NEURAL-NETWORK BASED STATELESS ICE DETECTION IN ERS SCATTEROMETER DATA

Xavier Neyt, Pauline Pettiaux, Nicolas Manise and Marc Acheroy

Royal Military Academy, 30 av de la Renaissance, B-1000 Brussels, Belgium

ABSTRACT

A new scheme to perform stateless ice / sea discrimination in ERS scatterometer data is proposed. This method consists in combining several methods proposed in the literature using a Bayesian framework. Each of the combined method is first reviewed in a consistent framework. In particular, the ice/sea probability according to each individual criterion is extracted using a neural network. The proposed method is shown to provide acceptable results even without taking into account historic data, i.e. without performing temporal averaging.

Key words: Scatterometer, ERS, ice detection, neural network.

1. INTRODUCTION

The data acquired by the scatterometer on-board the ERS spacecraft can be used to determine sea-surface wind speed provided the data was acquired over open water. Wind information is extracted from the σ^0 triplets σ_f^0, σ_m^0 and σ_a^0 measured in 3 different look directions. The model used to compute the wind speed and direction from the σ^0 triplets is only valid over open water. Wind extraction performed using σ^0 triplets measured over ice results in aberrant wind speed and directions which in turn disturb the wind ambiguity removal process. As part of the upgrade of the ground-processing of the ERS-2 scatterometer data (Ref. 1, 2), it was deemed necessary to perform a real-time ice detection in order to remove bogus wind vectors. Moreover, due to operational constraints, a stateless approach is required. This imply that the temporal coherence of the ice mask cannot be considered to increase the classification accuracy.

While it is relatively easy to discriminate land from sea trough the use of maps, discriminating between sea-ice and open water is more challenging. Several methods have been proposed to discriminate between ice and open water. These methods are reviewed in Section 2. In Section 3, the neural-network framework that will be used to compare these methods is presented. In this section, the conditions under which the different methods are discriminant are exposed under the form of an ice probability according to the neural network output. Finally, we propose to combine the different methods together in a Bayesian framework in order to increase classification accuracy and discuss the performance of the overall scheme.

2. REVIEW OF EXISTING SEA/ICE DISCRIMINATION CRITERIA

2.1. Introduction

As can be seen in figures 1 and 2, the distribution of the σ^0 measurements over sea is different than that of measurements over ice. This difference in distribution makes it, in principle, possible to separate the measurements in two classes: sea and ice.

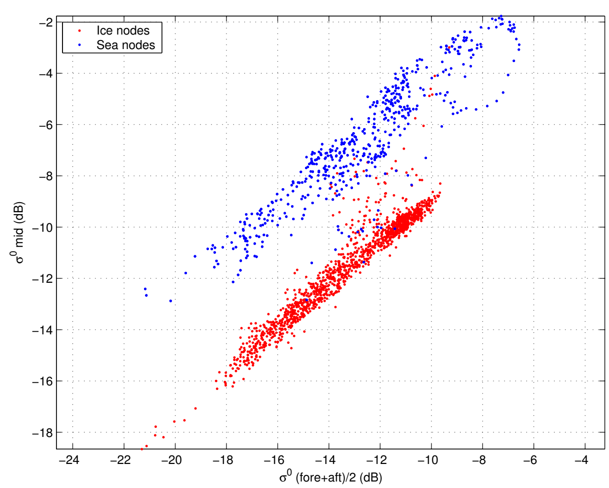

Figure 1. Scatter plot of the σ^0 over ice and over sea for node 5 (low incidence angle). Projection in the $\sigma_f^0 = \sigma_a^0$ plane.

However, the σ^0 triplets corresponding to ice and sea are not always as well separated. As the incidence angle increases, both classes get closer to each other.

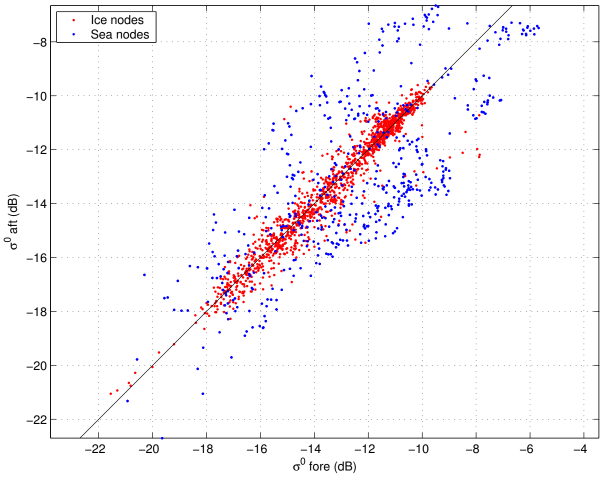

Figure 2. Scatter plot of the σ^0 over ice and over sea for node 5 (low incidence angle). Projection in the $\sigma_f^0 = \sigma_a^0$ plane.

