PIXEL-BASED SEA ICE CLASSIFICATION USING THE MAGSIC SYSTEM

Philippe MAILLARD[a], **David A. CLAUSI**[b],

[a]Cartografia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil, philippe@cart.igc.ufmg.br

[b]System Design Engineering, University of Waterloo, Ontario, N2L 3G1, Canada, dclausi@engmail.uwaterloo.ca

KEYWORDS: Snow Ice, Segmentation, Reasoning, Classification, Texture, RADARSAT, Navigation.

ABSTRACT

MAGSIC is an operation-oriented system in development dedicated to map-guided classification of sea ice for navigation route planning and meteorological modelling. It has already produced promising results in difficult situations such as the Gulf of Saint-Lawrence in late winter. The Canadian Ice Service (CIS) produces ice maps made of large regions with relatively homogeneous concentrations of different ice types. MAGSIC uses the information of these maps to produce a pixel-based (rather than region-based) ice map by labelling a Markov random field (MRF) segmentation of RADARSAT-1 data along with its derived texture features. The system uses a novel implementation of MRF segmentation in combination with a unique labelling approach based on "cognitive reasoning". Although reasonably successful, the system often had difficulties identifying ice type that required cues based on the shape recognition of large ice floes or leads. This article aims at thoroughly testing the MAGSIC system using validation data acquired during the "2003 Gulf of Saint-Lawrence SAR Validation Field Program" performed by CIS. Some new features were also added to MAGSIC and were evaluated. Results suggest a reasonable success and that the errors can be partially attributed to the generalized nature of the analysts´interpretation and to the difficulties of obtaining concurring ground and image data. They suggest that classification metrics that can compare sample distributions were slightly superior for labelling purposes but this could not be confirmed statistically.

RÉSUMÉ

MAGSIC est un système à vocation opérationnel en développement dédié à la classification guidée de la glace de mer pour des fins de navigation et de modélisation météorologique. MAGSIC a déjà produit des résultats promettants dans des situation difficiles comme celle du Golf du Saint-Laurent à la saison d'hiver tardive. Le Service Canadien des Glaces (SCG) produit des carte de glaces composées de grande régions ayant des concentrations relativement homogènes de divers types de glaces. MAGSIC utilise cette information pour produire des cartes de glace détaillées (au niveau des pixels plutôt que des régions) basées sur l'étiquetage d'une segmentation Markoviennne (MRF) de données RADARSAT-1 et de texture dérivée. Le système profite d'une implantation originale de la segmentation Markovienne combinée à une approche unique de l'étiquetage basée sur le "raisonnement cognitif". Malgré les résultats encourageants, le système a souvent du mal à identifier les types de glace nécessitants la reconnaissance de grands floes ou de fractures. Le présent article vise à tester le système MAGSIC à l'aide de données provenant du "Programme de Validation sur le Terrain des Données RSO dans le Golfe du Saint-Laurent (2003)" prodigué par le SCG. Quelques nouveaux aspects ont été incorporés à MAGSIC et sont également évalués. Les résultats suggèrent un taux de succès raisonnable et que les erreurs peuvent souvent être attribuées à la nature généralisée de l´interprétation des analistes ainsi qu´aux difficultés d´obtention de données de terrain et d´image simultanées. Ils suggèrent également que les métriques de classification qui tiennent compte de la distribution des données sont légèrement supérieures pour les opération d´étiquettage mais cela ne pu être confirmé statistiquement.

