scene 2 (Figures A.2 and A.3): RADARSAT image from East Coast of Greenland. The image is more complex vdue to many different surface classes: First and multi-year ice floes of varying concentrations as well as regions of calm and turbulent water. The surface types are mixed.

Scene 3 (Figure A.4): RADARSAT image from West Coast of Greenland. Moderately complex image due to turbulence in the water regions and over areas of the first year ice floes. The image show poor contrast of the sea ice areas in the far range of the image.

Scene 4 (Figure A.5): RADARSAT image from West Coast of Greenland. Less turbulence over the ice covered areas makes this image fairly easy to interpret. Large tonal variance from the near to the far range of the image.

Scene 5 (Figure A.6): RADARSAT image from West Coast of Greenland. Moderate amount of turbulence, similar to scene 4. Some areas of thin ice (dark).

No ground truth data was available to support the interpretation of the images, however, trained ice experts from the Danish Meteorological Institute (DMI) assisted in the interpretation of the SAR images and in the selection of training data.

3. Method

Classifying sea ice from SAR images is not a trivial task. Selection and extraction of features, selection of training data, and the choice of classification method, all influence the outcome and accuracy of the classification. Furthermore, the interplay between the radar system parameters (wavelength, polarization, incidence angle, number of looks, etc.) and the surface properties, i.e. how the ice and water surfaces reflect the signal, is an important consideration in designing a robust and successful classification scheme.

It is difficult to dessign a universally valid sea ice concentration algorithm operating on single-polarized SAR data. This is due to the large variabilities in the environment in which sea ice occur and the mode of operation of the SAR sensor. Often classification schemes are limited to operate on images from a particular sensor, in a particular season of the year, over a particular geographical area. At worst, the classification method will be limited to providing useful results on a single image only.

In this work the goal was to design a method for obtaining accurate ice concentration from a range of different images and areas. Therefore the selection of features and training data as well as the parameters of the neural network were not based on a single reference image only.

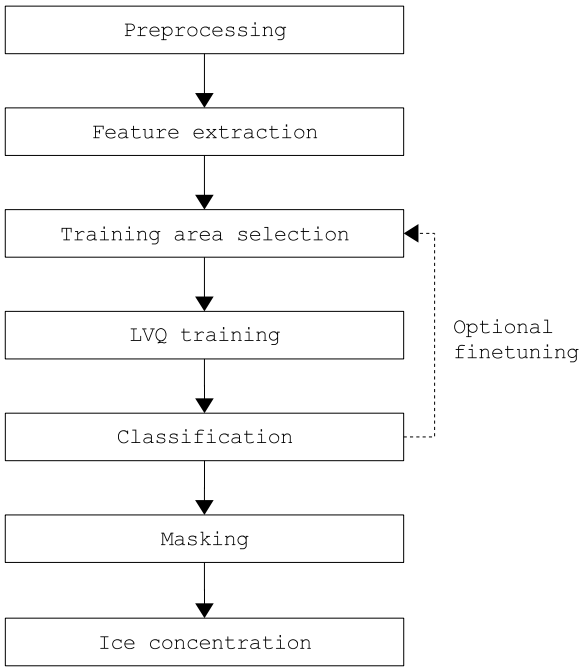


Figure 4.1: The neural network classification scheme.

The classification scheme was designed to allow for fine-tuning to accommodate to the individual challenges of each image. The manual fine-tuning is carried out by an ice expert, and allows for changing the number of surface classes as well as selecting different training areas. A flow chart of the classification scheme is shown in Figure 4.1.

Preprocessing

RADARSAT data, having a spatial resolution of 100 m, but a pixel size of 50 m, was pre-processed by a simple downsampling to a pixel size of 100 m. This decreases the computation time as well as suppresses some of the speckle noise inherent to SAR imagery. ENVISAT ASAR data was not downsampled prior to feature extraction due to a lower original pixelsize of 75 m which is comparable to the downsampled 100 m RADARSAT resolution.

SAR imagery, and especially wide swath SAR imagery, often display a tonal variation in the range direction of the image (see Figure 4.2, left). This is due to the varying incidence angle between radar rays and the Earth's surface as well as the changing distance from the radar to the target. These effects can partially be corrected for by absolute calibration of the SAR image. However, for this work a pragmatic solution to range correction was applied: The average variation of backscatter across range for sea ice was modelled by an exponential curve and subsequently applied to the SAR images to correct for this effect. This correction is not used for RADARSAT images but only ENVISAT images, due to the fact that ENVISAT displays a more pronounced range-brightness variation due to it's wider range of incidence angle. In Figure 4.2 the effect of the range correction can be seen.

Feature selection and extraction

Due to speckle inherent in SAR data it is not possible to accurately classify SAR data using the amplitude or the normalized backscattering cross section only. Therefore an important issue to consider is the extraction of reasonable features from the data.

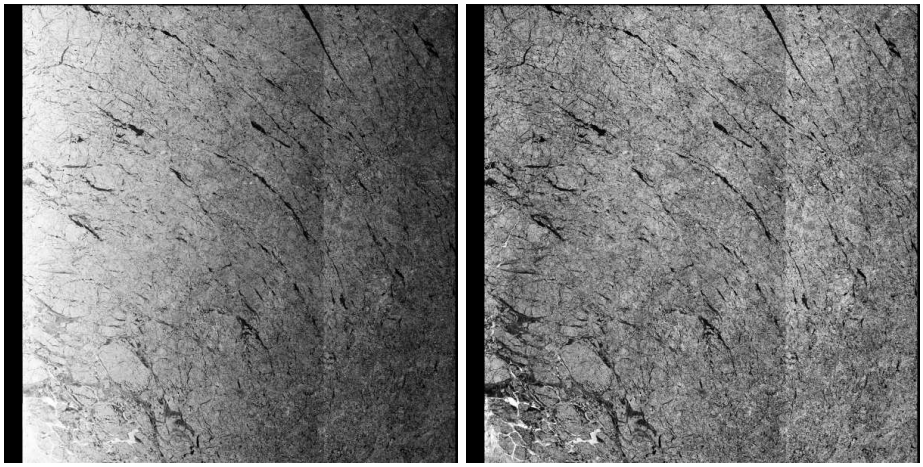


Figure 4.2: Left: Uncorrected ENVISAT WSM scene. Right: Corrected ENVISAT WSM scene.

Many different features and their combinations have been investigated in the field of sea ice classification, but no universal features, invariant to data resolution, seasonal or weather conditions, has been formulated. A common approach is to use region based features, i.e. to calculate a measure of a local region around each pixel. Examples of this approach is the use of second-order grey level statistics, distribution matching and adaptive thresholding methods.

Texture patterns from second-order grey level statistics have been shown to provide valuable, but not always unambiguous, information about the surface classes. Several studies present successful results in discriminating sea ice in the texture feature space, e.g. Shokr (1991), Barber and LeDrew (1991) and Kaleschke and Kern (2002). It was chosen to elaborate on these methods by using second-order grey level statistics derived from the Grey-Level Co-occurrence Matrix (GLCM) as defined originally by Haralick et al. (1973).

The texture features were computed using the Grey Level Co-occurrence Linked List, which makes fast computation of second-order grey level textures possible. The algorithm was described by Clausi and Jernigan (1998) and later used for sea ice classification by Kaleschke and Kern (2002). This implementation allows for the computation of eight second-order statistics as listed in table 4.1, where C_{ij} is the probability for the cell ij in the GLCM, σ_x is the standard deviation of row i, σ_y is the standard deviation of column j, μ_x is the mean of row i, μ_y is the mean of column j, G is the number of grey levels.

The different texture measures reflect the degree of smoothness and the homogeneity of the surface and is normally grouped into three groups as indicated in table 4.1. Within the smoothness and the homogeneity groups the texture measures tend to be correlated. Therefore, a common approach in selecting texture features for classification is to choose only one or a few features from each group.

There are a number of parameters to consider before calculating the GLCM for the SAR images: quantization of the input data, size of the GLCM and the direction in which the GLCM is calculated.

For this project, after some initial tests, it was chosen to quantize the original 256 grey level (8 bit) images to 64 levels (6 bit). This vgives a considerable faster computation of the grey-level co-occurrence matrices without loss of information. Clausi (2002) reported that a too high or too low quantization resulted in decreased classification accuracies.

Homogeneity texture measures		
ASM	Angular Second Moment (or Energy)	$ASM = \sum_{i,j=0}^{G-1} C_{ij}^2$
ENT	Entropy	$ENT = -\sum_{i,j=0}^{G-1} C_{ij} log C_{ij}$
Smoothness texture measures		
CON	Contrast	$CON = \sum_{i,j=0}^{G-1} C_{ij}(i-j)^2$
DIS	Dissimilarity	$DIS = \sum_{i,j=0}^{G-1} C_{ij}\,\lvert i-j\rvert$
HOM	Homogeneity	$HOM = \sum_{i,j=0}^{G-1} \frac{C_{ij}}{1+(i-j)^2}$
INV	Inverse Difference Moment	$INV = \sum_{i,j=0}^{G-1} \frac{C_{ij}}{1+\lvert i-j\rvert}$
Other texture measures		
COR	Correlation	$COR = \sum_{i,j=0}^{G-1} \frac{(i-\mu_x)(j-\mu_u)C_{ij}}{\sigma_x \sigma_y}$
MU	Mean	$MU = \sum_{i,j=0}^{G-1} iC_{ij}$

Table 4.1: The eight Grey Level Co-occurrence Texture Measures.

The window size of the GLCM has been discussed by for example Shokr (1991) and an appropriate box size of 9x9 pixels was chosen after some trails using larger and small window sizes. The spatial scale of the textures, as well as the spatial resolution of the RADARSAT ScanSar Wide images were also taken into account. Too large a window causes blurring of the image, especially coming into effect at the boundaries between ice and water, whereas too small a window size does not allow the detection of the desired textures. A smaller window size is preferable if computation time of the GLCM is of importance.

The GLCM can be calculated for a range of different directions, allowing the detection of textures that are aligned at specific directions. Barber and LeDrew (1991) concludes that a GLCM calculated with an orientation parallel to the look direction gives maximum discrimination of sea ice, however for this work this statement has not been investigated and it was chosen to use an averaging of directions by making the GLCM rotationally invariant.

Number of features

From the body of research into classification of sea ice using second-order statistics it is, as mentioned ealier, not possible to conclude which feature measures are the best. Clausi (2002) uses second order statistics on RADARSAT ScanSAR images and concludes that if only three features are to be used, the contrast (CON), the entropy (ENT), and the correlation (COR) measures are

preferred. Kaleschke and Kern (2002) uses quite similar features on ERS-2 SAR data, recommending the use of the entropy (ENT), mean (MU), contrast (CON) as well as a lee-filtered image (speckle reducing filter).

However, the optimal choice of features might be suspected to be quite closely connected to the specific data set used. On this background it was decided, for the development of this method, to investigate the optimal combination and number of features over a range of typical images.

Using the full set of texture features in one single classification is not expected to provide good results, i.e. a high classification accuracy. This is because using all of the features will make it difficult for the training samples to model the full feature space. Furthermore, many of the GLCM texture features are correlated and will therefore not improve on the clustering of the data but only make the feature space more noisy. The investigation on the topic of optimal feature selection and its results is presented in the later chapter 5, Results and Discussion.

Examples of the texture images can be seen in Figure B.1 in Appendix B where all nine features for scene 2 is shown.

Training data

Selection of training areas for the classification has to be carried out with special care to ensure a good statistical data material for the neural network training and classification. Campbell (1996) mentions studies showing that the selection of training data is equally important to the classification accuracy as the choice of the classification algorithm and its parameters itself.

In the selection of training data it was assured that homogeneous areas, containing only one surface class were selected. Furthermore, areas were selected evenly distributed over the entire image (if possible). This ensures that all the variations within the class is represented within the training data.