Actually, around mid-swath, the ice class is totally inside the sea class. This is illustrated in figures 3 and 4.

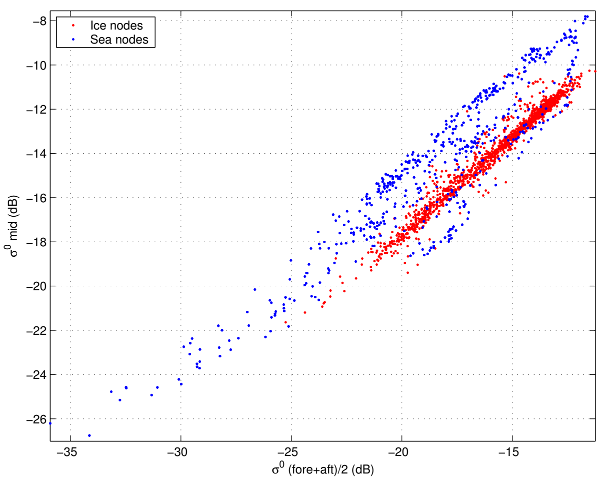

Figure 3. Scatter plot of the σ^0 over ice and over sea for node 15 (high incidence angle). Projection in the $\sigma_f^0 = \sigma_a^0$ plane.

Due to the complex shape of the two classes, classification based on the raw σ^0 triplets would be quite complex. Hence other classification metrics have been proposed in the literature. These will be reviewed below.

The graphs in this section correspond to data acquired during two passes over the North pole on December 30th 1999. The classification is based on the IFREMER ice mask, considered as reference data. Misclassification is obvious for some data points, as can for instance be seen in figure 1.

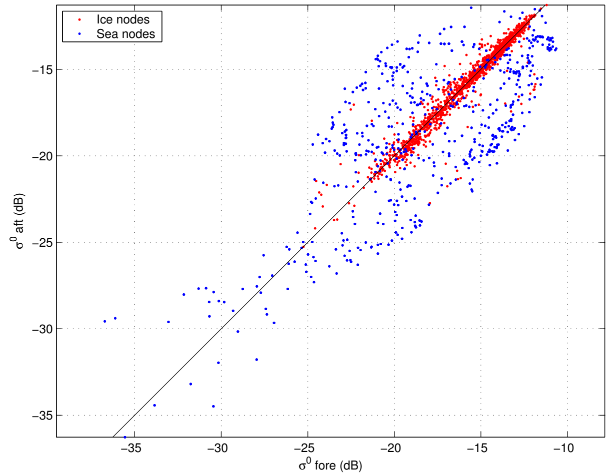

Figure 4. Scatter plot of the σ^0 over ice and over sea for node 15 (high incidence angle). Projection in the $\sigma_m^0 = 0$ plane.

2.2. Isotropy

In (Ref. 3, 4), a measure of the isotropy of the backscattering is proposed as discriminating criterion. The isotropy factor is defined as

$$\mathcal{A} = \left| \frac{\sigma_f - \sigma_a}{\sigma_f + \sigma_a} \right| \tag{1}$$

where the σ are provided in dB.

Ice is supposed to be isotropic as far as EM backscattering is concerned and hence typically has a low isotropy factor. On the other hand, open sea typically results in a high isotropy factor. However, low isotropy values can be observed over sea. This happens when the wind is blowing parallel or perpendicular to the ground track. Indeed, in this case, the capillary waves generated by the wind are propagating perpendicular or parallel to the satellite ground track hence the fore and aft measurements will have similar values.

Figure 5 illustrates the distribution of the isotropy values for the sea and the ice classes in function of the incidence angle. Clearly, the isotropy values corresponding to ice are typically small. However small isotropy factor over open sea are indeed possible too. As written above, this occurs when the wind is blowing parallel or perpendicular to the satellite ground track.