1. INTRODUCTION

Sea ice recognition represents an important remote sensing application throughout the high northern latitudes. For many years it has provided vital maps to ships navigating in ice infested seas (Hall, 1998). More recently, sea ice mapping is playing a growing role in modelling global climate and in providing northern communities with timely information on the sea ice / open water edge (Agnew et al., 1999). In Canada, the Canadian Ice Service (CIS) produces and discloses ice maps daily using RADARSAT images (HH C-band 5.3 GHz) as their primary source of information. The ScanSAR wide mode provides a swath of 500 kilometers and can supply the necessary daily coverage of Canadian territory. Apart from the SAR images, the ice analyst has also access to a wealth of information including helicopter and ship reports, maps from previous days along with other satellite products (*i.e.* MODIS, AVHRR, etc.). The ice analysts prepare World Meteorological Organization (WMO) maps in which regions containing homogeneous concentrations of sea ice types (between one and four out of the possible 14 types) for which a standard codification (called an "egg code") is defined. These regions can have thousands of square kilometers and contain no indication as to the location of each ice type (see Figure 1 for an example of an egg code symbol and map). Even though the analysts work in a highly computerized environment using geographic information system (GIS) and image processing technology, their interpretation work is completely manual and no automated segmentation/classification routines are used to assist them. The main reason for such situation is due to the inherent difficulties in interpreting SAR images due to speckle noise (Raney, 1998), incident angle variations (Mäkynen et al., 2002) and because geometrical and electrical factors vary simultaneously.

Various systems have been proposed to produce ice maps from SAR images in a semi-automated fashion (Soh et al., 2004; Karvonen, 2004) but none are expected to be adopted by important ice services to assist their ice analysts. MAGSIC (Maillard

and Clausi, 2005b) is also one such system under development. Although still embryonic, it has shown promising results and is being encouraged by CIS. The validation of sea ice classification is always problematic and needs to be assessed seriously for any sea ice classification system to be taken seriously.

1.1. Objective

At present, sea ice analysts do not have the time or technology to prepare ice maps at pixel resolution. Such a pixel-based product would be of great value for a number of applications including navigation and meteorological modelling. In a previous paper we have described the development of MAGSIC (MAp-Guided Sea Ice Classification), an automated system that performs pixel based sea ice classification using a SAR image and an egg code map. In this paper, we are using validated data to test the performance of the MAGSIC system to which we have added a few more features. In particular, we are testing a total of five classification metrics for their respective performance. Finally, a voting scheme in which each metric represents a vote is also being tested.

2. PREVIOUS WORK: THE MAGSIC SYSTEM

2.1. System description

MAGSIC is a modular system being developed as an operational tool to be inserted into CIS operations. It does not aim at replacing sea ice analysts but rather to work in parallel to produce added value documents such as pixel based ice maps and more precise statistics on ice type proportions. The originality of the system lies in the fact that instead of trying to imitate the analyst's work, the system uses the manually produced egg code map as *a priori* information to feed a segmentation algorithm and considerably reduce the number of classes within a limited region. This map-guided approach can be regarded as a symbiosis between machine-aided interpretation and ice-analyst: while the first benefits from a simplification, the second sees its work taken one step further.

There are four main components to the system: 1) a map scanning module to store the region polygons and their specific ice types, 2) a Markov random field (MRF) based segmentation module that processes each "egg code" region and splits it into a number of segments that (ideally) correspond to each ice type, 3) a module to compute the statistics of each segment and each egg code region, and 4) a labelling module based on cognitive reasoning. The system and its components are fully described elsewhere (Maillard and Clausi, 2005b) but a short description of the segmentation and labelling components follows.

The segmentation algorithm is described in Deng and Clausi (2005) and is based on an original implementation of Markov random fields (MRF): the "Modified adaptive Markov random field segmentation" (MAMSEG). MRF models inherently describe spatial context: the local spatial interaction among neighboring pixels. This is most appropriate since neighboring pixels are generally not statistically independent but are linked by spatial correlation. Furthermore, MRFs can effectively combine the relative importance of the pixel being considered and its neighborhood. MAMSEG was adopted because of its good performance with SAR data and sea ice. MAMSEG is innovative because it does not fix *a priori* the relative weight of the central pixel and its neighborhood but rather lets it vary with each iteration in the simulated annealing solution. After the image has been segmented in a satisfactory manner (with relation to the number of iterations which typically vary between 50 and 100), statistics are calculated for each segment of all the egg code regions. The statistics include 1) mean, 2) standard deviation (of each feature), 3) covariance matrix (when more than one feature is used) and 4) histogram of each feature.