In the Figures A.1 to A.6 the distribution of the training areas for each RADARSAT scene can be seen as colored rectangles.

The method is not dependent on the ability to select equal number of training pixels to be selected for each class. After the manual selection of training areas, an equal amount of training data is extracted randomly from the manual selected areas. For scene 1, 2a, 2b, 3 and 4, a total of 2000 training pixels per class were extracted, whereas for scene 5 only 1300 pixels per class were extracted.

For the validation of the classification, the training data is divided randomly into two subsets of data. One for *training* the neural network and one for *testing* the accuracy of the classification by calculating how well the neural network classifies the data. This is described in section 4, Classification Accuracy.

Traning data selection software

Selecting training data from remote sensing images requires the experience of a trained image analyst. For this method development, ice charting experts carried out the work.

A software extension for the Erdas Imagine image analysis system was developed (see Figure 4.3). The ice experts are used to working in Erdas Imagine for ice charting which was the reason this software was chosen. The GUI-based extension enables the user to load a SAR image and annotate it with a number of training classes. The training classes are displayed with different colors and are assigned a class name. The annotated image is eventually exported into a file format readable to the

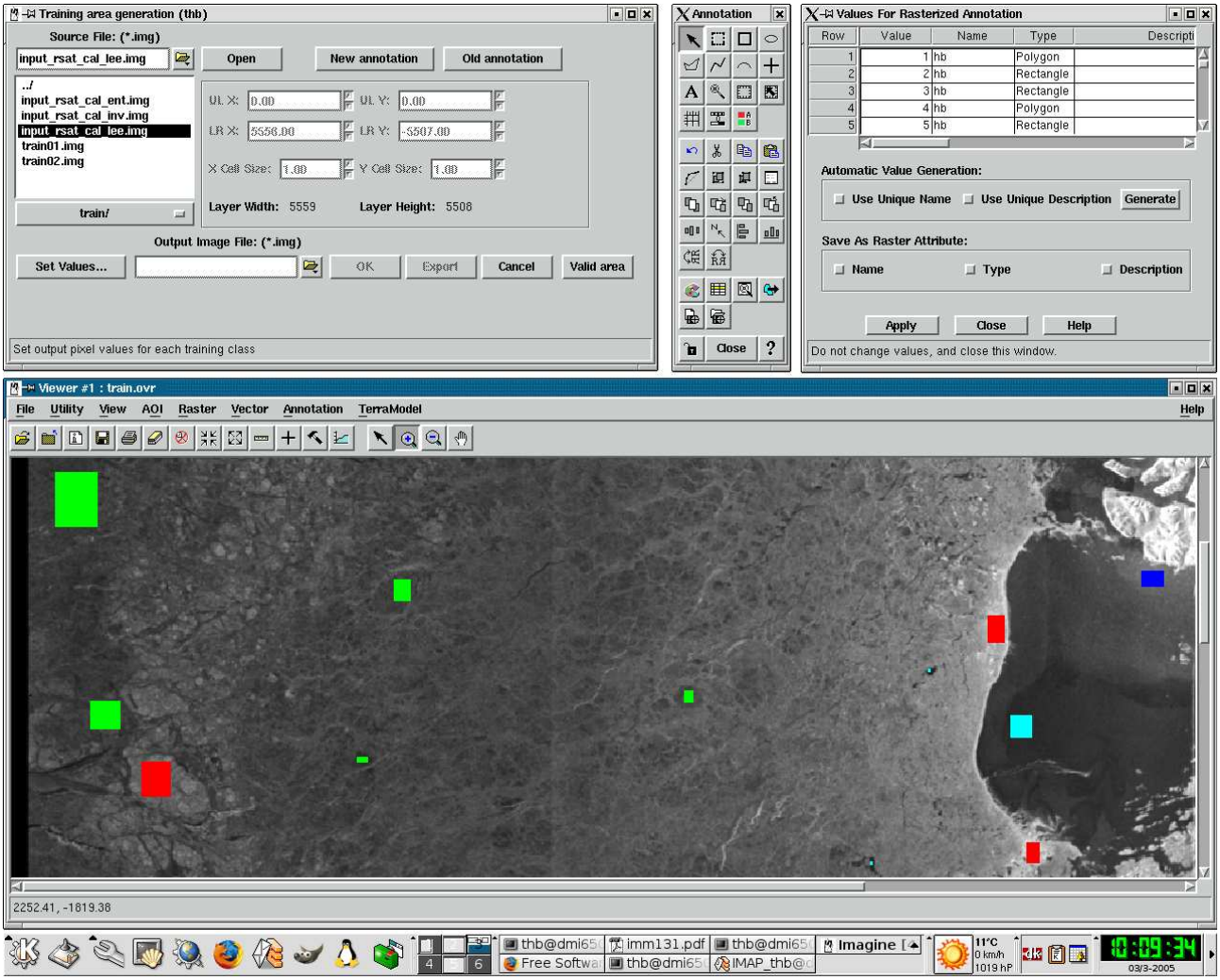


Figure 4.3: Training data selection. An extension to Erdas Imagine. Top left: File management. Top middle: Annotation tools. Top right: Training class management. Bottom: SAR image with training data overlay.

classification software.

Number of classes

For the purpose of sea ice concentration retrieval the number of surface classes can be kept relatively small, as it is not necessary to be able to discriminate between different ice types, e.g. first-year ice and multi-year ice. However, for the clustering of the data in the feature space it is often valuable to have individual classes for each ice type. Later these are combined into one ice class for achieving sea ice concentration.

The choice of classes for the five images analyzed in this study typically consist of two sea ice classes and one or two open water classes. The sea ice classes are: 1) smooth, low backscattering ice, and 2) rough, high backscattering ice. The water regions were either chosen to consist of calm water and turbulent water as separate classes, or in the cases of scene 3, 4, and 5 to only include one water class and ignore the turbulent water class. This is done because of a frequent overlap between turbulent open water and the rough ice classes was found. This is discussed in the Results and Discussion, chapter 5.

Classification method

The choice of classification algorithm is, like the choice of features, not trivial either. This method uses a supervised classification scheme by the means of an artificial neural network. Many statistical models for classification, suffer from the assumption that the data follows a certain probability distribution (Benediktsson and Sveinsson, 1997). Neural networks allows for *model-free* classification that are both fast and flexible. However, there is a trade-off in that the method works much like a black box where the relations between the features and the class is not easily deductible. Neural networks are also known to perform well on data sets of noisy and poorly defined data (Silipo, 1999), which makes it reasonable to use the neural network on SAR data.

A widely applied neural network in the fields of statistical pattern recognition and classification is the Learning Vector Quantization (LVQ) as formulated by Kohonen (2001). Hastie et al. (2001) groups LVQ together with *K-means Clustering* calling them *Prototype Methods*. The training data is represented by a set of points in the feature space, and the classification is based on a distance measure (often the Euclidean distance).

Remote Sensing applications of LVQ have found the algorithm to be accurate and flexible as reported by for example Kaleschke and Kern (2002). For this project the LVQ_PAK implementation Kohonen et al. (1996) of Learning Vector Quantization was used.

The LVQ algorithm is based on the description of the classes by so called *codebook vectors*, which are essentially training pixels viewed as vectors (see Figure 4.4).

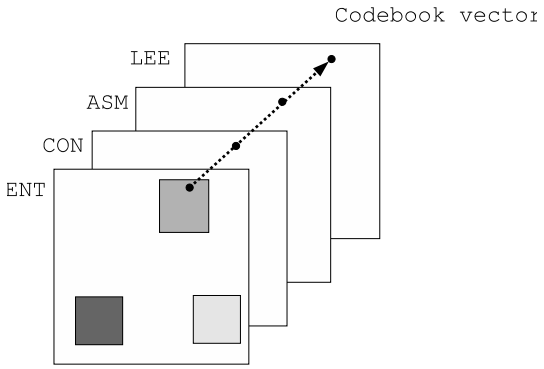


Figure 4.4: Four features and the codebook vector.

LVQ - neural network training

The codebook of vectors is constructed through a *learning or training phase* in which the feature vectors are adjusted and positioned in an optimal way, so that they do not intermingle and cause overlapping class boundaries.

For the actual training of the LVQ a total of 10 % of the number of training samples per class were chosen as the number of codebook vectors. Thus, the number of codebook vectors was 200 for scene 1-4, but only 130 for scene 5. After initialization of the codebook, the learning phase adjusts the position of the codebook vectors on the basis of input vectors fed from the training data. If the class of the input vector agrees with the class of the closest codebook vector, then the codebook vector is moved *towards* the input vector. This training stage is repeated a predefined number of steps and the distance by with the codebook vectors are moved is decreased simultaneously as specified by the learning rate. Ideally, this iterative process reaches stability.

The codebook which is the result of the learning phase can be illustrated by the *Voronoi tessellation* in which the boundaries between the Voronoi cells constitute the class boundaries. This is shown in Figure 4.5. However, this is the 2D case. In the normal case the boundaries between classes are hyperplanes.

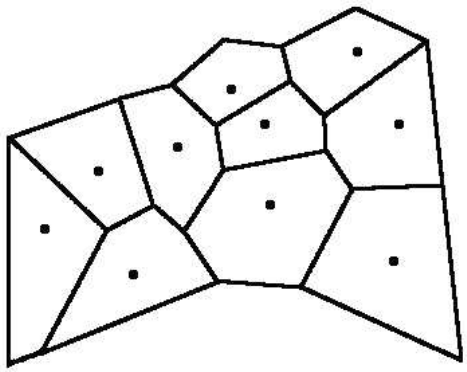


Figure 4.5: The Voronoi tessellation. Dots are optimized codebook vectors, lines are boundaries between classes.

A common phenomenon in the field of neural networks is *over-learning*. Over-learning is what happens when the neural network becomes fine-tuned to the training data and thus will not be able to generalize to new data (Silipo, 1999). However, by using the LVQ3 variant of the LVQ algorithm, over-learning is not of concern. LVQ3 is *self-stabilizing* as described in Kohonen et al. (1996), meaning that the optimal placement of the codebook vectors do not change with extended learning. The number of learning steps was therefore set to a high value of 500.000, the learning rate set to a low value of 0.03, the relative learning rate parameter to 0.1 and the window width parameter to 0.3. The optimization of these parameters for the classification accuracy has so far not been investigated fully in the work of this project.

Classification and its accuracy

After the training by Learning Vector Quantization, the final codebook can be used to classify unlabeled feature vectors. An unlabeled feature vector is assigned the class of the closest codebook vector given by the Euclidean distance between them. This is used to compute the final classified image product which is used for ice concentration calculation.

Assessment of the classification accuracy was done by establishing the *confusion matrix* of the training and the validation vectors. From this an overall accuracy for each classification was computed. In Appendix C the confusion matrices for each classified RADARSAT image can be seen.

A visual way to evaluate on the classification's ability to classify the data, is to make *Sammon plots* (Kohonen et al., 1996). A Sammon plot is a method for mapping high dimensional data onto a lower dimensional space, usually 2-D. By plotting the trained codebooks and their labels in this way, it can be evaluated whether the clusters of classes are overlapping or not. Figure 4.6 shows a Sammon plot of the trained codebook vectors of scene 1. Four well-separable classes are seen, which corresponds well with the high recognition accuracy of that specific classification.