2.3. Derivative of backscatter in function of the incidence angle

In (Ref. 5), the derivative of the σ^0 w.r.t. the incidence angle was proposed as discriminating criterion. Since the fore-/aft- and mid-beam measurements are made at different incidence angles, the derivative can

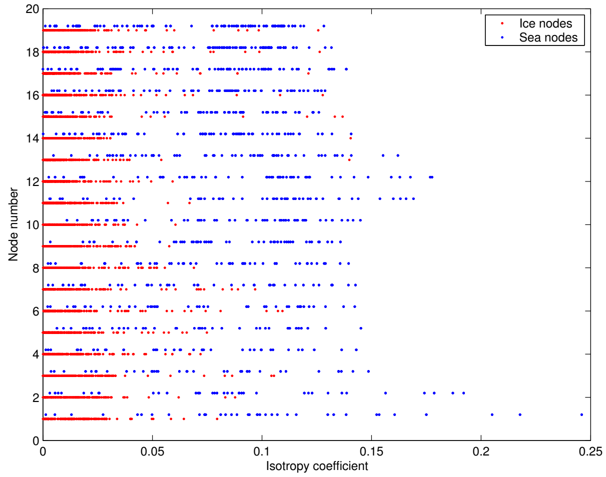

Figure 5. Scatter plot of the isotropy factor illustrating the distribution of the sea and ice classes.

be approximated by

$$\mathcal{D} = -\frac{(\sigma_f + \sigma_a)/2 - \sigma_m}{(\theta_f + \theta_a)/2 - \theta_m} \qquad (2)$$

where the σ^0 values are in dB and the incidence angle values in degrees. Notice that, due to the use of the zero-gyro mode following the gyroscopes anomaly, the incidence angles of the fore and of the aft beam cannot be assumed equal anymore (Ref. 1) hence the modification of the original formula given in (Ref. 5). When the incidence angle is not too high, σ^0 measurements over ice typically result in a lower $\mathcal{D}$ than measurements over sea.

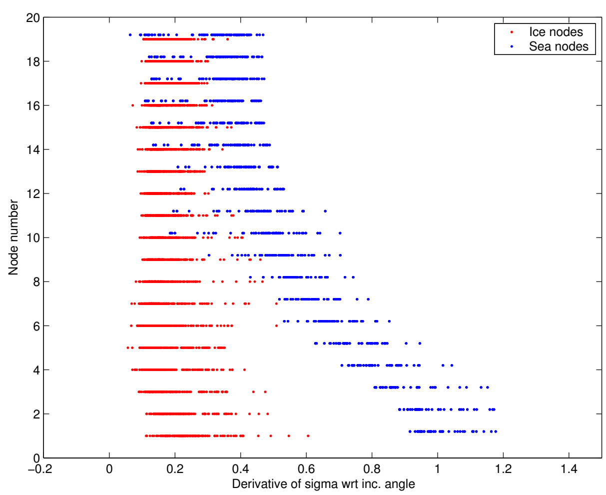

Figure 6. Scatter plot of the derivative of sigma, illustrating the distribution of the sea and ice classes.

Figure 5 illustrates the distribution of the derivative of sigma between sea and ice nodes in function of the incidence angle. The figure tend to indicate that the criterion will be more discriminant at low incidence angles (low node number, near swath).

2.4. Distance to wind model

The (Euclidean) distance to the wind model was also proposed (Ref. 6) as a criterion. Indeed, σ^0 measurements acquired over sea should lie close to the wind model, while measurements performed over ice might be located further away from the wind model. However, since the σ^0 measurements over ice are actually very close to the wind model at mid swath, this criterion will not be very discriminating at mid swath.

This is illustrated in figure 7.

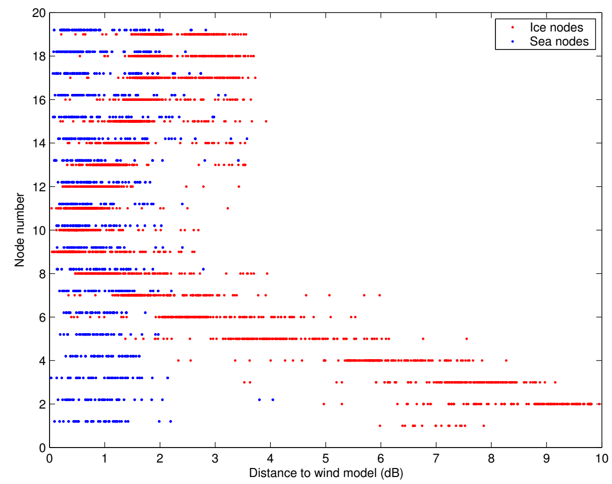

Figure 7. Scatter plot of the distance to the wind model, illustrating the distribution of the sea and ice classes.

2.5. Distance to ice model

Several models of the σ^0 over ice where proposed (Ref. 7, 8). Similarly as what was done in the previous section, the idea is to use the euclidean distance to the ice-model as discriminating criterion. The model selected in this article is the ERS ice model described in (Ref. 8). This model can however be generalized to other types of scatterometers. The model consists in an incidence angle-dependent line. For the details of the model, the reader is referred to (Ref. 8).

