SAR SURFACE ICE COVER DISCRIMINATION USING DISTRIBUTION MATCHING

Rashpal S. Gill

Danish Meteorological Institute, Ice charting and Remote Sensing Division, Lyngbyvej 100, DK 2100, Copenhagen Ø, Denmark.Tel. (+45) 39157342, Fax. (+45) 39157300, email.: rsg@dmi.dk

ABSTRACT

Discrimination between open water and sea ice in SAR imagery can still pose a problem to the ice analysts during manual interpretation. To help them in this task, new algorithm have been tested which is based on the user first manually identifying a particular surface type in a SAR image (e.g., open water area or sea ice of particular concentration or ice type) then the program will automatically determine similar regions in the remainder of an image. The algorithm is based on matching the statistics of the known and unknown regions using either (a) Kolmogorov-Smirnov (KS), or (b) Chi-Square (CS) distribution matching test. The main advantage in using these distribution matching tests is that the knowledge of the probability distribution functions (pdf) of the regions are not needed. Both KS and CS tests determine whether the two data sets belong to the same or different, yet undetermined, distributions. The main difference between KS and CS tests is that they are valid for un-binned and binned data respectively.

The KS and CS were tested on the amplitude SAR image and the image products: (a) Power-to-Mean Ratio (PMR), and (b) Gamma-pdf which are computed from it. Both PMR and Gamma-pdf are useful tools for discriminating between open water and sea ice type in SAR images. The results presented in this article shows that the KS test is efficient (both reliable and computationally fast) at identifying similar surface types. It performed best with the amplitude data and Gamma-pdf while results using the PMR images were more prone to ambiguities. CS test did not perform as well as the KS test. This is because the data first has to be arbitrarily binned which results in some information being inevitably lost. It was also found to be many times slower to run on the computer. For these reasons it was decided not to use the CS test for matching known and unknown regions in a SAR image.

The information obtained using the KS tests can be considered as the 'best statistical guess' during situations when the ice analysts have difficulty in interpreting parts of a SAR image.

Keyword: Sea ice, RADARSAT, distribution matching, Kolmogorov-Smirnov test, Chi-Square test, Greenland.

1. INTRODUCTION

The Danish Meteorological Institute (DMI) is responsible for the operational charting of the sea ice in the waters around Greenland for the safety of ship navigation using primarily RADARSAT ScanSAR Wide (RSAT-SCW) data since autumn 1998. To provide fast and easy to interpret additional tools/products to the ice analysts that would help them to analysis SAR images during routine operations, a number of algorithms based on the first and second order statistical parameters, probability distributions, wavelet transform and constant false alarm rate have been tested at DMI in the last 7 - 8 years [1], [2]. A suitable method to estimate the performance of these products was found to be to display their values on a computer screen (after appropriate scaling) and use manual interpretations. This method was preferred as it was found to be more accurate than the traditional methods which usually involve making graphical or scatter plots of the parameter for different ice types or concentration. Further, using this approach it was possible to make the use of ice analyst's image interpretation skills. By displaying these grey tone 'images' of the parameters along side the original image it was possible for the ice analysts to evaluate the performance (and the limits) of the various parameters over different regions of sea ice and open water.

Results from the investigations on the first order statistics have already been reported [1]. These investigations showed that all first order parameters were ambiguous for extremely 'noisy' images (usually those containing very complex surface wind patterns). Nevertheless, it was found that the grey tone 'images' of the normalised second moment of the probability distribution, the Power-to-Mean-Ratio (PMR), given by the simple expression

$$PMR = \frac{< I^2 >}{< I >^2} \qquad (1)$$

where I is the intensity value, were in most cases very useful to discriminate between the different regions of ice and water and for determining the positions of possible icebergs. Currently, PMR grey images are routinely used by the ice analysts during ice charting operations as supplement to the original contrast enhanced RADARSAT images.

The grey tone 'images' of the Gamma probability distribution for open water - sea ice discrimination was also investigated [2]. It was found that these grey tone images also contain useful information for sea ice and open water discrimination and supplement the PMR 'images'. Gamma probability distribution was found to be especially useful for detecting calm water regions. This distribution for the average intensity, $\bar{I}$, over m pixels is given by

$$P_\gamma(\bar{I}) = \left(\frac{mL}{\mu_B}\right)^{mL} \frac{\bar{I}^{(mL-1)}}{\Gamma(mL)} \exp\frac{-mL\bar{I}}{\mu_B} \qquad (2)$$

where L is the number of looks (L=7 for ScanSAR Wide image product), Γ is the Gamma function and μ_B is the mean intensity of the background [2].