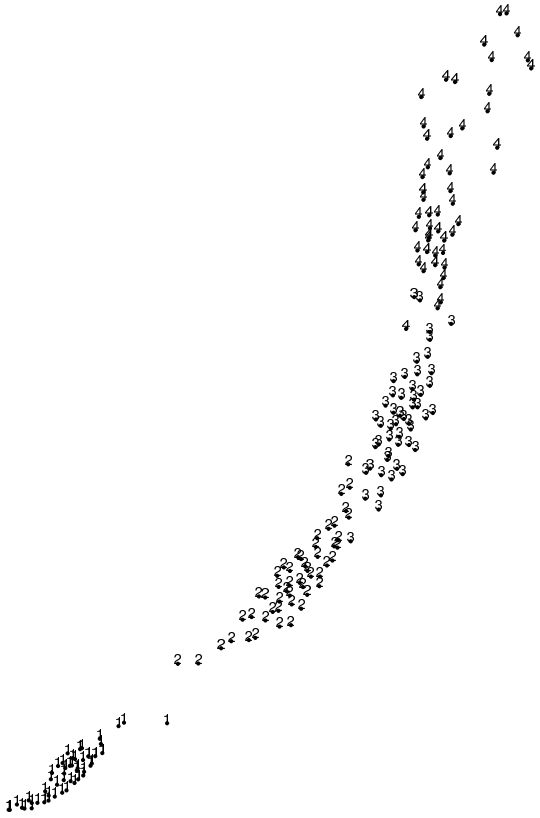


Figure 4.6: Sammon plot for scene 1. The four classes are well-clustered and non-overlapping

Assessment of classification and fine-tuning

After the classification phase, the classified image is assessed by an ice expert by comparing the classified image with the original image and the texture images. If the classification is erroneous, the operator can choose to fine-tune the number and placement of training areas as illustrated in the classification flowchart, Figure (4.1). Signs of the classification being in error can be the overlapping of classes, e.g. poor discrimination between for example high backscattering ice and turbulent water. This is a common problem for sea ice classification from single-polarized SAR data as also reported by Gill (2002).

Choosing different combinations of features, adding or removing training areas, changing the number of surface classes can improve the classification. After the fine-tuning the classification is performed again.

Classification mask

The final step in the classification procedure is to make a classification mask. The mask separates areas of well classified areas from misclassified areas and is based on manual image interpretation. Obvious misclassified areas such as areas of wind roughened water confused with high backscatter sea ice and very calm water confused with low backscatter ice are the main target for the masking. Furthermore, land areas are also masked out, producing an image showing only well classified areas of sea ice and water.

In Figure 4.7 and example of the masking procedure is shown.

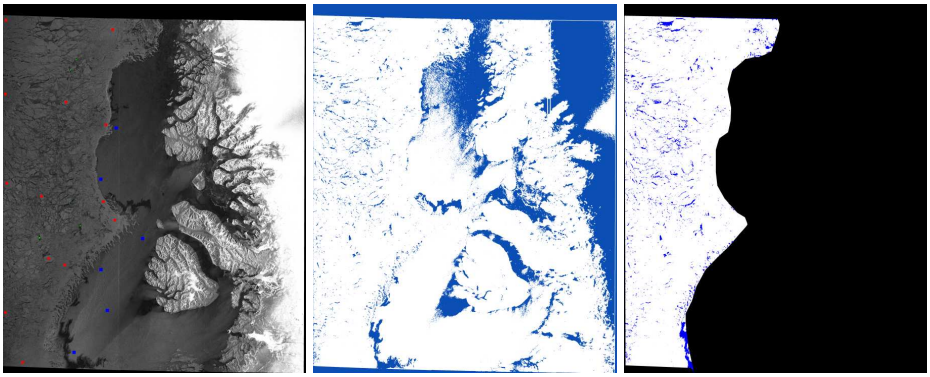


Figure 4.7: Left: Original SAR scene. Middle: Classification result. Right: Masking out misclassified areas.

4. Results and Discussion

This chapter reports on the results of the six different classifications, but firstly the investigations of the choice of texture features and their impact on the classification accuracy is vdescribed.

Selecting optimal number of features and their combinations

One of the experiments in the development of this method was to investigate the effect of the number of texture features and their combinations on the classification accuracy. The experiment was carried out by training the neural network for all possible combinations of the nine features: Angular Second Moment (ASM), Entropy (ENT), Contrast (CON), Dissimilarity (DIS), Homogeneity (HOM), Inverse Difference Moment (IDM), Correlation (COR), Mean (MU) and Lee. This equals 512 different feature combinations for which the recognition accuracy was computed for each class as well as for the average over all the classes. The average over all classes can be considered a good measure for the ability of the individual feature combination to classify the image.

To get an overview of the classification results, the classification accuracy was plotted for each feature (Figure 5.1). A distinct and similar pattern can be seen for each different image. There are two groups of classification accuracies for each image. One group showing a high accuracy and one group showing a lower accuracy. Surprisingly these groups consist of the same combinations of features for all classifications. This means that certain combinations of features consistently yield high accuracies while some combinations consistently yield lower accuracies.

The figure 5.1 also shows that as the number of features increases, the maximum classification accuracy decreases. This support the recommendations of Clausi (2002) to limit the number of texture features in sea ice classification. On this basis it was chosen to analyse the feature combinations of four or less features belonging to the group of high classification accuracies. This group is indicated in Figure 5.1 as the areas in the top left dotted box. These areas each contain the same 153 combinations of features. Thus, the optimal feature combination for the five classified images should on this background be chosen from those groups.

Assigning a score to each combination, by weighing it with the classification accuracy from each image, it was possible to make a list of the top 20 best feature combinations for all five images (Figure 5.1). In Appendix D the top 20 accuracy scores are listed for each individual classification.

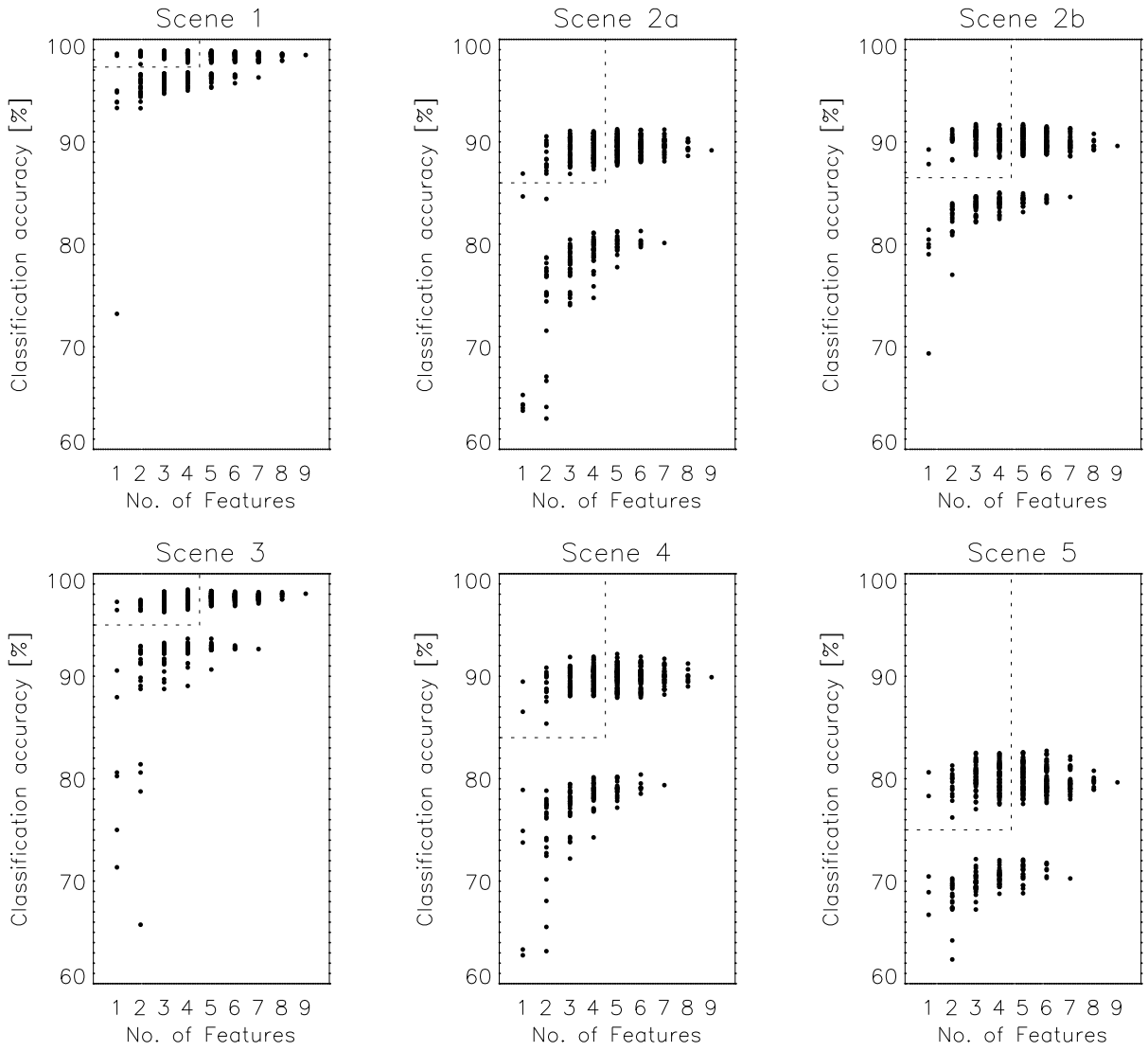


Figure 5.1: Classification accuracy for classifications using different number and combinations of features.

From Appendix D (Table D.1 to D.7) to it can be seen how the contrast (CON), angular second moment (ASM), entropy (ENT) and the lee filtered image (LEE) are consistently useful for classifying the SAR data. Occasionally, the inverse difference moment (INV) and the correlation (COR) features are useful.

Classification and sea ice concentration results

To illustrate the sea ice classification capabilities of the described classification system, this section presents the results of the classification of each of the five RADARSAT scenes. The classified images can be seen in Appendix E, figures E.1 to E.6, where the classes have been merged into only two classes, water (blue) and ice (white), to make the result easier to interpret.

The images were classified with the combination of features which gave the highest recognition score as listed in the top 20 scores for each image in Appendix D.

Image 1. The image was accurately classified with an average accuracy of almost 98%. The four

MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	0	1	1	0	1	1	92.09
0	0	0	1	1	1	0	1	1	92.02
0	1	0	0	1	0	1	1	1	91.97
1	1	0	0	1	0	0	1	1	91.80
0	1	0	0	1	0	0	1	1	91.76
1	0	0	0	0	0	0	1	1	91.74
0	0	0	0	1	0	0	1	1	91.72
1	0	0	0	1	0	0	1	1	91.70
0	0	0	1	1	0	1	1	1	91.68
1	0	0	1	1	0	0	1	1	91.57
1	1	0	0	0	0	1	1	1	91.54
1	1	0	0	0	1	0	1	1	91.51
0	0	0	0	1	1	1	1	1	91.51
0	0	0	1	0	1	0	1	1	91.51
1	0	0	1	0	1	0	1	1	91.48
0	1	0	1	1	0	0	1	1	91.48
0	1	0	0	0	0	1	1	1	91.48
1	0	0	0	1	0	1	1	1	91.46
0	0	0	0	1	0	1	1	1	91.44
0	1	0	1	0	0	1	1	1	91.44

Table 5.1: Top 20 averaged accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

surface classes chosen for the classification (two ice classes and two water classes) proved to be able to classify the image successfully. There were no problems with overlapping classes and very few areas seem to be misclassified.

Image 2a. The image was classified twice with different training areas (see Figure A.2 and A.3) to illustrate the fine tuning capabilities of the classification scheme. The first classification (scene 2a) used training samples from the entire image and as can be seen, this resulted in poor recognition of vboth sea ice and the open water in between the ice floes. The reason for this is probably overlapping of the rough ice and turbulent water classes. The sea ice areas are underestimated and therefore the sea ice concentration would not be representative for the area.

Image 2b. In the second attempt to classify the image, the wind roughened water area in the west of the image was ignored from the classification and no training pixels were selected in this area. Furthermore, additional training areas for the low concentration sea ice were added to strengthen the recognition of this class. This resulted in an improved recognition of the sea ice class (Figure E.3), and the recognition accuracy was increased from 90.55% of the scene 2a to 98.38% for the scene 2b classification.