As can be seen in figure 8, this metric is essentially discriminant at low incidence angles. Indeed, for higher incidence angles, the distribution of ice and sea nodes tends to be closer to each other as can be seen in figures 1 and 3.

3. SEA ICE DISCRIMINATION

3.1. Introduction

In order to perform a classification between open sea or ice, the criteria described in the previous section

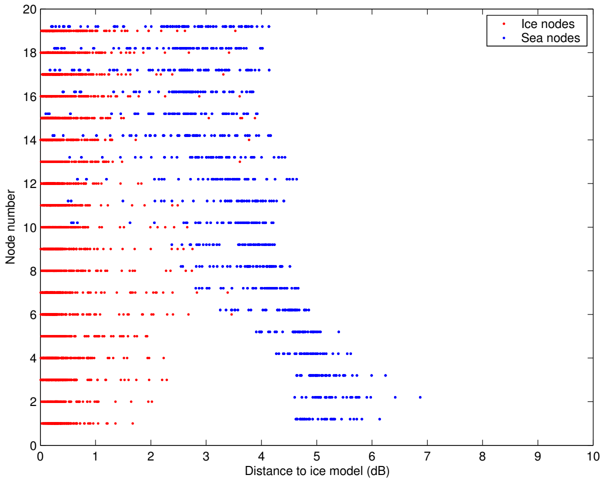

Figure 8. Scatter plot of the distance to the ice model, illustrating the distribution of the sea and ice nodes.

are typically thresholded (Ref. 3, 8). The threshold used is also typically dependent on the incidence angle. However, a binary decision does not take into account the fact that, in ambiguous cases, it is not possible to make a decision based only on one single σ^0-triplet measurement. In (Ref. 8), a decision algorithm using 4 classes (sea, ice, mixed and not valid) is proposed. This algorithm is based on an incidence-angle dependent thresholding of the distances to the ice model and to the wind model. Measurements close to the ice model and far from the wind model are classified as ice, measurements close to the wind model but far from the ice model are classified as sea and measurements close to both the ice model and the sea model are classified as mixed. The existence of the mixed class acknowledges the fact that some nodes cannot clearly be classified as ice or sea.

These decision methods provide a binary (quaternary) answer regarding the status of the considered measurement point (sea or ice). No information is provided on the (un)certainty of the classification and the classification accuracy trade off is difficult to master. In the following section, we define an ice probability based on the criteria defined in the previous section. This ice probability can then be thresholded to perform a classification. Furthermore, the ice probability computed from the different criteria presented above can be combined together, as individual experts to provide a combined ice probability.

3.2. A neural-network based classification

Our goal is thus to compute the ice probability given the measurements, $P(H_1|m_{c,i})$, where H_1 is the hypothesis "the measurement i corresponds to ice", $m_{c,i} = (C_{c,i}, n_i)$ is the measurement vector composed of the numerical value $C_{c,i}$ of the criterion c for measurement i and of the across-track node number n_i at which measurement i was made.

It is well known (Ref. 9) that this probability can be learned by a Multi-Layer Perceptron (MLP). The learning process consists in feeding the MLP with the measurements made $m_{c,i}$ and imposing as desired output 1 if H_1 is true for measurement i and 0 else. In order to avoid biasing the MLP output due to a differing a priori probability we must ensure $P(H_1) = P(\overline{H_1})$ over the training set. This imply that the number of measurements corresponding to ice in the training sets must be equal to the number of measurements corresponding to sea.

We considered an MLP with two inputs, one for the value of the criterion $C_{c,i}$ and the other for the node number n_i. It has one single hidden layer counting 5 neurons. The number of hidden layers and the number of neurones in these layers govern the complexity of the non-linear function that the MLP will be able to approximate. We are actually seeking to approximate reasonably "simple" functions, hence the single hidden layer and the small number of neurons in that hidden layer.

The results[1] of the learning of the probability $P(H_1|m_{c,i})$ by the MLP for the different criteria are shown in figures 9 to 12. For clarity, the 3D-surface was thresholded. Blue corresponds to $P < 20\%$ (open sea with 80% probability), red to $P > 80\%$ (ice with 80% probability) and green to values in between (mixed or unknown).