The labelling process is carried out by an innovative algorithm that alternatively cumulates evidences about class membership of the segments and labels the segments one at a time. If there is not enough evidence to label one particular segment, the algorithm goes on to the next segment. After having tried to label all the segments of all the regions, the algorithm executes another pass and tries once more to label the remaining segments and so on. The basic idea behind the algorithm comes from the fact that the regions have a limited number of classes (between one and five) and there are generally enough evidence to deduce a class membership for all the segments. A special terminology has been proposed for the type of evidence (Maillard and Clausi, 2005a;b):

- First-degree evidence falls into two categories: 1) the egg code region contains only one class or 2) the egg code region contains several classes and all but one have already been solved and assigned. In either case, the association is straight-forward and no additional information is needed to solve the association.
- Second-degree evidence is characterized by the fact that although all or some classes of a region have previously been solved (in other regions), the program still has to find which set of associations is the most likely. For a total of n_c classes, there are $n_c!$ permutations of matching each segment of a particular region to one of the n_c classes. The objective is then to determine which permutation is more likely according to some metric.
- In third-degree evidence, reasoning is based on the fact that while comparing two egg code regions, although no association was previously solved, if only one class is common to both egg code regions (intersection), then one can deduce which is more likely by calculating a distance metric between all pairwise possibilities. The optimal result is retained as the correct association.

2.2. Improvements

MAGSIC is still in its "work in progress" phase but its usefulness has already been demonstrated and has received encouragements by the CIS. MAGSIC has created a "pipeline" between the ice analyst's interpretation and a detailed (pixel-based) ice map through a novel implementation of MRF segmentation and an original labelling solution. Some improvements have been added to MAGSIC since Maillard and Clausi (2005b) have first published its description. Three such improvements are described below.

- **Additional classes**: Egg code symbols can only have a limited number of ice classes. They generally have, at most, three classes but since the percentage of ice is not always 100%, a fourth class, open water, is often considered. It does happen, however, that a fourth ice class is indicated outside (on the right side) the egg code symbol bringing the total number of possible classes to five. This has a direct effect on the results as it has been observed that the greater the number of classes, the greater the risk of error. But it was also observed that some egg code regions actually include more classes than there are in the

egg code symbol because the ice analyst generalizes its interpretation in the final version of the ice code map. We are therefore considering using the non-generalized version of the egg code map to ensure the segmentation algorithm splits the region in the correct number of classes.

- **Classification metrics**: The initial version of MAGSIC included three classification metrics to which we have added two more for a total of five. The list of classification metrics follow.

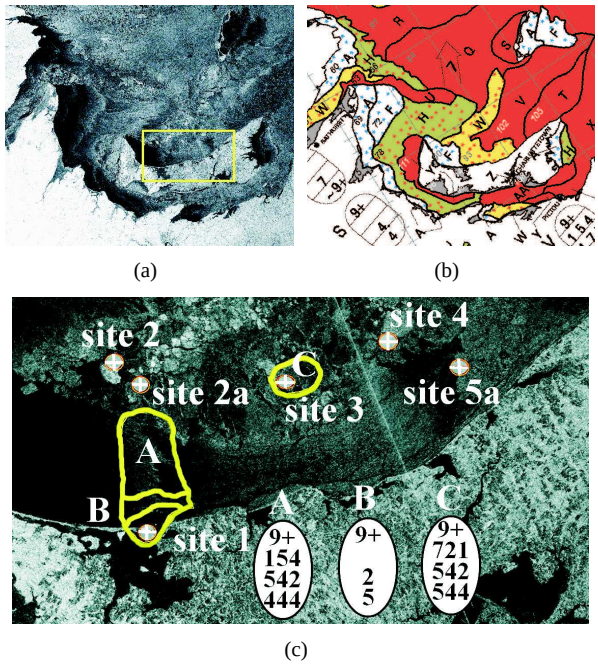
  1. Mahanalobis distance - MD
  2. Fisher criterion - FC
  3. Chi-square test - χ^2
  4. Kolmogorov-Smirnov test - KS
  5. Student's t-test - t