This article essentially consists of three sections. In the next section the method used to evaluate the KS and CS tests is presented. The results from the evaluations of the two tests are presented in section 3. Finally, section 4 concerns with discussion and conclusions. Full details of the results presented in this article are given in [3].

2. EVALUATION METHOD

The algorithms reported in here are given in the "Numerical Recipes in C by Press et. al., Second Edition, University of Cambridge Press, pages 620 – 627, and thus will not be described in here. As mentioned above the main difference between KS and CS tests is that they are valid for binned and un-binned data, respectively. What this means is that in the KS case there is no restriction on the type of data sets to be compared; they can be integers or floats. However this is not the case with CS test, here data must be binned which means some sort of arbitrary histograms of the data must first be created. This inevitably results in loss of information. Thus the binning of the data sets is critically important for the CS test. In particular, for the amplitude averaged products this is not important as the amplitude values are naturally binned and are in the range 0.0 – 255.0. However, in the case of PMR and Gamma-pdf it is not so simple. PMR values are typically in the range 1.0 – 3.0, while Gamma-pdf probabilities lies between 0.0 – 1.0. Thus given the narrow ranges of these parameters, comparing just their integer values from the different surface types is inadequate (as then most of the information is lost) and it is essential to use their decimal values. In particular, the following amplifications to the PMR and Gamma-pdf (GAM) values are made when binning in the CS algorithm: $PMR_{binned} = 10^6(PMR_{original} - 1.0)$ and $GAM_{binned} = 10^5(GAM_{original})$. From these two expressions it is clear that PMR and GAM values are binned into 10^6 and 10^5 different bins, respectively. These multiplications factors were obtained after extensive trials.

The above three float products are generated from the original ScanSAR Wide 8-bit amplitude data files. In particular, the AMPLITUDE product is obtained by averaging window of size 4 × 4 pixels, and fixing the distance between two of these consecutive windows at 4 pixels in both directions. Similarly, the GAMMA product is obtained by using Eq. 2, above, the size of computation window and inter window spacing is same as for the AMPLITUDE product. Finally, the PMR product is obtained by using Eq. 1, for window size of 20 × 20 pixels, with inter spacing of two consecutive windows again fixed at 4 pixels in both directions. The averaging of windows is necessary to compute the parameters used in Eqs.

1 and 2 and it helps to reduce the background speckle noise. The following computational procedure is used to test the KS and CS methods:

1. Training areas of different surface types (e.g., calm and turbulent water, sea ice of low and high concentration, multi- and first- year sea ice, and if appropriate sea ice in different stages of development, etc.) are generated by displaying the AMPLITUDE product on a computer terminal. In particular, regions of size ~ 50 × 50 pixels of different surface types are manually identified in the AMPLITUDE image. In Fig. 1 only two regions have been identified: open water (green box) and sea ice (red box). In practice there is no limit to the number of different regions of open water or sea ice of different concentration or type that can be manually identified and hence used for matching tests.
2. The mask file generated above is then used to identify and store in a buffer the data points pertaining to different surface types from each of the three products: AMPLITUDE, GAMMA and PMR. For example, these are the points within the colour boxes shown in Figs. 1 (left side).
3. Then KS and CS tests are used to match the data points in each of the three products to their respective data points from the different surface types to determine the probabilities. To compute these probabilities a test window of size 4 × 4 pixels, which was determined by carrying out trials with windows of different sizes and was found to be both computationally optimal and statistically sufficient, is slid across the three products.
4. The probabilities computed above are then linearly amplified so that they could be displayed on a computer terminal for manual interpretation.

3. RESULTS

This consists of, for the purpose of illustrating the method, a classification of the RSAT-SCW image from Disco Bay off West Greenland from 2000-05-24 shown in fig. 1 (top left). The figure also shows the GAMMA and PMR products computed form the SAR image on a grey tone scale (figs. 1b, 1c). As can be seen from this figure, the GAMMA product is efficient at identifying regions of calm water, including those within the ice pack, and fast ice [2]. Similarly, the PMR product is good at detecting sea ice boundaries, icebergs and also at delineating the structure, such as ice ridges, within ice floes. In the figure typical region of sea ice and open water (the red and blue boxes respectively) are manually identified for distribution matching tests. The data points from three different products within these (and only these !) boxes, shown as histogram plots in Fig.1 (right side), are stored in a buffer and are used in the distribution matching tests as described in the last section. As can be seen from each pair curves shown in Fig. 1, they appear distinctively different for sea ice and open water. The KS and CS techniques are thus exploiting this information to locate regions with similar data distributions.