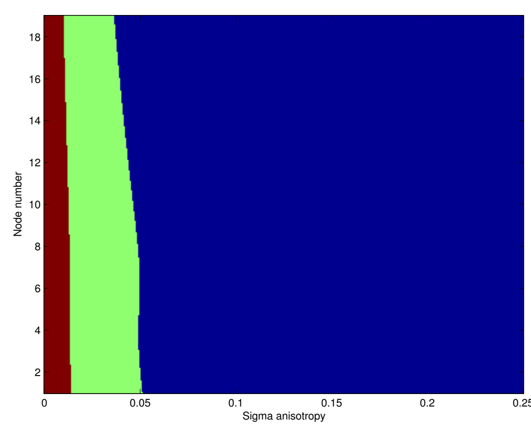

Figure 9. Thresholded probability $P(H_1|m_{c,i})$ for the isotropy criterion.

[1] learning was performed on data acquired during one day, on December 30th 1999.

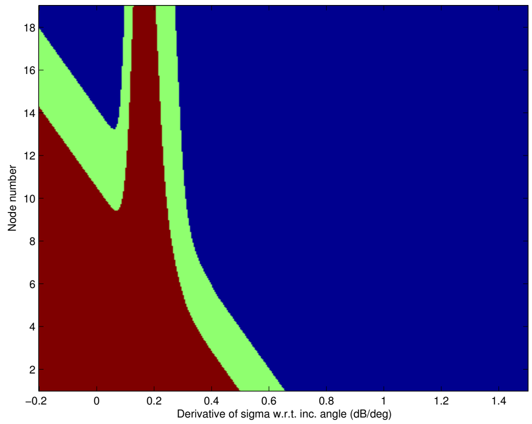

Figure 10. Thresholded probability $P(H_1|m_{c,i})$ for the Derivative of sigma criterion.

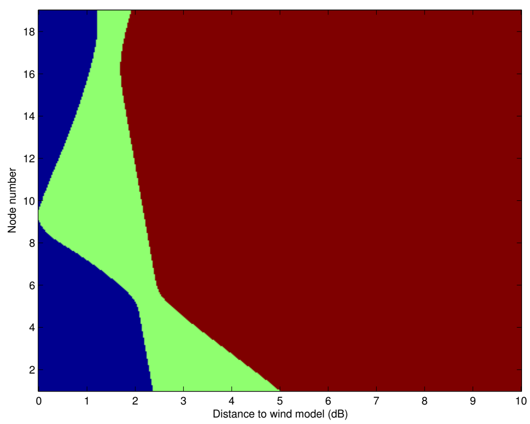

Figure 11. Thresholded probability $P(H_1|m_{c,i})$ for the Distance to wind model criterion.

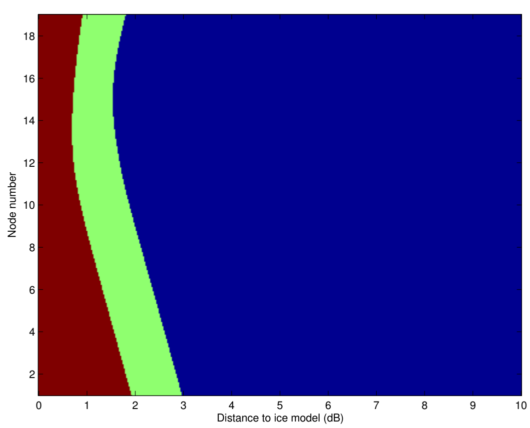

Figure 12. Thresholded probability $P(H_1|m_{c,i})$ for the Distance to ice model criterion.

As can be seen from these figures, the uncertain areas for the Isotropy and for the Distance to wind model criteria are quite large. In itself a large uncertain area is not bad, as long as no measurement samples fall inside that area. As can be seen by comparing figures 9 to 12 with the figures from Section 2, there are actually a lot of samples that fall inside that uncertain area for the Isotropy and for the Distance to the wind model criteria. Consequently, these methods exhibit a high rate of "No Decision" answers.

3.3. Performance comparison

The output of the MLP provides the probability that a given input corresponds to ice (H_1). By thresholding the output probability, a decision can be taken regarding the class to which the provided input belong. For a given threshold, it is possible to compute the True Ice rate (measurements classified as ice and actually corresponding to ice), False Sea rate (measurements classified as ice, but actually corresponding to sea) and unclassified ice (measurement not classified although it was actually ice). Figure 13 shows the ROC for ice, where the independent variable is the threshold used in taking the decision. As can be seen, the "Distance to ice model" criterion is the most discriminant, closely followed by the "Derivative of sigma". The two other criteria would imply a much lower True Ice rate if a low False sea rate was to be achieved. Figure 14 shows the unclassified