- **A voting scheme**: It has been suggested that selecting the "best" classifier is not necessarily the wisest decision since valuable information may still be produced by less successful classifiers (Wolper, 1992). Considering that a specific classification metric is better adapted to a specific situation, all classifiers can have the potential of being the "best" for some situations. One approach to preserving this information is to pool the output of all classifiers and base the final decision on that stack of information. There are numerous ways to combine the output of different classifiers (Xu et al., 1992) but in the present work, we have chosen, the voting principle as a first experience. According to Xu et al. (1992), there are three level of output from classifier: 1) the abstract level produces a single label, 2) the rank level ranks all possible labels and 3) the measurement level produces a value of likeliness (*e.g.* distance, probability, etc.) for all possible labels. In this initial effort, we have opted for the simple abstract level.

3. STUDY AREA AND DATA

In 2003, the CIS undertook a validation campaign in the Gulf of Saint-Lawrence near Prince Edward Island (PEI) in order to have near real time ground truth data which could serve to calibrate SAR data and assess its real potential in sea ice recognition. Helicopter oblique photos were also acquired along with airborne SLAR, ENVISAT data, RADARSAT-1 data and a wealth of other radar and optical satellite data. The ground sampling was done for three dates: 01, 02 and 06 March 2003. Only 01 March matched the RADARSAT image and, considering the very dynamic nature of sea ice in the St.-Lawrence, data from the two other dates could not be used. Figure 1 shows the location of the study area and the ground sampling sites in the Gulf of St.-Lawrence near PEI.

During winter and spring, strong currents and winds make the Gulf of St.-Lawrence as one of the marine routes that is most dangerous to navigate. The ice can shift so fast that data collected in the morning can be out-dated in the afternoon (D. Flett, CIS, personal communication, October 2004). For these reasons, obtaining good ground data for sea ice in the St.-Lawrence is extremely difficult and the opportunity to use such data bears great value. Ground data was collected in two fashion: punctual sites gave exact ice types for an area of limited extension and zonal sites were recorded as egg code regions with different ice type concentrations. Figure 1c shows both of these and Figure 2 shows the



(a) (b) (c)

Figure 1: Study area: (a) RADARSAT-1 sub image near PEI (validation area in yellow), (b) corresponding egg code map, (c) enlargement of the study area identifying the validation sites. Legend of an egg code: the letter "A" indicates that the symbol applies to region "A": first row indicates the total proportion of ice ($9+ = 90\%$); second row indicates the proportion of each ice type (3 in this case); third row gives the ice code of each type; last row indicates the relative floe size (http://ice-glaces.ec.gc.ca) and (C) validation sites (1 to 5a) and areas (A to C).

helicopter photographs matched with a RADARSAT sub-image for the punctual sampling sites.

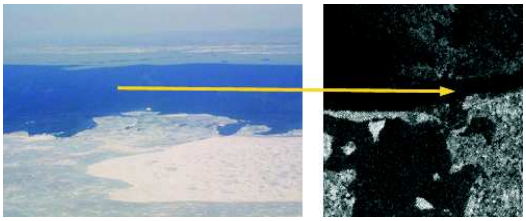
The MAGSIC system was conceived to produce a pixel-based ice map from the combined use of an ice code maps and a SAR image. The ice code map serves a double purpose: 1) it provides information on the number and type of sea ice and 2) it is used as a series of masks (one for each egg code region) for the segmentation algorithm. The sea ice information is stored in a special spreadsheet while the masks are in a raster format constructed from a digital vector version of the egg code map supplied by CIS. The segmentation can incorporate numerous channels of information and it has been found that optimum results were obtained when two texture channels were derived from the SAR image and incorporated in the segmentation process (Deng and Clausi, 2005). The two texture channel are *contrast* and *entropy* produced from the co-occurrence matrix texture paradigm (Haralick et al., 1973).