Image 3, 4 and 5. These three images were classified with the purpose of retrieving good ice concentration measures from the ice covered areas of the images. Therefore the classification result over the open water regions were not considered to be of importance to the success of the classifications. The result for scene 3 (Figure E.4) suffered from very low contrast in the far range of the image, but still gave a high recognition accuracy of 98%. Scene 4 (Figure E.5) shows trouble in classifying the sea ice correctly in the far range (left hand side) of the image due to similar low contrast of intensity values. However, for most of the ice covered area of the image a good estimate of the sea ice concentration is achieved. The recognition accuracy was 89%. Scene 5 gave a low recognition accuracy (80%) because of misclassified areas of turbulent open water. This should be

corrected in a fine-tuning of the classification, but still the classification result is satisfying over the ice covered areas.

Robustness of classification

A SAR scene was classified twice to test the robustness and precision of the classification algorithm in regards to the selection of training areas. It is important that the ice concentration retrieval method is not dependent on the subjective selection of training areas.

The RADARSAT ScanSAR Wide scene from June 21st 2004 was classified twice by two different ice experts. The difficulty of the scene was medium, i.e. no wind-roughened water and high ice concentration. Three classes were used: calm water (class 1), low backscatter ice (class 3), and high backscatter ice (class 4). The comparison of the two classifications can be seen from the error matrix in table 5.2.

Table 5.2: Error matrix, two classifications. Number of pixels for classes 1-4.

	1	3	4	Totals
1	8375543	29206	2	8404749
3	354177	10639648	277474	11271299
4	267422	682735	10131684	11081843
Totals	8997142	11351589	10409160	30757891

From the error matrix the inter-classification precision can be computed:

ice/water confusion: (354177+267422+29206+2)/30757891 = 2.1%

The results show that the algorithm is very robust towards the choice of training data in that the ice and water classes are classified with a precision of 2.1%.

5. Conclusion

The classification system described in this report allows for retrieval of high accuracy sea ice concentration for at least parts of single-polarized RADARSAT ScanSAR Wide scenes.

Second-order grey-level statistics are used as features in a supervised classification scheme based on learning vector quantization. Training data is selected manually by expert ice analysts using a custom training selection tool. After optional fine-tuning of the training data a binary mask is produced to mask out areas of misclassified pixels.

A special investigation into the classification accuracy for a wide range of texture feature combinations was carried out. Certain features consistently yield high classification accuracies, namely the contrast, the angular second moment, the entropy, and the lee filtered image. Other features occasionally proved useful: the inverse difference moment and the correlation. Using too many features gave rise to lower recognition accuracies.

From the classification of the five RADARSAT images it was shown that special care has to be taken in the selection of training data. The image analyst must be aware of the limitations of the data, that is, know which classes are ambiguous in the feature space. In particular the rough sea ice class and

the turbulent water classes, but also thin ice and calm water classes, constitute potential class pairs of confusion.

A recommended solution is to ignore areas of turbulent water and concentrate the classification on the areas of sea ice and not too wind roughened water. Subsequently, high resolved sea ice concentration can be retrieved from these subsets of the image.

References

Barber, D. G. and LeDrew, E. F. (1991). Sar sea ice discrimination using texture statistics: A multivariate approach. *Photogrammetric Engineering and Remote Sensing*, 57(4):385–395.

Benediktsson, J. and Sveinsson, J. (1997). Feature extraction for multisource data classification with artificial neural networks. *International Journal of Remote Sensing*, 18(4):727–740.

Campbell, J. B. (1996). *Introduction to Remote Sensing*. Taylor & Francis, London, United Kingdom, second edition.

Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62.

Clausi, D. A. and Jernigan, M. (1998). A fast method to determine co-occurrence texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1):298–300.

Gill, R. S. (2002). SAR ice classification using fuzzy screening method. Scientific Report 02-12, Danish Meteorological Institute, Copenhagen, Denmark.

Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on System, Man, and Cybernetics*, 3(6):610–621.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning; data mining, inference, and prediction*. Springer-Verlag, Berlin, Heidelberg, Germany.

IOMASA (2002a). IOMASA - integrated observing and modeling of the artic sea ice and atmosphere. description of work. Technical report.

IOMASA (2002b). IOMASA - phase 1 report for part 4 (sea ice concentration). Technical report.

Kaleschke, L. and Kern, S. (2002). ERS-2 SAR image analysis for sea ice classification in the marginal ice zone. In *Proceedings of the 2002 International Geoscience and Remote Sensing Symposium (IGARSS '02), Toronto*.

Kern, S., Kaleschke, L., and Clausi, D. A. (2003). A comparison of two 85 GHz SSM/I ice concentration algorithms with AVHRR and ERS-SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10):2294–2306.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer-Verlag, Berlin, Heidelberg, Germany, third edition.

Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J., and Torkkola, K. (1996). LVQ_PAK: The learning vector quantization program package. Technical Report A30, Laboratory of Computer and Information Science. Helsinki University of Technology.

Shokr, M. E. (1991). Evaluation of second-order texture parameters for sea ice classification from radar images. *Journal of Geophysical Research*, 96(C6):10625–10640.

Silipo, R. (1999). Neural networks. In Berthold, M. and Hand, D. J., editors, *Intelligent Data Analysis: An Introduction*. Springer-Verlag, Berlin, Heidelberg, Germany.

Acknowledgments

The work was carried out under the IOMASA-project (Integrated Observation and Modeling of Arctic Sea ice and Atmosphere) which is a 'Research and Technological Development'-project of the European Commission under contract EVK3-CT-2002-00067. The authors would like to thank Lars Kaleschke of the University of Bremen, Germany, for providing source code for the LVQ classification scheme for ERS SAR data.

Appendix

A. Images and training areas

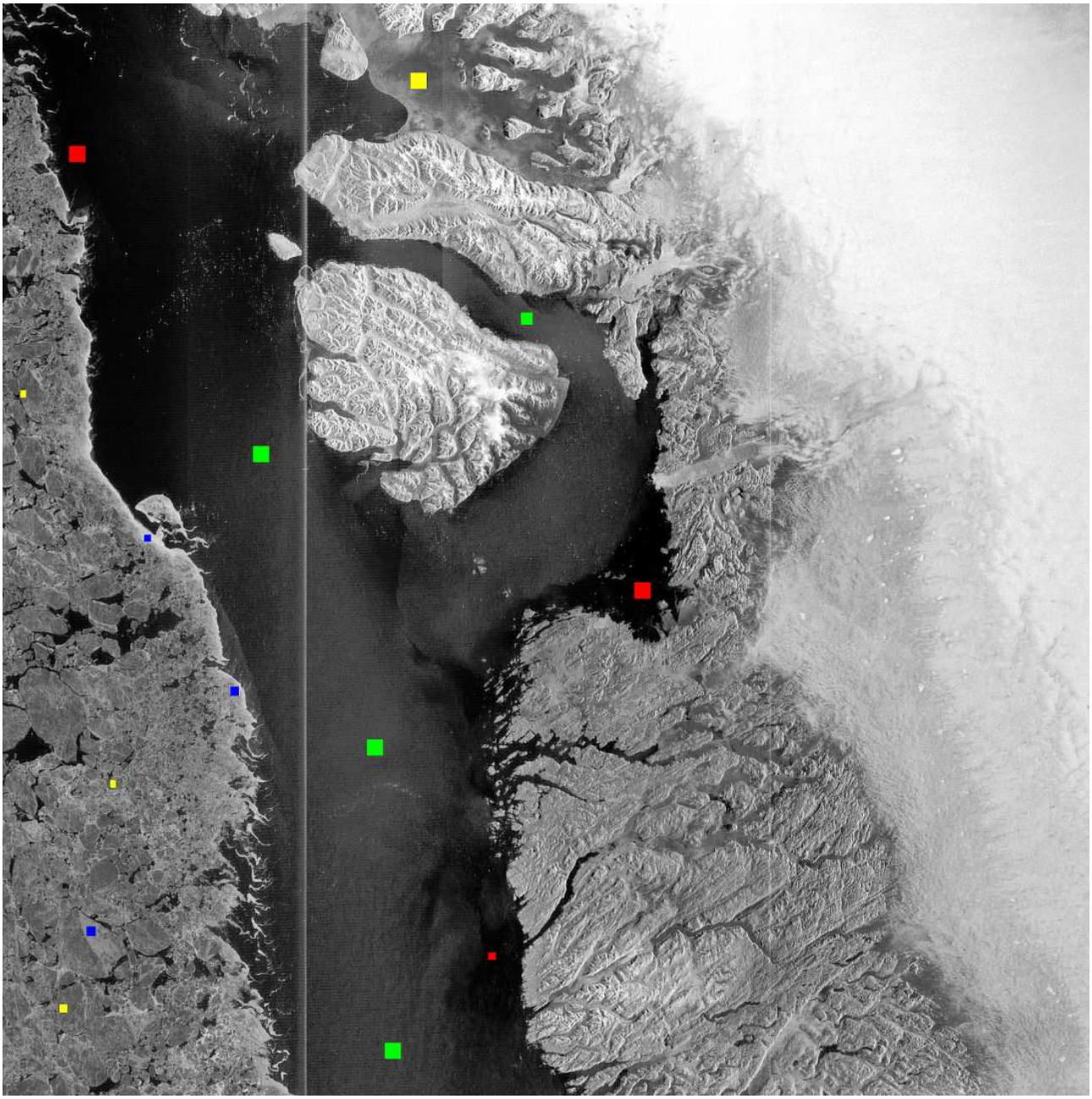


Figure A.1: Scene 1. Amplitude image with training areas. Red = calm water, green = turbulent water, yellow = sea ice smooth, blue = sea ice rough.