In fig. 2 the probabilities computed using the KS test from the products AMPLITUDE (top row), GAMMA (second row) and PMR (third row), amplified by a factor of 10^9, $10^5$and 10^{10} , respectively, for both sea ice and open water are shown on a grey tone scale. From fig. 2a it can be seen from the high grey values that the amplitude pixels from the ice regions matches well with the manually located (those within the red box), thus indicating that they belong to the same distribution. On the other hand, low grey values in the figure indicates that the amplitude pixels from these regions (mainly in the open water and land regions) does not belong to the same distribution as sea ice pixels found in the red box. Similarly, fig. 2b shows the results from the KS test for determining how well the amplitude image pixels matches to those from the water region indicated by the blue box. Here high probabilities of a distribution match are obtained for part of the water region. Low values are obtained for not only the sea ice and land regions in the image, but more significantly, also for some part of the water region. From this it can be concluded that the water pixels from the latter regions does not belong to those water pixels within the blue box. One possible interpretation of this could be that the water in these regions have a different ‘backscatter state’ which physically could be turbulent water compared to the relatively calm water found within the blue box.

The pair of figures (2c, 2d) and (2e, 2f), middle and bottom rows, respectively, shows the KS probabilities computed from the GAMMA and PMR products shown in fig. 1. A comparison of these figures with the pair (2a, 2b), discussed above shows that they are similar to each other thus indicating that both the GAMMA and PMR products are also efficient at separating between the 2 surface classes. It is important to note that the fast ice at middle top of the image (marked on fig. 1a) is not accurately classified in figs. 2a, 2c, 2e and 2f. This implies either that the 2 classes (red and blue box) chosen to represent a typical sea ice and water classes cannot be used to characterise this particular type of sea ice or that the KS test is insensitive to discriminate between these classes. Nevertheless, for more accurate surface type discrimination more surface types must be used in the distribution matching tests.

Figs. pair 2g, 2h shows the CS probabilities computed using the AMPLITUDE product. Comparing these probabilities with their counterpart computed using the KS test (figs. 2a, 2b), discussed above, shows that the latter are on the whole are accurate.

From the example presented in fig. 2 it is clear that KS distribution matching test is quite successful at locating similar surface types. However, this was not generally found to be true when using the CS test. Even in the cases when the CS test performed well it did not appear to contain any more information that already was not available from the KS test. The reason for the poor performance of the CS test are clear and have been outlined above. Furthermore, the CS algorithm was found to be 4 - 5 times slower than the KS algorithm, which is a very important parameter in any operational environment such as the national sea ice charting services. For these reasons it was decided not to use the CS test in any further evaluation of this technique.

4. DISCUSSION AND CONCLUSIONS

In this article the technique of matching the data from an unknown region with that of previously identified areas were explored to determine if useful information could be obtained about the former. Two type of data matching tests were examined; (a) Kolmogorov-Smirnov and (b) Chi-Square. The first test, KS, works on the original floating point data while the CS test is used with binned data. The data sets used in the tests were the floating point AMPLITUDE, GAMMA and PMR products.

The results of the KS and CS tests were presented as 'images' which are nothing more than displaying their respective probabilities on a grey tone scale. This method of presenting the results was chosen as it lends easily to manual interpretation. The evaluation of the results shows that both KS and CS distribution matching techniques contain some very useful information at matching the unknown surface regions to known surface types, for example sea ice and open water. In particular, this type of information is very useful during operational ice charting where even the most experienced ice analyst can sometime be in doubt in interpreting part of a SAR image. The information available from the KS and CS tests can thus be considered as a 'best statistical guess'.

Concerning the relative performance of the KS and CS tests, it was found that KS test was far more superior than the CS test at locating surfaces of same types. In terms of computer execution times KS test was found to be 4 - 5 times faster than the CS test. For these reasons it was concluded not to use the CS test for further evaluations. However, it must be pointed out that the Kolmogorov-Smirnov test has its own limitations. In particular, KS test is most sensitive around the median of the cumulative distribution function, P(x), it is less sensitive at the tail ends where P(x) is approaching 0 or 1. In the C-recipe book, modified version of the KS test are proposed, these were tested by author and results were not too different from the original when displayed on a grey tone scale. Further, the KS test cannot discriminate between all types of distributions, such as a distribution with 2 maximums.

In the subsequent articles it will be reported how the information obtained from these products using the KS test is used in the fuzzy logic rules called Multi Experts – Multi Criteria Decision Making (ME-MCDM) to classify a SAR image both semi-automatically and fully automatically i.e., without the need for any operator supervision or input.