4. RESULTS AND DISCUSSION

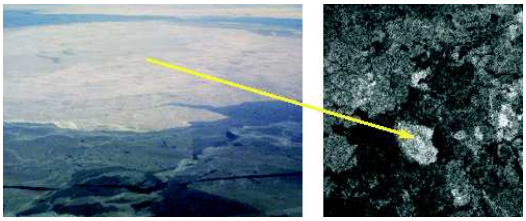
The results are divided in three main topics: the segmentation, the labelling using cognitive reasoning and the labelling using the voting principle.

4.1. Evaluation of image segmentation

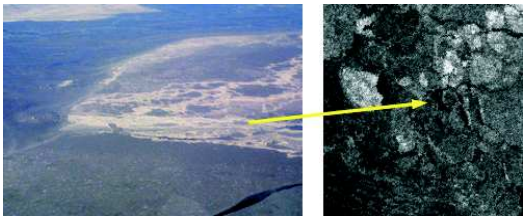
Evaluating the results of a segmentation with sparse validation data is a subjective task. Since the labelling results are evaluated



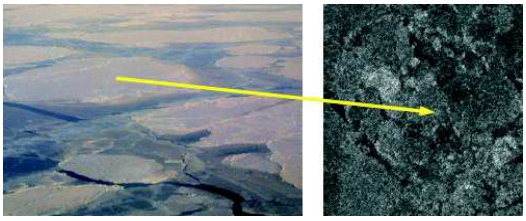
(a) Site 1



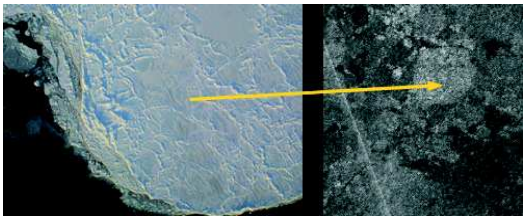
(b) Site 2



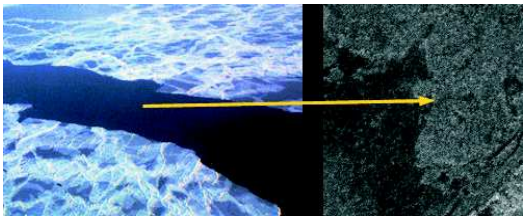
(c) Site 2a



(d) Site 3



(e) Site 4



(f) Site 5a

Figure 2: The six punctual validation sites represented as an helicopter oblique photo and RADARSAT subimage pair: (a) site 1 (fast ice, open water and nilas), (b) site 2 (first year ice and nilas), (c) site 2a (grey and grey-white ice), (d) site 3 (grey, grey-white ice and nilas), (e) site 4 (first year ice and open water) and (f) site 5a (leads in nilas).

using ground truth, the segmentation received only a visual evaluation. Figure 3 shows the results of both the segmentation and the labelling. The first observation is that spatial consistency is present and that no discontinuities can be observed. This is not trivial since the SAR image (plus texture) has been segmented as a series of egg code regions with different number of classes: of the 42 regions, 10 had a single class and were not segmented but directly assigned a class, one region had two classes, 13 regions had three classes, 14 had four classes and four had five classes. Despite this stratification, the segmented image appears as one continuum and, apart from the land areas, no break lines can be observed.

The second observation is that, unlike land use maps, the ice map is usually made of relatively small bundles of pixels having the same class. This is rather normal considering the broken nature of sea ice (especially in the Gulf of St.-Lawrence) and the spatial resolution of 100 meters. It should also be noted that the image reflects the end of the winter season when the ice situation is usually the most complex. It was however also noted that that the number of classes has an effect on the segmentation results and that the number of iterations of the segmentation process should take this information into account. The effect of that problem can be sensed in the egg code regions with four or five classes which tend to be more speckled than the two- or three-classes regions. The relation is obviously not linear and further study is needed to establish the nature of the relation and its quantification.