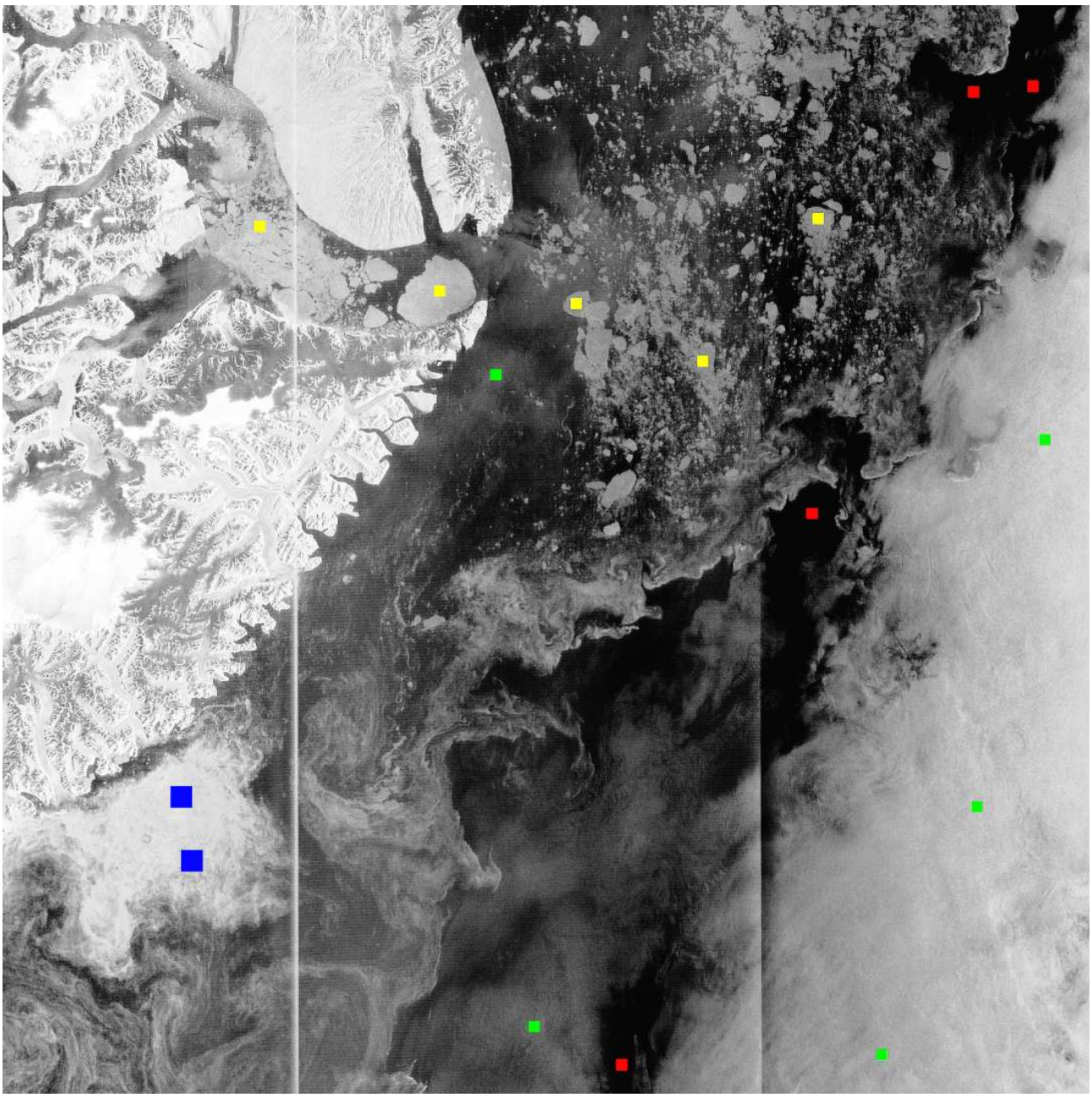


Figure A.2: Scene 2a. Amplitude image with training areas. Red = calm water, green = turbulent water, yellow = sea ice low concentration blue = sea ice high concentration.

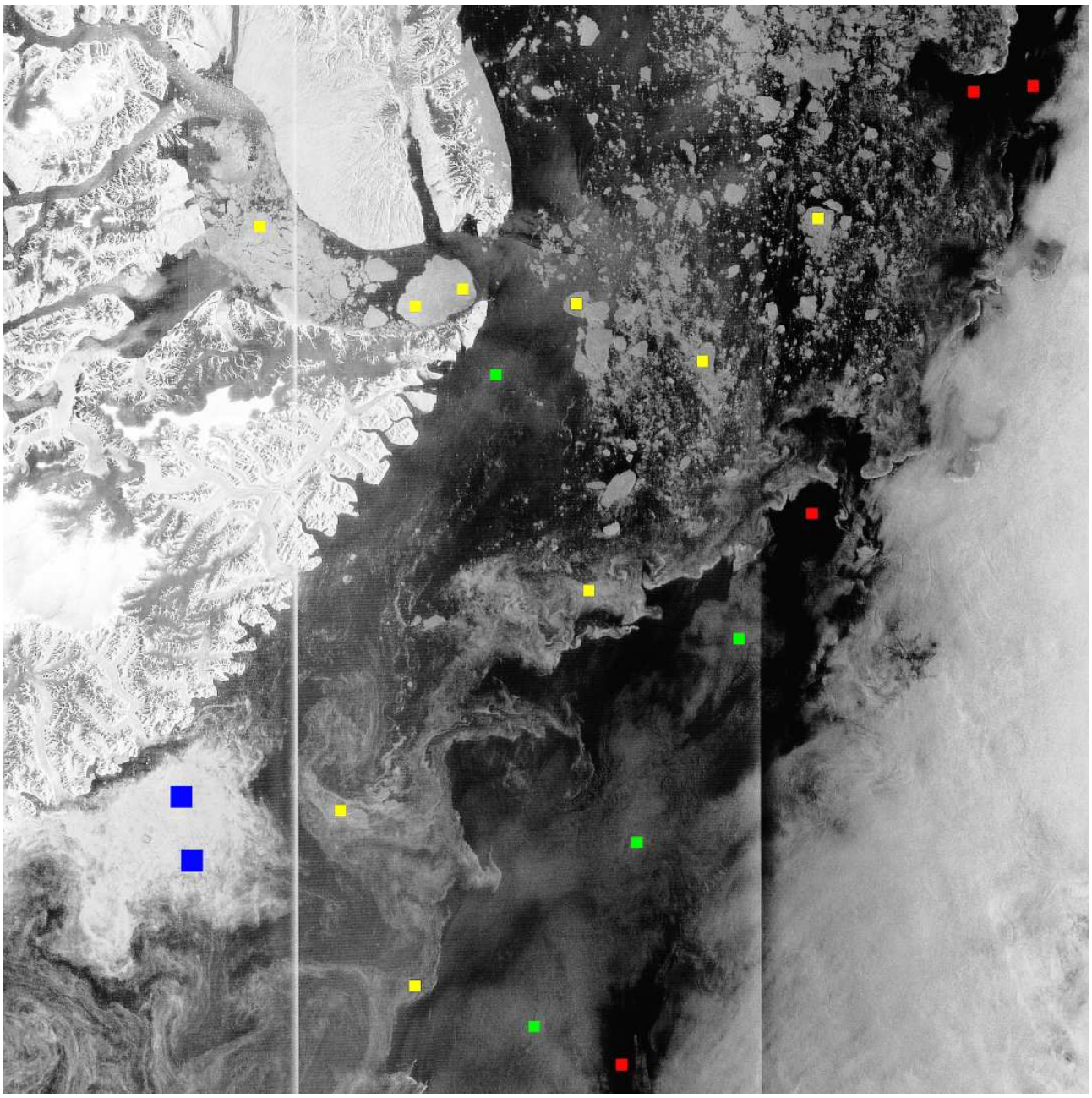


Figure A.3: Scene 2b. Amplitude image with training areas. Red = calm water, green = turbulent water, yellow = sea ice low concentration blue = sea ice high concentration.

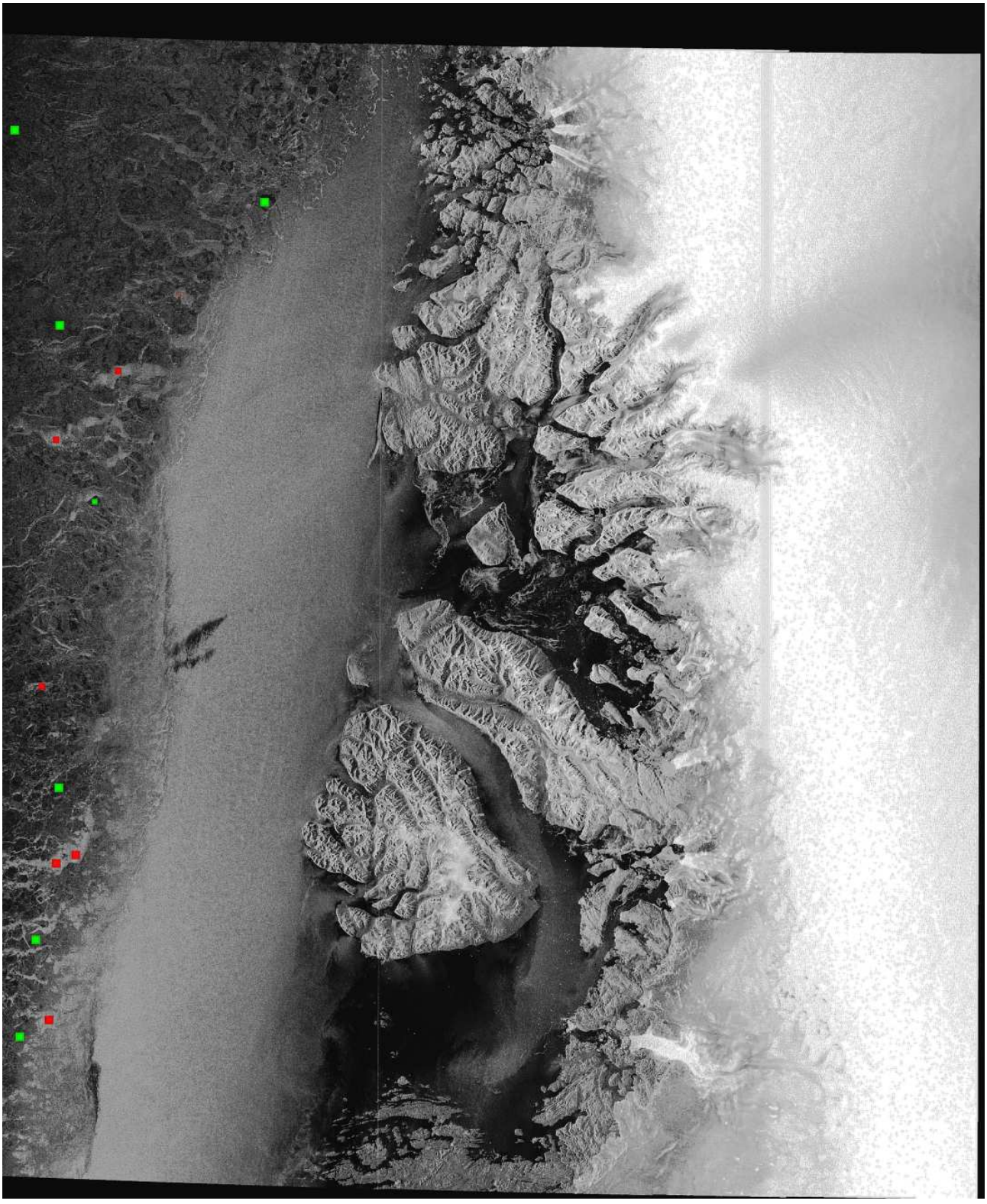


Figure A.4: Scene 3. Amplitude image with training areas. Red = water (turbulent), green = sea ice.