Acknowledgements. The support of the European Union DG-XII Climate Environment Programme under the CONVECTION project (contract EVK2-CT-2000-00058) is acknowledged.

5. REFERENCES

1. Gill R. S. and Valeur H. H. "Ice cover discrimination in the Greenland waters using first-order texture parameters of ERS SAR images", *Int. J. Remote Sensing*, vol. 20, no. 2, pp. 373 – 385, 1999.

2. Gill R. S., "Sea Ice Edge and Icebergs Detection using routine operations". *Canadian J. REMOTE SENSING*, special issue on Sea Ice and Icebergs, vol. 27, no. 5, pp 411 – 432, 2001.

3. Gill R. S., "SAR surface classification using distribution matching", *DANISH MET. Scientific report*, 02 -08, 2002.

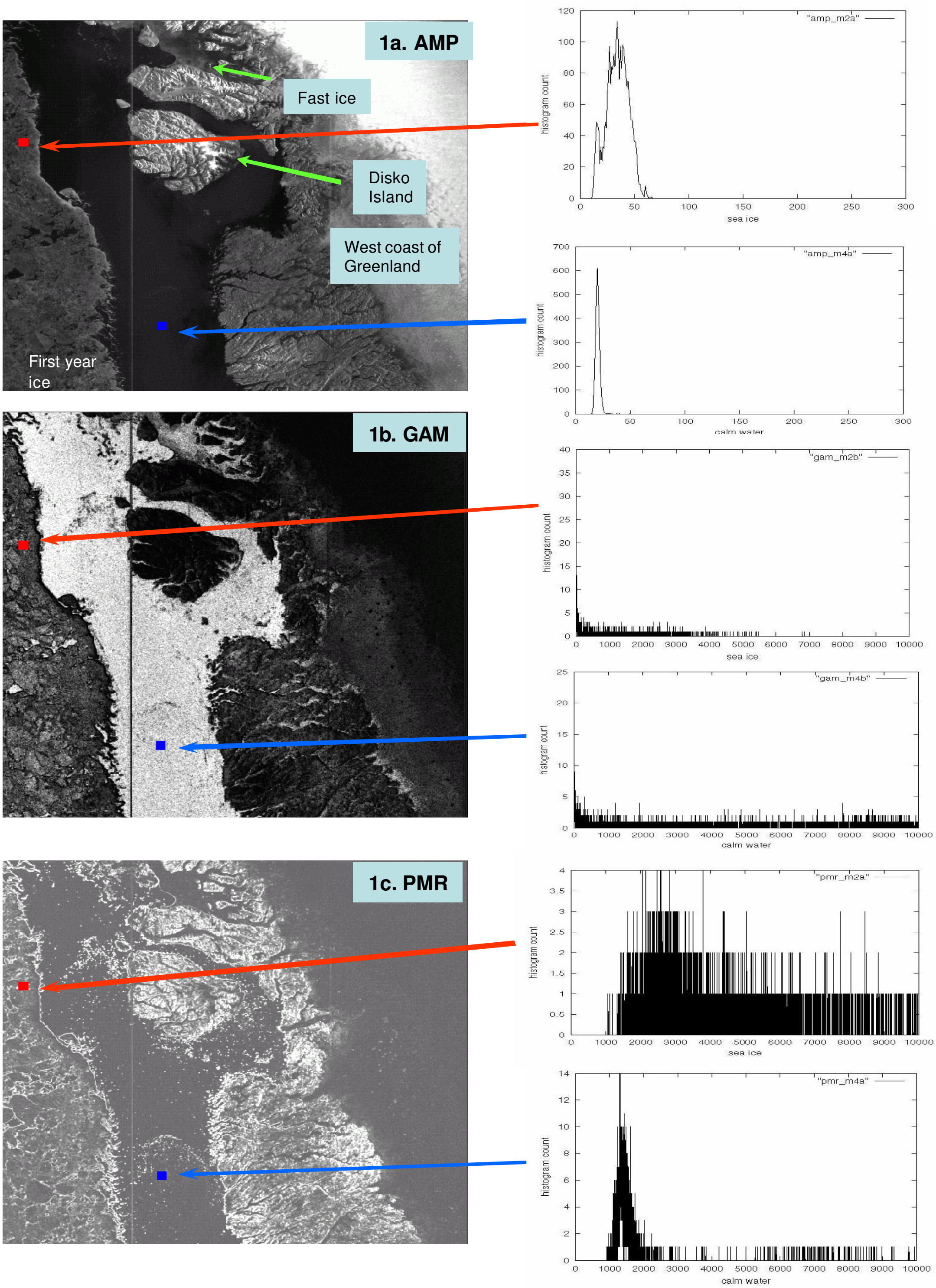

Fig. 1 Shows the AMPLITUDE, GAMMA-pdf and PMR (figs. 1a – 1c, respectively) image products and their binned data values of typical sea ice and open water, in the red and blue boxes, respectively