4.2. Evaluation of class labelling

The results of the labelling are evaluated using ground truth from the 2003 validation campaign of CIS and quantified in a series of confusion matrices (Tables 1-5) showing the performance of the different classification metrics. The labelling algorithm is based on what we have called a “cognitive reasoning” algorithm that uses evidence from the egg codes to deduce the labels of the segments inside the egg code regions. In the present case, first-degree evidence came from the 10 one-class regions all of which were “fast ice”. It then turned out that third-degree evidence (two regions with a unique common class) was not possible so that the algorithm could not solve the remaining 32 regions based only on the information on a single class. The cognitive reasoning algorithm needs little information but it needs to be sufficiently diversified. To increase the quantity of evidence, a single three-class egg code region was manually solved to provide sufficient evidence to solve the rest of the regions in two passes.

Errors in the labelling results can either be attributed to the classification, the labelling or both. To separate these three kind of errors, one needs extensive ground truth which is not the case in this study. The following paragraph describe the results obtained with the five classification metrics without specification of the source of errors. It should be kept in mind that misclassification is two-fold and that their combined effects are reflected in these tables.

Mahanalobis distance: The MD is a popular classifier in the remote sensing community. It measures the probabilistic distance between and single vector and a sample through its mean and covariance matrix. In our particular case it is not the most appropriate because it can not measure the distance between two populations: the samples and the vectors composing the segments. To compensate this, the mean of the population is used. Despite this drawback, the MD classifier has proven valuable in previous experiments (Maillard and Clausi, 2005b). Table 1 shows

the confusion matrix obtained after labelling using the MD classifier. Although the overall results are rather low (53.88%), they are comparable with the other classifiers. It should be noted that much of the confusion is between grey-white ice and first-year ice, two ice classes that can be considered alike in terms of tone. It is also notable that various ice types were confused with open water. These errors might be attributable to a time lag between the ground truth and image acquisition during which the ice has shifted; this is further discussed the conclusions section.

Table 1: Confusion matrix for the MD classifier
N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference	Classified Data						
Data	N	G	GW	FY	F	OW	**Total**
N	907	1	40	0	0	294	1242
G	0	408	0	0	0	0	408
GW	342	2	313	0	0	625	1282
FY	117	118	593	224	0	28	1080
F	45	0	0	0	354	0	399
OW	0	0	0	0	0	370	370
Total	1411	529	946	224	354	1317	4781

Percent correct = 53.88% Kappa statistic = 0.4157

Fisher criterion: Unlike MD, the FC classifier is capable of taking into consideration the spread of both populations being compared. Coincidentally, the FC yielded the exact same results as the MD. This is not such an extraordinary coincidence since labelling is applied to whole segments and not single pixels.

Table 2: Confusion matrix for the FC classifier
N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference	Classified Data						
Data	N	G	GW	FY	F	OW	**Total**
N	907	1	40	0	0	294	1242
G	0	408	0	0	0	0	408
GW	342	2	313	0	0	625	1282
FY	117	118	593	224	0	28	1080
F	45	0	0	0	354	0	399
OW	0	0	0	0	0	370	370
Total	1411	529	946	224	354	1317	4781

Percent correct = 53.88% Kappa statistic = 0.4157

Chi-square: The $\chi 2$ "goodness-of-fit" test is usually employed to compare a sample's distribution to a theoretical distribution such as the Gaussian function. In the present case, it is used to compare two independent sample distributions through their shape. With 59.07% of overall classification success, the χ^2 yielded the highest score of the five classifiers. Here, all the open water pixels have been misclassified.

Table 3: Confusion matrix for the χ^2 classifier
N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference	Classified Data						
Data	N	G	GW	FY	F	OW	**Total**
N	907	39	262	0	0	34	1242
G	0	408	0	0	0	0	408
GW	342	2	931	0	0	7	1282
FY	117	118	72	224	0	549	1080
F	45	0	0	0	354	0	399
OW	0	305	65	0	0	0	370
Total	1411	872	1330	224	354	590	4781

Percent correct = 59.07% Kappa statistic = 0.4814

Kolmogorov-Smirnov: The KS test can be used as a classifier in the same manner as the χ^2 as it also measure the difference between two distributions but in their cumulative version. This is probably why it produced a similar result than the χ^2 with 58.96% overall success. It succeeded slightly better in classifying open water but at the expense of grey-white ice (being like water relatively dark).