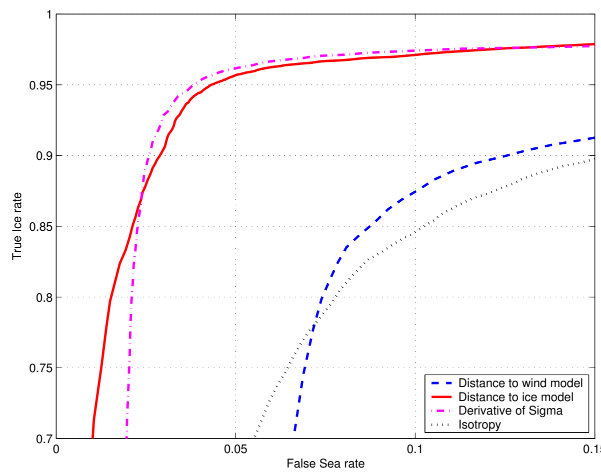

Figure 13. Comparison between the different criteria: ROC curve for the 4 criteria.

ice rate in function of the across track node-number (which corresponds approximatively to an incidence angle). As can be seen, the "Distance to wind model" criterion fails to discriminate between ice and sea at mid swath. This is due to the fact that at those incidence angles, σ^0 corresponding to ice are also very close to the wind model. Also, the σ^0 "Isotropy" criterion does not prove to be very decisive at any incidence angle. This confirms the deductions made in Section 2.

4. COMBINATION

When several sources of information are available, they can be combined to reduce imprecision and uncertainty and increase completeness(Ref. 10). A good review of some of the existing combination

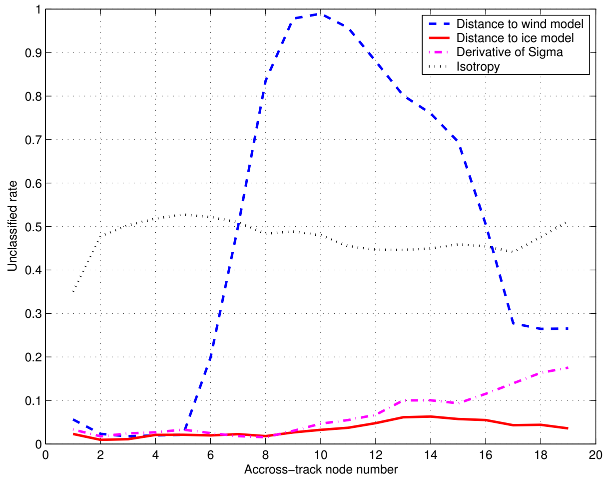

Figure 14. Comparison between the different criteria: "No Decision" rate for a False Sea rate of 3%.

methods is to be found in (Ref. 10). It only makes sense to combine the "best" sources of information, hence we will combine the "Distance to ice model" and the "Derivative of sigma" criteria. We will respectively refer to these criteria as c_1 and c_2.

The neural networks described in the previous section provide the posterior probability $P(H_1|m_{c,i})$. However, we would like to compute the Ice probability given measurements made by two criteria $P(H_1|m_{c_1,i}, m_{c_2,i})$. We will follow a development similar as in (Ref. 11), where it was performed for classification in several classes. One has

$$P(H_1|m_{c_1,i}, m_{c_2,i}) = \frac{P(m_{c_1,i}, m_{c_2,i}|H_1)P(H_1)}{P(m_{c_1,i}, m_{c_2,i})} \quad (3)$$

or since

$$\begin{aligned} P(m_{c_1,i}, m_{c_2,i}) = \quad & P(m_{c_1,i}, m_{c_2,i}|H_1)P(H_1) \\ & + P(m_{c_1,i}, m_{c_2,i}|\overline{H_1})P(\overline{H_1}) \end{aligned}$$

one can rewrite (3) as

$$P(H_1|m_{c_1,i}, m_{c_2,i}) = \frac{1}{1+e^{-a}} \quad (4)$$

in which we recognize the sigmoidal activation function of the considered neurons and where

$$a = \ln\left(\frac{P(m_{c_1,i}, m_{c_2,i}|H_1)P(H_1)}{P(m_{c_1,i}, m_{c_2,i}|\overline{H_1})P(\overline{H_1})}\right). \quad (5)$$

This last expression can further be decomposed, assuming independence of $m_{c_1,i}$ and $m_{c_2,i}$,

$$\begin{aligned} a = \quad & \ln\left(\frac{P(m_{c_1,i}|H_1)}{P(m_{c_1,i}|\overline{H_1})}\right) + \ln\left(\frac{P(m_{c_2,i}|H_1)}{P(m_{c_2,i}|\overline{H_1})}\right) \\ & + \ln\left(\frac{P(H_1)}{P(\overline{H_1})}\right). \end{aligned} \quad (6)$$