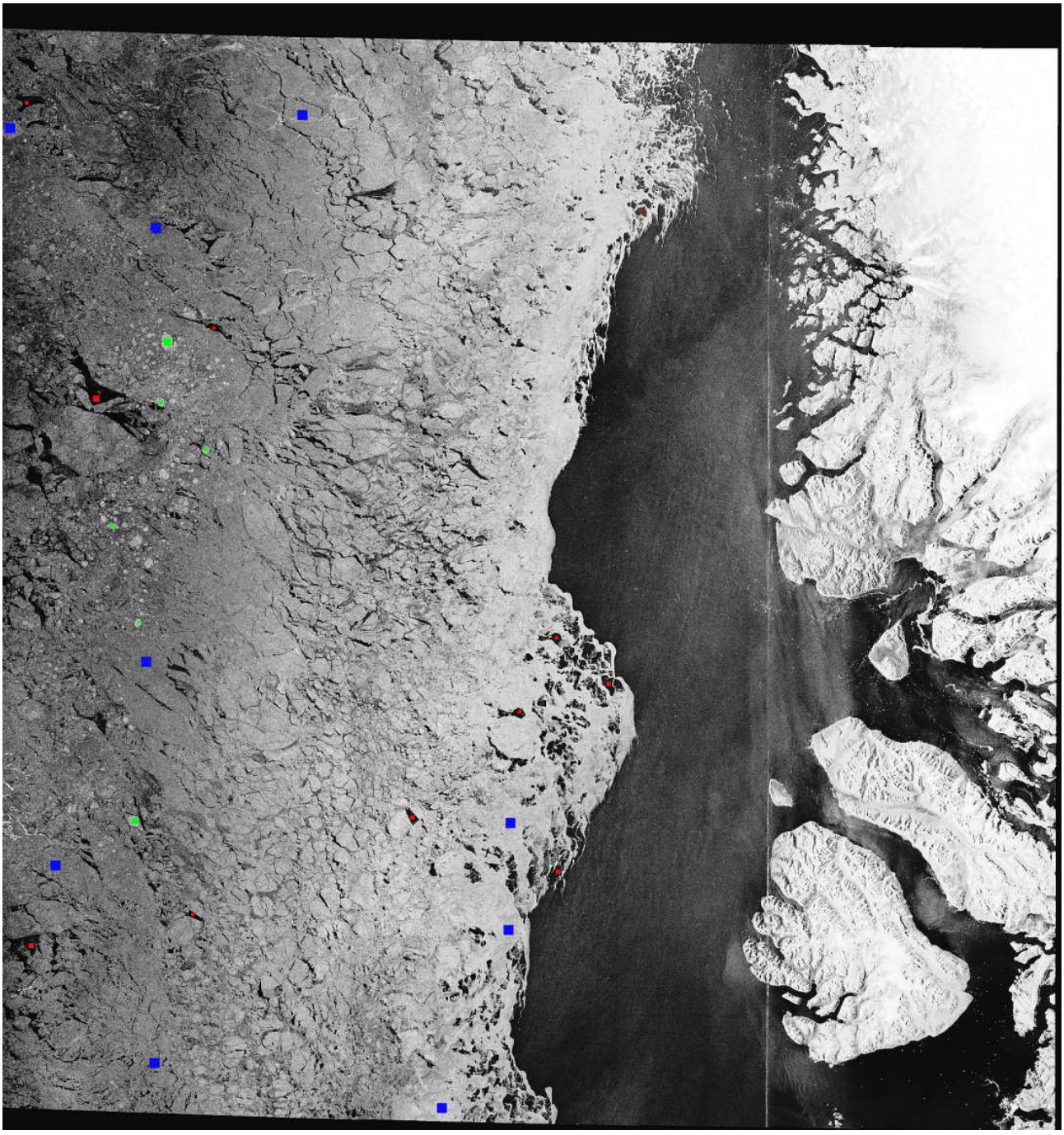


Figure A.5: Scene 4. Amplitude image with training areas. Red = calm water, green = sea ice rough, blue = sea ice smooth.

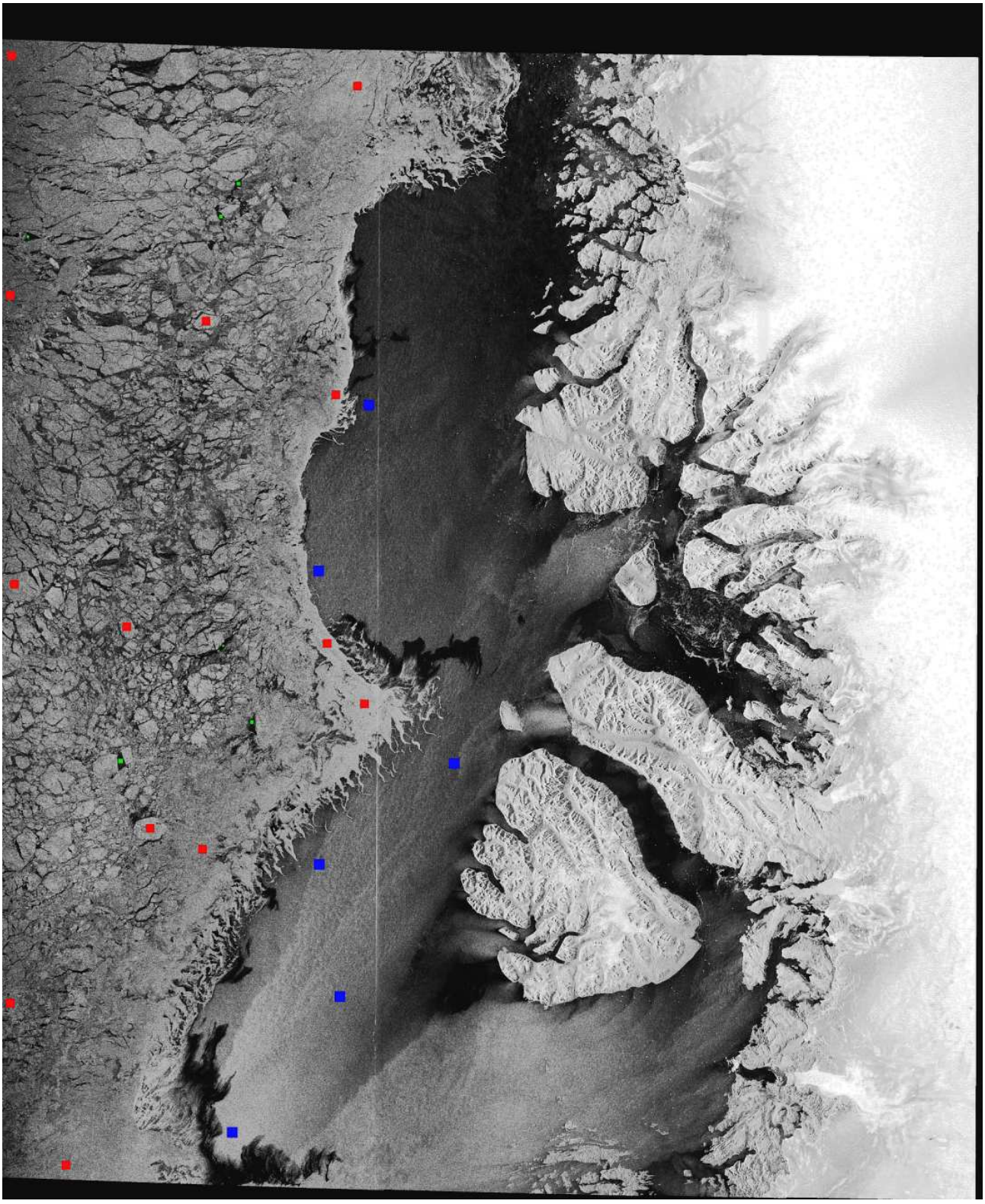


Figure A.6: Scene 5. Amplitude image with training areas. Red = calm water, green = sea ice high backscatter, blue = sea ice low backscatter.

B. Samples of texture images

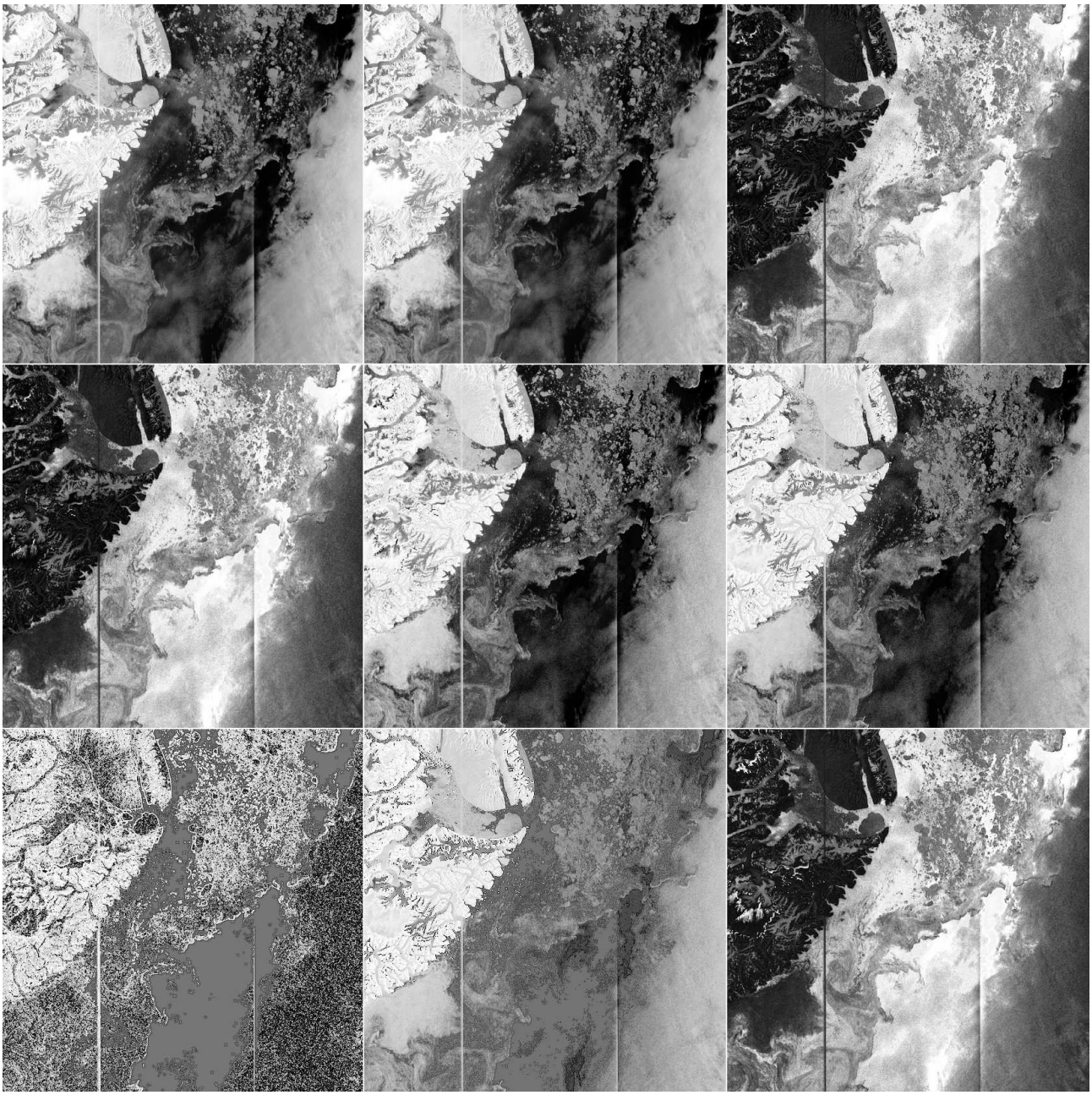


Figure B.1: Feature images for scene 2. Top row from left to right: MU, LEE, INV. Middle row: HOM, ENT DIS. Bottom row: COR, CON, ASM

C. Classification accuracies

```
SCENE 1
              1: 1000 entries 100.00 %
              2: 1000 entries  98.30 %
              3: 1000 entries  93.90 %
              4: 1000 entries  99.60 %

Total accuracy:  4000 entries  97.95 %

-------------------------------------------------------

SCENE 2a
              1: 1000 entries 100.00 %
              2: 1000 entries  80.60 %
              3: 1000 entries  97.60 %
              4: 1000 entries  84.00 %

Total accuracy:  4000 entries  90.55 %

-------------------------------------------------------------
SCENE 2b
              1: 1000 entries 100.00 %
              2: 1000 entries  99.20 %
              3: 1000 entries  98.20 %
              4: 1000 entries  96.10 %

Total accuracy:  4000 entries  98.38 %

-------------------------------------------------------------
```

```
SCENE 3
              1: 1000 entries  96.50 %
              2: 1000 entries  99.50 %

Total accuracy:  2000 entries  98.00 %

-------------------------------------------------------------

SCENE 4
              1: 1000 entries  99.20 %
              2: 1000 entries  86.80 %
              3: 1000 entries  81.10 %

Total accuracy:  3000 entries  89.03 %

-------------------------------------------------------------

SCENE 5
              1:  650 entries  67.69 %
              2:  650 entries  99.69 %
              3:  650 entries  73.08 %

Total accuracy:  1950 entries  80.15 %

-------------------------------------------------------------
```

D. Top 20 accuracy and feature combinations scores

Table D.1: Top 20 averaged accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