Table 4: Confusion matrix for the KS classifier
N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference	Classified Data						
Data	N	G	GW	FY	F	OW	**Total**
N	889	1	274	34	0	44	1242
G	0	408	0	0	0	0	408
GW	625	2	342	271	0	42	1282
FY	0	118	369	521	0	72	1080
F	45	0	0	0	354	0	399
OW	65	0	0	0	0	305	370
Total	1624	529	985	826	354	463	4781

Percent correct = 58.96% Kappa statistic = 0.4801

Student's t-test: The t-test verifies if the difference between two sample's mean is significant to a certain degree of confidence or if the difference can be attributed to the sample's pooled variance. The t-test assumes a Gaussian distribution but this assumption can usually be relaxed if N is large. The t-test yielded the lowest score (55%) of the five metrics. This is probably due to the fact that it assumes a Gaussian distribution and that it considers the mean as fully representative of the samples.

Table 5: Confusion matrix for the t classifier
N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference	Classified Data						
Data	N	G	GW	FY	F	OW	**Total**
N	907	1	6	34	0	294	1242
G	0	408	0	0	0	0	408
GW	342	2	42	7	0	889	1282
FY	117	118	296	549	0	0	1080
F	45	0	0	0	354	0	399
OW	0	0	0	0	0	370	370
Total	1411	529	344	590	354	1553	4781

Percent correct = 55.01% Kappa statistic = 0.4300

4.3. Voting principle

The voting principle adopted here only considered the final labels of each classification metric and the label with the most votes was considered a "winner". In cases where no definite winner was found (*e.g.* two votes for label 1, two for label 2 and one for label 3) the segment was left unclassified. This occurred in 17 of the 127 segments (or 13.4%) and this explains why the total number of samples in the confusion matrix (Table 6) is smaller than in the other matrices. It can be observed form Table 6 that the total score of $\sim 61\%$ is higher than any of the five classification metrics. This confirms the fact that different metric can be better adapted to certain situations and that different such situations do occur in classifying sea ice from SAR data.

To compare the individual success of the five metrics between themselves, Table 7 shows the percentage of each of the five classification metrics matches the final label of the voting principle. The most striking observation is the relatively small number of matches between the voting principle solution and the χ^2 metric. This is somewhat troubling since the χ^2 yielded the highest score of the five metrics. The MD and FC classifiers appear to be very consistent with the previous results and show a very similar behavior.

Finally, we have tested the significance (at 90% level of confidence) of the difference between the six confusion matrices and found that only the t-test results and the voting proinciple results

Table 6: Confusion matrix for the voting principle (multi-classifier).

N=nilas, G=grey, GW=grey-white, FY=first year, F=fast, OW=open water

Reference Data	Classified Data N	G	GW	FY	F	OW	**Total**
N	907	1	6	0	0	294	1208
G	0	408	0	0	0	0	408
GW	342	2	306	0	0	625	1275
FY	117	118	72	224	0	28	559
F	45	0	0	0	354	0	399
OW	0	0	0	0	0	370	370
Total	1411	529	384	224	354	1317	4219

Percent correct = 60.89% Kappa statistic = 0.5006

Table 7: Number of matches between the voting principle and the five classification metrics. The one-label regions (which were not segmented) were withdrawn to compute the corrected values.

Metric	total matches	corrected matches
MD	81.10%	79.49%
FC	84.25%	82.91%
χ^2	37.80%	32.48%
KS	52.76%	48.72%
t	59.84%	56.41%

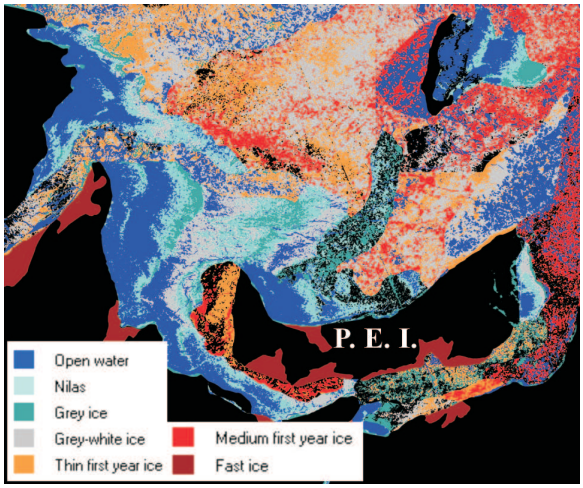


Figure 3: Classified image-map of the study area using the voting principle.

can be considered significantly different so that no real conclusion can be drawn from the different scores.