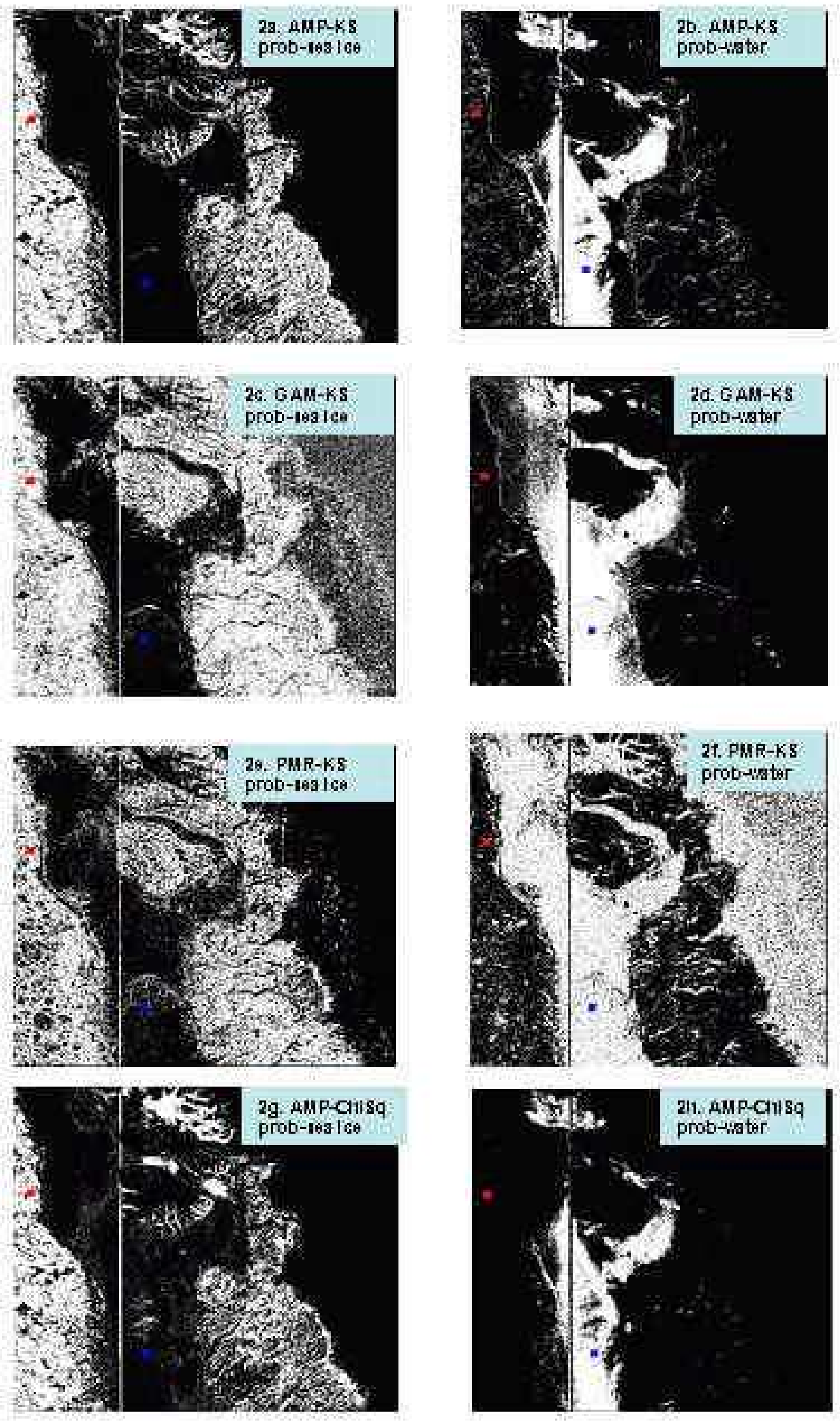

Fig. 2 The Kolmogorov – Smirnov probabilities (figs. 2a – 2f) and Chi Square probabilities (figs. 2g, 2h) computed by matching the AMPLITUDE (figs. 2a, 2b, 2g, 2h), GAMMA-pdf (figs. 2c and 2d) and PMR (figs. 2e, 2f) values for sea ice in the red box and water in the blue box with the rest of the pixels in the respective 'image' products shown on a grey tone scale. High grey tone values indicate high probabilities and vice-versa.