ALL									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	0	1	1	0	1	1	92.09
0	0	0	1	1	1	0	1	1	92.01
0	1	0	0	1	0	1	1	1	91.97
1	1	0	0	1	0	0	1	1	91.80
0	1	0	0	1	0	0	1	1	91.76
1	0	0	0	0	0	0	1	1	91.74
0	0	0	0	1	0	0	1	1	91.71
1	0	0	0	1	0	0	1	1	91.70
0	0	0	1	1	0	1	1	1	91.68
1	0	0	1	1	0	0	1	1	91.57
1	1	0	0	0	0	1	1	1	91.54
1	1	0	0	0	1	0	1	1	91.51
0	0	0	0	1	1	1	1	1	91.51
0	0	0	1	0	1	0	1	1	91.51
1	0	0	1	0	1	0	1	1	91.48
0	1	0	1	1	0	0	1	1	91.48
0	1	0	0	0	0	1	1	1	91.48
1	0	0	0	1	0	1	1	1	91.46
0	0	0	0	1	0	1	1	1	91.45
0	1	0	1	0	0	1	1	1	91.44

Table D.2: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 1									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	1	1	0	0	0	1	1	98.90
0	1	0	0	1	0	0	1	1	98.88
0	0	1	1	0	0	0	1	1	98.88
0	0	1	0	1	0	0	1	1	98.88
0	1	0	1	1	0	0	1	1	98.85
0	0	0	1	1	0	0	1	1	98.85
0	0	0	0	1	0	0	1	1	98.83
0	1	1	0	1	0	0	1	1	98.80
0	0	0	1	1	1	0	1	1	98.80
1	1	0	0	1	0	0	1	1	98.78
1	0	0	1	0	1	0	1	1	98.78
1	0	0	0	1	0	0	1	1	98.78
0	0	1	1	1	0	0	1	1	98.78
1	1	0	0	0	1	0	1	1	98.75
1	0	1	0	0	0	0	1	1	98.75
0	0	0	0	1	1	0	1	1	98.75
1	0	0	0	0	1	1	1	1	98.72
0	1	0	0	1	1	0	1	1	98.72
0	0	1	0	0	0	1	1	1	98.72
0	0	0	1	1	0	0	1	0	98.72

Table D.3: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 2a									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	0	1	1	0	1	1	91.23
0	1	0	0	1	0	1	1	1	90.77
0	0	0	1	1	1	0	1	1	90.77
0	0	0	0	0	1	1	1	1	90.77
1	1	0	0	0	0	1	1	1	90.53
0	0	0	1	1	0	1	1	1	90.53
0	0	0	0	0	0	1	1	1	90.53
1	0	0	0	0	1	1	1	1	90.50
0	1	0	0	1	0	0	1	1	90.47
1	1	0	0	1	0	0	1	1	90.43
1	0	0	1	1	0	0	1	1	90.43
1	0	0	0	1	0	0	1	1	90.43
0	0	0	1	0	0	1	1	1	90.37
0	0	0	0	1	0	0	1	1	90.37
0	0	0	0	0	1	0	1	1	90.30
1	0	0	1	0	1	0	1	1	90.27
1	0	0	1	0	0	1	1	1	90.27
1	0	0	1	0	0	0	1	1	90.27
0	1	0	0	0	0	1	1	1	90.27
1	0	0	0	0	0	1	1	1	90.23

Table D.4: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 2b									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	1	1	0	0	1	1	91.35
0	0	0	1	1	0	1	1	1	91.35
0	0	0	1	1	1	0	1	1	91.33
0	0	0	1	0	0	1	1	1	91.33
0	1	0	1	0	0	1	1	1	91.30
1	0	0	0	1	0	0	1	1	91.25
0	1	0	0	1	1	0	1	1	91.20
1	0	0	1	1	0	0	1	1	91.17
0	1	0	0	1	0	0	1	1	91.12
1	0	0	1	0	1	0	1	1	91.08
0	0	0	0	1	1	1	1	1	91.08
1	0	0	0	0	0	0	1	1	91.05
0	0	0	0	1	1	0	1	1	91.05
0	0	0	0	1	0	0	1	1	91.03
1	0	0	0	1	1	0	1	1	91.00
1	0	0	1	1	0	0	1	0	90.97
1	0	0	1	0	0	0	1	1	90.95
0	1	0	0	1	0	1	1	1	90.95
1	1	0	0	0	0	0	1	1	90.92
0	1	0	1	0	1	0	1	1	90.92

Table D.5: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 3									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	1	0	0	0	1	1	1	98.30
0	0	1	0	1	1	0	1	1	98.25
0	1	1	0	0	0	0	1	1	98.15
0	1	1	1	0	0	0	1	1	98.10
0	1	1	0	0	1	0	1	1	98.10
0	0	1	0	0	1	1	1	1	98.10
1	1	1	0	0	0	0	1	1	98.05
1	0	1	0	0	0	1	1	1	98.05
0	0	1	1	0	0	1	1	1	98.05
0	1	1	1	0	0	0	1	0	97.95
0	0	1	0	0	0	1	1	1	97.95
0	0	1	0	0	0	1	1	0	97.95
0	0	0	0	1	1	1	1	1	97.95
1	0	1	1	0	0	0	1	1	97.90
1	0	1	1	0	0	0	1	0	97.90
1	0	0	0	1	0	1	1	1	97.90
0	0	1	1	0	1	0	1	1	97.90
0	0	1	1	0	0	0	1	1	97.90
0	0	0	0	1	0	1	1	1	97.90
1	0	1	0	0	1	0	1	0	97.85

Table D.6: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 4									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	0	1	0	1	1	1	91.83
1	1	0	0	1	0	0	1	1	91.47
0	1	0	0	1	1	0	1	1	91.43
0	0	0	1	1	1	0	1	1	91.40
0	1	0	0	1	0	0	1	1	91.27
1	0	0	1	1	0	0	1	1	91.23
1	1	0	0	0	0	1	1	1	91.03
0	0	0	1	1	0	1	1	1	91.00
0	1	0	1	0	0	1	1	1	90.93
0	0	0	0	1	0	0	1	1	90.87
0	1	0	0	0	0	1	1	1	90.83
0	0	0	1	1	0	0	1	1	90.80
1	0	0	1	0	0	1	1	1	90.73
1	0	0	0	1	0	0	1	1	90.67
1	1	0	1	0	0	0	1	1	90.63
0	0	0	0	1	1	0	1	1	90.63
0	0	0	1	0	0	1	1	1	90.60
0	1	0	0	0	1	1	1	1	90.57
0	0	0	0	1	0	1	1	1	90.57
0	1	1	0	1	0	0	1	1	90.53

Table D.7: Top 20 accuracy scores and their feature combinations. 1 indicates that the feature was used, 0 indicated that it was omitted.

Scene 5									
MU	LEE	INV	HOM	ENT	DIS	COR	CON	ASM	ACCURACY
0	1	0	0	1	1	0	1	1	82.56
0	0	0	1	1	1	0	1	1	82.51
0	0	0	1	0	1	0	1	1	82.36
1	0	0	0	0	0	0	1	1	82.05
1	1	0	0	0	1	0	1	1	81.90
0	1	0	0	1	1	0	1	0	81.74
0	1	0	0	1	0	1	1	1	81.74
0	1	0	0	0	1	0	1	1	81.74
1	1	0	0	1	0	0	1	1	81.69
0	0	0	1	0	1	1	1	1	81.69
0	0	0	0	1	0	0	1	1	81.69
0	1	0	0	0	1	0	1	0	81.64
0	0	0	1	1	1	0	1	0	81.64
1	0	0	0	1	0	0	1	1	81.59
0	0	0	0	1	1	1	1	1	81.59
1	0	0	0	1	1	0	1	1	81.54
1	0	0	0	0	0	1	1	1	81.49
0	0	0	0	0	1	1	1	1	81.49
1	0	0	1	0	1	0	1	1	81.44
0	0	0	0	1	1	0	1	1	81.33

E. Classified images

Figure E.1: Scene 1. Final classification. Blue = water, white = sea ice.

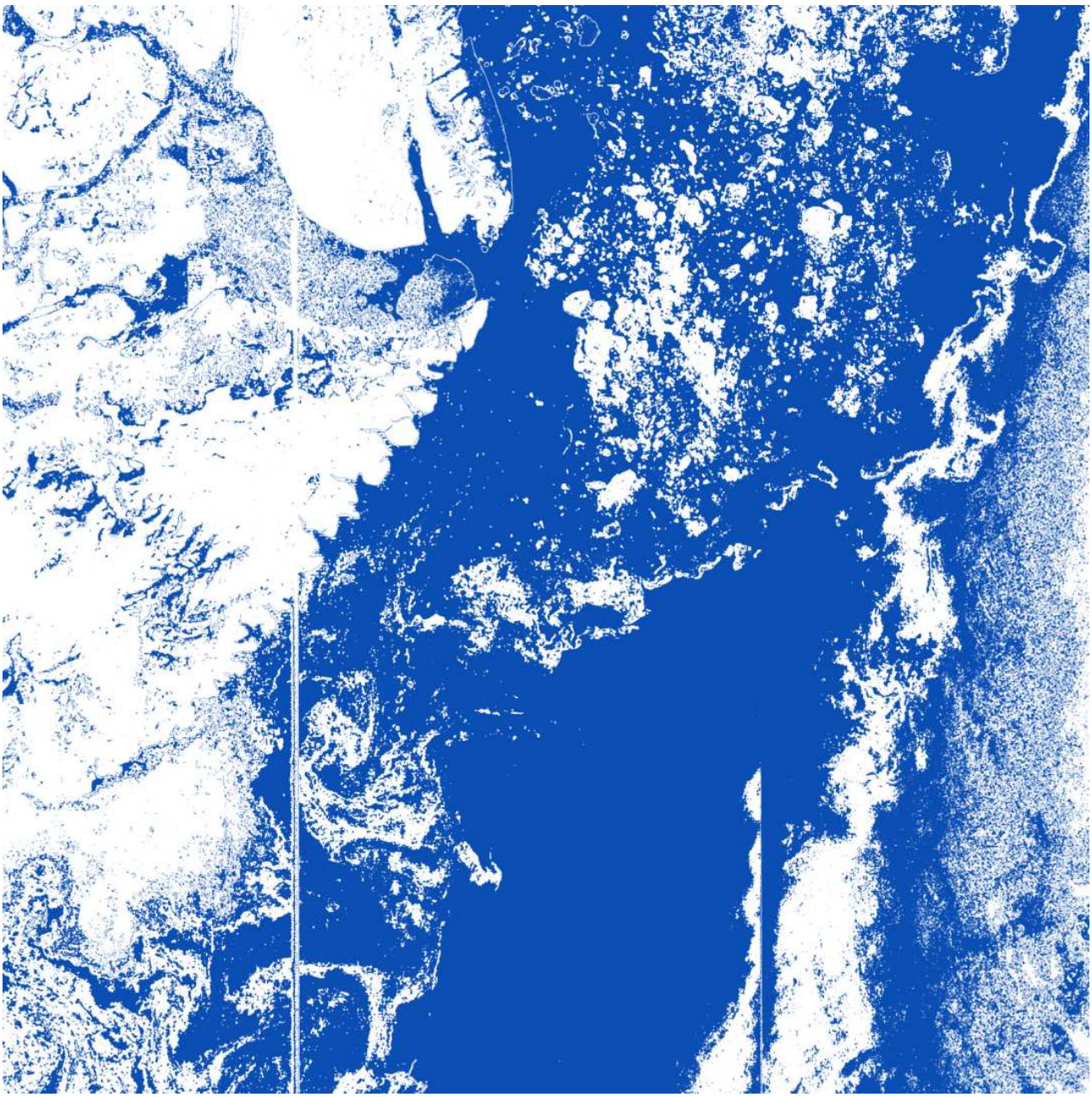


Figure E.2: Scene 2a. Final classification. Blue = water, white = sea ice.