5. CONCLUSIONS

This article presents the performance of the MAGSIC system using ground truth data collected *in situ* by CIS during a validation campaign. It also serves the purpose of testing and comparing five different classification metrics for the labelling phase as well as a voting scheme using all five metrics. Results show that, for different reasons including over-generalization in the egg-code maps and possible shifts of the ice condition between ground data and image acquisition, the percentage of success hardly surpasses the 60% mark. They also suggest that the χ^2 and the Kolmogorov-Smirnov classifiers appear to perform slightly better (maybe because they compare the whole distributions) but their superiority could not be confirmend statistically. The multi-classifier approach using a simple voting scheme performed at least as well as the best classifier and could be used as an alternative to choosing a single classifier.

ACKNOWLEDGMENTS

The authors are thankful to the Canadian Ice Service (CIS), - GEOIDE (http://www.geoide.ulaval.ca) and CRYSYS (http://www.crysys.ca) for their their financial support. As well, CIS is thanked for providing the SAR sea ice images (all RADARSAT images ©CSA - http://www.space/gc/ca).

REFERENCES

Agnew, T., R. Brown, G. Flato, H. Melling, and B. Ramsay (1999). Canadian contributions to GCOS - sea ice. *Atmospheric Environment Service (AES) - Environment Canada, Toronto, Ontario*.

Deng, H. and D. A. Clausi (2005). Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field models. *IEEE Trans. on Geoscience and Remote Sensing 43*(3), 528–538.

Hall, D. (1998). *Manual of remote sensing* (3rd ed.), Volume 2, Chapter Remote sensing of snow and ice using imaging radar, pp. 677–703. New York, NY: John Wiley and Sons.

Haralick, R. M., K. Shanmugam, and I. Dinstein (1973). Textural features for image classification. *IEEE Trans. Sys. Man Cybern 3*, 610–621.

Karvonen, J. A. (2004). Baltic sea ice sar segmentation and classification using modified pulse-coupled neural networks. *IEEE Trans. on Geoscience and Remote Sensing 42*(7), 1566–1574.

Maillard, P. and D. A. Clausi (2005a, May 9-11). Comparing classification metrics for labeling segmented remote sensing images. In *Proceedings of 2nd Canadian Conference on Computer and Robot Vision, Victoria, BC, Canada*.

Maillard, P. and D. A. Clausi (2005b). Operational map-guided classification of sar sea ice imagery. *IEEE Trans. on Geoscience and Remote Sensing 43*(12), 2940–2951.

Mäkynen, M., A. Manninen, M. Similä, J. Karvonen, and M. Hallikainen (2002). Incidence angle dependence of the statistical properties of C-band HH-polarization backscattering signatures of the baltic eea ice. *IEEE Trans. on Geoscience and Remote Sensing 40*(12), 2593–2605.

Raney, R. (1998). *Manual of Remote Sensing* (3rd ed.), Volume 2, Chapter Radar fundamentals: technical perspective, pp. 9–130. New York, NY: John Wiley and Sons.

Soh, L. K., C. Tsatsoulis, D. Gineris, and C. Bertoia (2004). ARKTOS: An intelligent system for sar sea ice classification. *IEEE Trans. on Geoscience and Remote Sensing 42*(1), 229–248.

Wolper, D. H. (1992). Stacked generalisation. *Neural Networks 5*, 241–259.

Xu, L., A. Krzyzak, A., and C. Y. Suen (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transaction on Systems, Man, and Cybernetics 22*(3), 418–435.