On the other hand, the neural networks described in the previous section provide

$$P(H_1|m_{c,i}) = \frac{1}{1+e^{-a_c}} \quad (7)$$

where

$$a_c = \ln\left(\frac{P(m_{c,i}|H_1)}{P(m_{c,i}|\overline{H_1})}\right) + \ln\left(\frac{P(H_1)}{P(\overline{H_1})}\right). \quad (8)$$

where c denotes the criterion considered. Since the training set is balanced, one has $P(H_1) = P(\overline{H_1})$ and the last term in (6) and (8) vanishes. Combining (8) in (6), one obtains

$$a = a_{c_1} + a_{c_2}. \quad (9)$$

This means that the posterior probability conditioned on two measurements $P(H_1|m_{c_1,i}, m_{c_2,i})$ can be obtained using the existing neural networks each being trained on one criterion. The combined posterior probability is obtained by adding the output of the existing neural network *before* the sigmoidal activation function of the output neuron and by applying a sigmoidal activation function on the resulting sum.

Interestingly enough, from (7), one has

$$-a_c = \ln\left(\frac{P(\overline{H_1}|m_{c,i})}{P(H_1|m_{c,i})}\right) \quad (10)$$

and if we combine this equation with (9) and (4), we obtain

$$P(H_1|m_{c_1,i}, m_{c_2,i}) = \frac{P(H_1|m_{c_1,i})P(H_1|m_{c_2,i})}{P(H_1|m_{c_1,i})P(H_1|m_{c_2,i}) + P(\overline{H_1}|m_{c_1,i})P(\overline{H_1}|m_{c_2,i})}. \quad (11)$$

If we denote $P_1 = P(H_1|m_{c_1,i})$ and $P_2 = P(H_1|m_{c_2,i})$, this last equation can be rewritten as

$$s(P_1, P_2) = \frac{P_1 P_2}{P_1 P_2 + (1-P_1)(1-P_2)} \quad (12)$$

which is precisely the expression of a particular symmetric associative sum operator well known in the fusion theory (Ref. 10). Independently of probability considerations, this operator has a behavior that depends on the input data P_1 and P_2. It will tend to a compromise if the inputs do not agree. On the other hand, if the inputs do agree, it will reinforce the agreement.

We will compare this combination method to he mean operator. The mean operator simply outputs the mean of its inputs, i.e. $s(P_1, P_2) = (P_1 + P_2)/2$, where P_1 and P_2 have the same definitions as above. It is obvious that this operator always makes a compromise between its inputs.

Figure 15 compares the performance of the two selected criteria ("Distance to ice model" and "Derivative of sigma") with those obtained after combination of these two criteria using each of the two fusion operators considered. The ROC curves of the combinations are above that of the best single criterion, which means that the True Ice rate will be higher for the same False Sea rate. There is hardly any difference between the two fusion operators considered. This might be explained by the fact that the

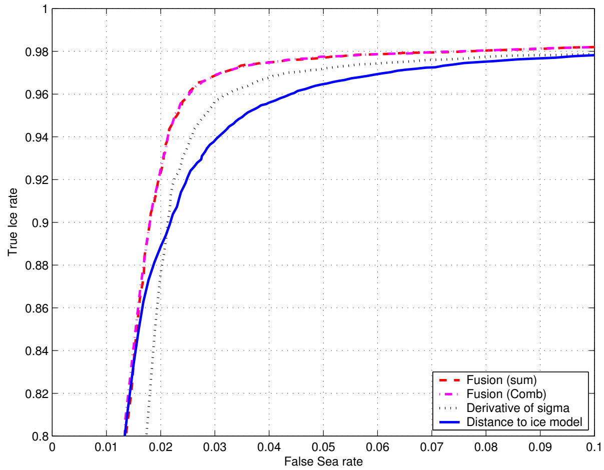

Figure 15. Comparison of the performance of the fusion: ROC curve for a False Sea rate of 3% (right).

input data are not totally independent with as consequences that the Symmetric associative sum under performs.

Figure 16 compares the "No Decision" rate of the same decision methods in function of the across-track node number. The combination has the effect of reducing the "No Decision" rate but for large node numbers. This is particularly true for the Symmetrical associative sum operator. The lower performance at large node numbers is due to the fact that the performance of the "Derivative of sigma" criterion has a lower decisiveness for large node numbers.