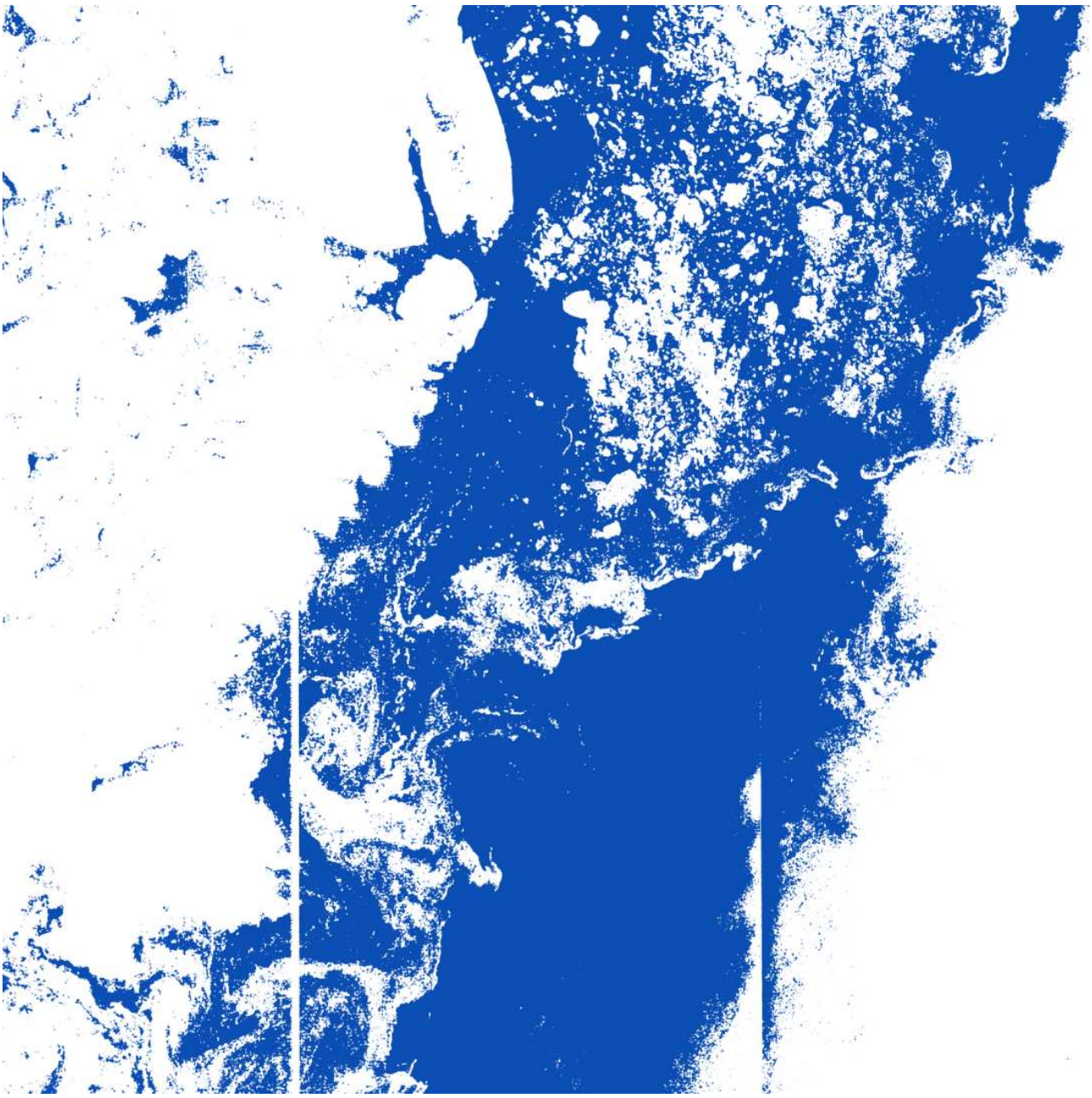


Figure E.3: Scene 2b. Final classification. Blue = water, white = sea ice.

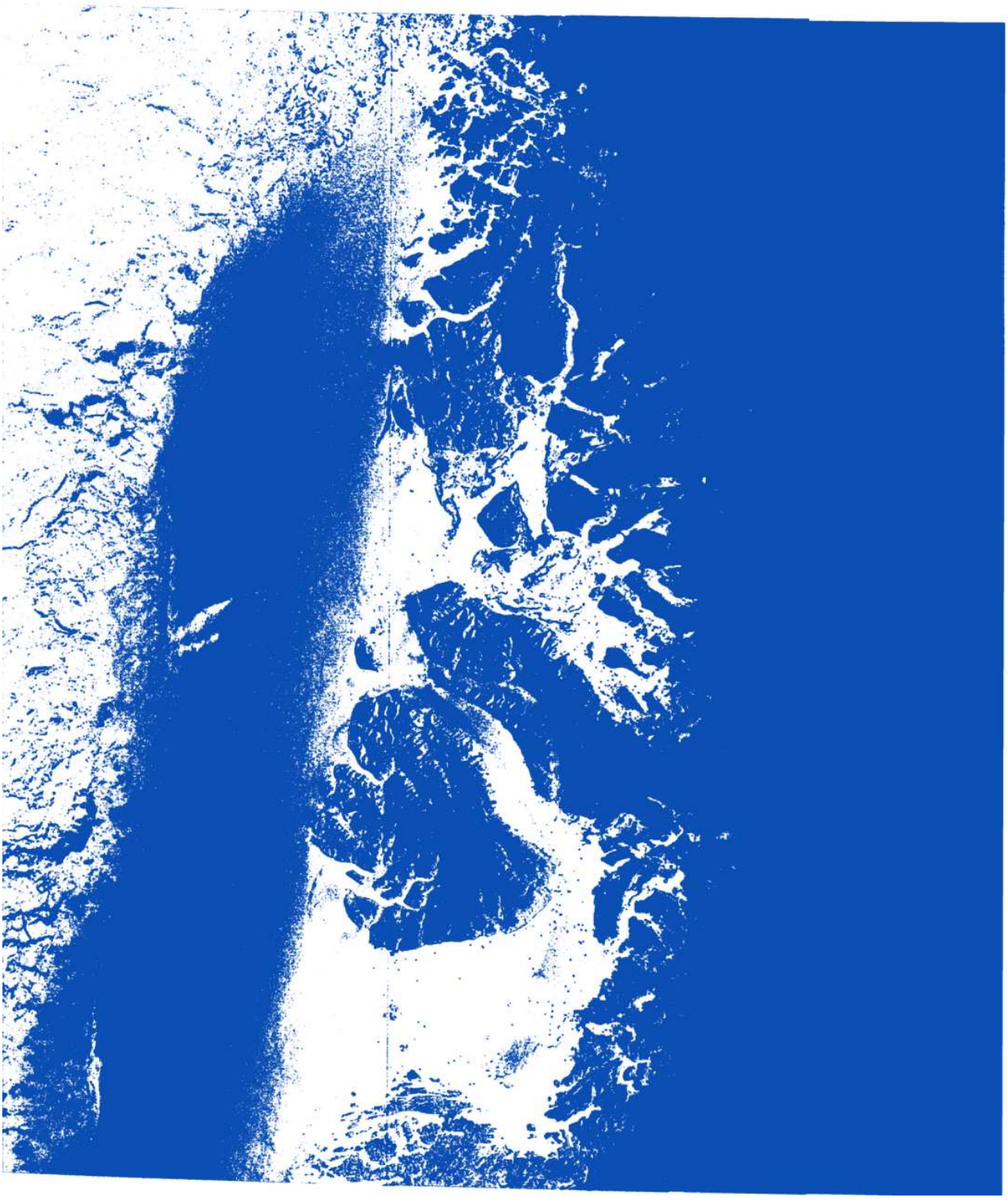


Figure E.4: Scene 3. Final classification. Blue = water, white = sea ice.

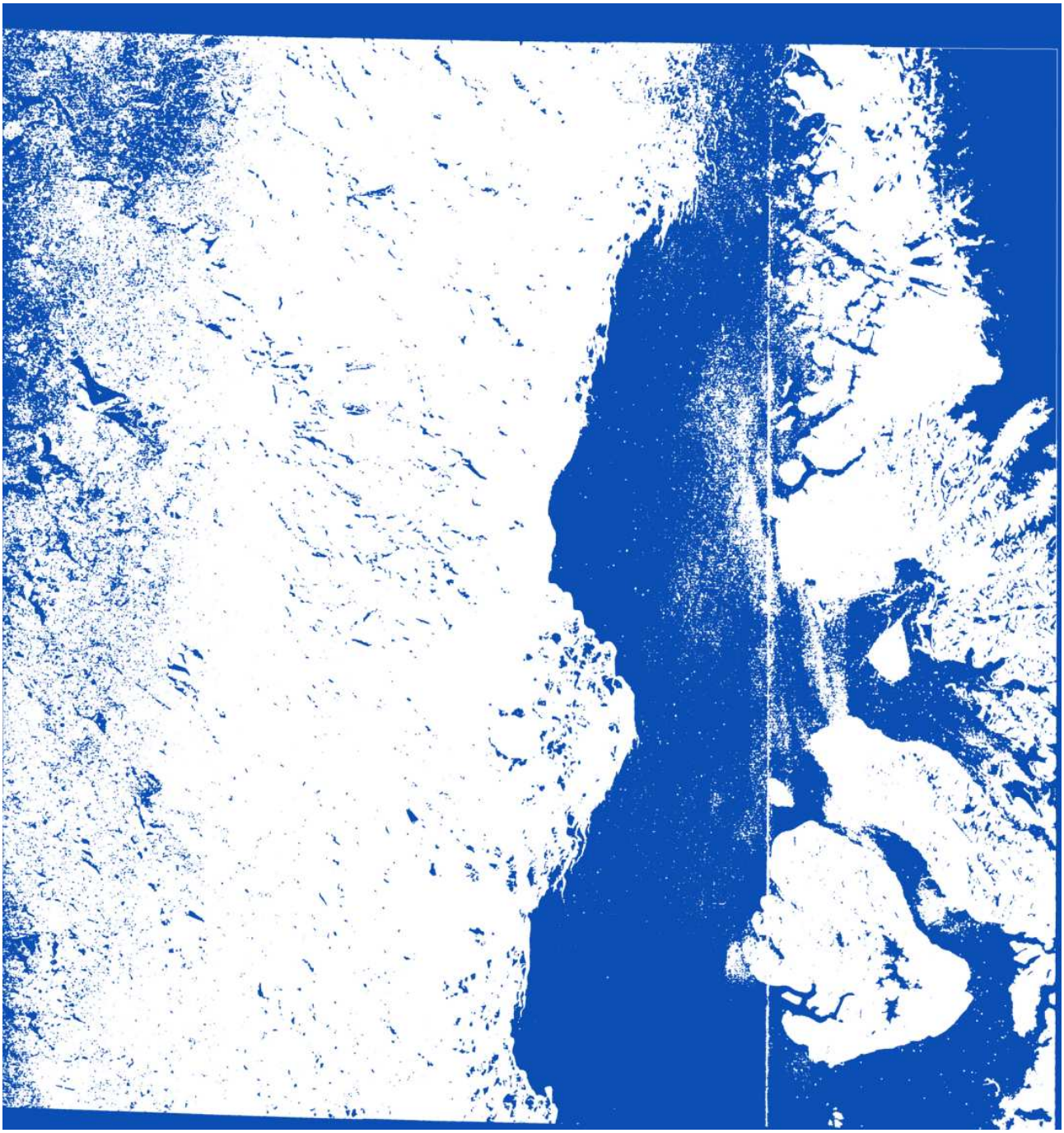


Figure E.5: Scene 4. Final classification. Blue = water, white = sea ice.

Figure E.6: Scene 5. Final classification. Blue = water, white = sea ice.

Previous reports

Previous reports from the Danish Meteorological Institute can be found on:
http://www.dmi.dk/dmi/dmi-publikationer.htm