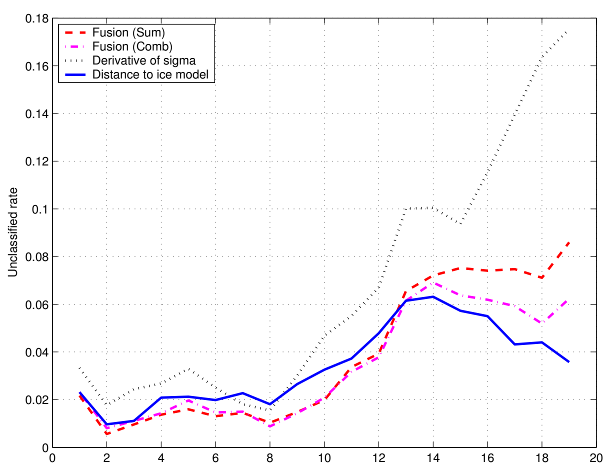

Figure 16. Comparison of the performance of the fusion: and "No Decision" rate for a False Sea rate of 3%.

Table 1 shows the performance figures for the two individual criteria and the proposed combination methods.

	True Ice	Unkn. Ice
Deriv. of σ	95.6%	1.4%
Dist. to ice model	93.8%	3.3%
Mean	96.8%	0.2%
Sym. Ass. Sum	96.9%	0.2%

Table 1. Performance comparison at a False Sea rate of 3%.

5. CONCLUSIONS

We reviewed the existing sea/ice discriminating criteria found in a literature by comparing them using a Neural-network framework. The advantage of the Neural-network framework is that sea/ice decision is seen as a thresholding of a conditional ice probability with clear trade offs. The results of the comparison clearly show the limits of each criterion.

The use of the conditional ice probability further makes it straightforward to perform a fusion of the Neural-network output. The obtained results (both in terms of classification accuracy as in terms of decisiveness) are enhanced by the fusion. These performances where obtained without considering state information, i.e. without relying on temporal coherence of the ice mask.

6. ACKNOWLEDGMENTS

This work was performed under European Space Agency (ESA) contracts. We would like to thank the ESA for the use of data and the provision of industry-confidential information. This work would not have been possible without the help of Pascal Lecomte and Raffaele Crapolicchio from the European Space Agency/ESRIN.

REFERENCES

1. Neyt, X., Pettiaux, P., and Acheroy, M. Scatterometer ground processing review for gyroless operations. In *Proceedings of SPIE: Remote Sensing of the Ocean and Sea Ice 2002*, volume 4880, Crete, Greece, September 2002.
2. Crapolicchio, R., Lecomte, P., and Neyt, X. The advanced scatterometer processing system for ERS data: Design, products and performances. In *Proceedings of the Envisat Symposium*, Salzburg, Austria, September 2004.
3. Lecomte, P. Wind scatterometer processing requirements for ice detection. Technical Report ER-SA-ESA-SY-1122, ESA ESRIN, September 1993.
4. Cavanié, A., Gohin, F., Quilfen, Y., and Lecomte, P. Identification of sea ice zones

using the AMI-wind: Physical bases and applications to the FDP and CERSAT processing chains. In *Proc. Second ERS-1 Validation Workshop*, pages 1009–1012, Hamburg, Germany, October 1993. ESTEC.

5. Gohin, F. and Cavanié, A. Some active and passive microwave signatures of antarctic sea ice from mid-winter to spring 1991. *International Journal of Remote Sensing*, 16(11):2031–2054, 1995.

6. Gohin, F. and Maroni, C. ERS scatterometer polar sea ice grids: User manual. Technical Report C2-MUT-W-03-IF, IFREMER, March 1998.

7. Cavanié, A. An empirical C-band backscatter model over arctic sea ice from ERS-1 AMI-wind data. In *Proc. of a Joint ESA-Eumetsat Workshop on Emerging Scatterometer Applications*, pages 99–106, Noordwijk, The Netherlands, November 1995. ESTEC.

8. de Haan, S. and Stoffelen, A. Ice discrimination using ERS scatterometer. Technical report, KNMI, September 2001.

9. Bishop, C. *Neural Networks for Pattern Recognition.* Carendon Press, 1995.

10. Bloch, I. Information combination operators for data fusion: A comparative review with classification. *IEEE Transactions on Systems, Man and Cybernetics*, 26(1):52–67, January 1996.

11. Druyts, P. and Acheroy, M. A modular multi layer perceptron (MMLP) to identify objects composed of characteristic sub-parts. In *Proc. of the '97 Artifical Neural Networks in Engineering conference.* University of Missouri - Rolla, November 1997.