The RFI 2022 Workshop at ECMWF, 14 - 18 February 2022
2022 URSI International Union of Radio Science

# INTRODUCING MACHINE LEARNING IN THE GROUND RFI DETECTION SYSTEM (GRDS)

R. Onrubia(1), and R. Oliva (1)
(1) Zenithal Blue Technologies, Barcelona, Spain, https://zenithalblue.com/

## Abstract

Radio Frequency Interference is becoming a global threat to Earth Observation missions, particularly in passive microwave remote sensing. The Ground RFI Detection System is a novel concept for Radio Frequency Interference detection in remote sensing products, which combines multiple techniques in order to maximize the RFI detection. The software makes use of external information such as a land/sea mask so set the proper thresholds for the detection techniques. Initial results from this technique applied to SMOS products showed that a lot of false alarms were raised when misclassification of the land/sea mask occurred due to presence of sea-ice. For this purpose, an artificial neural network has been developed to estimate sea-ice concentration from L1c SMOS products. The new results show an important improvement in the reduction of false alarm rates when detecting RFI.

***Index Terms***— Sea ice detection, SMOS, artificial neural networks, passive microwave radiometry, Radio Frequency Interference

## 1 Introduction

Radio Frequency Interference (RFI) is a worldwide growing threat particularly affecting microwave passive Earth Observation (EO) missions. Since these sensors estimate the natural electromagnetic emission from ground, any signal over the noise floor is susceptible to disrupt the measurements. For instance, RFI contaminated data would lead Numerical Weather Prediction (NWP) systems to weather forecasting errors. The contaminated data must be flagged, and if possible mitigated, in order to avoid erroneous scientific results.

Ground RFI Detection System (GRDS) is a novel concept for RFI detection and mitigation being developed by Zenithal Blue Technologies. The software is currently able to ingest products from the Soil Moisture and Ocean Salinity (SMOS) and the Advanced Microwave Scanner Radiometer 2 (AMSR2) missions at different product levels. Then it applies multiple RFI detection techniques, such as maximum intensity, outlier detection or maximum spatial variability, with adjustable threshold levels. These thresholds can be defined as function of multiple parameters, e.g., the ground type, the latitude, and the incidence angle. The thresholds are also adjusted depending on the proximity to previously known RFI sources. The original ingested EO products are modified in order to add the RFI contaminated flags. Finally, the ground pixels flagged as contaminated are added to a database, which can be used to monitor the RFI scenario in thar particular band.

The system has been tested with SMOS products in collaboration with the European Center for Mid-Range Weather Forecast (ECMWF). One month of data was processed by the GRDS system and then fed to ECMWF integrated forecasting system. The Community Microwave Emissivity Modelling platform (CMEM) is used to convert measured temperature, atmospheric pressure, humidity, etc. into expected brightness temperatures. These values are then subtracted to the measured ones obtaining thus the First Guess Departures (FGD) [1]. The statistics of these FGD, in particular the standard deviation, are studied to assess the quality of the SMOS data, and in this case the performance of the GRDS system. The presence of RFI signals increase the standard deviation of the FGDs. The results of this campaign showed promising results as shown in Figure 1~~Figure 1~~. Therein it is compared the standard deviation of the FGD of the SMOS data without any RFI screening (top), with the RFI screening implemented in the original SMOS data product flags (middle), and with the independent RFI screening carried out by GRDS (bottom).

Figure 2~~Figure 2~~ shows the accumulated RFI probability for SMOS YY polarization during July 2019 in ascending passes during the validation campaign with ECMWF. A problem that was detected at the time was a false detection of sea ice pixels as RFI as seen as yellow fringes near the poles in Figure 2~~Figure 2~~. The utilized SMOS L1C data does not provide information on whether sea ice is present or not, therefore the RFI detection techniques apply the same thresholds for sea and sea ice pixels. This causes the system to mistakenly consider the excess of brightness temperature from sea ice pixels with respect to the expected sea pixels as RFI contamination. To solve this issue, a sea ice detection system based on Artificial Neural Networks (ANN) has been implemented in GRDS to detect and classify sea and sea ice pixels in order to apply the proper RFI detection thresholds.

## 2 Methodology

SMOS measures the Earth's brightness temperature in snapshots with the shape shown in Figure 3~~Figure 3~~ a) [2]. Feeding a neural network with SMOS data is not straightforward since many difficulties are found: the ground points have different number of data observations, even between polarization of the same ground point, so data has to be regularized.

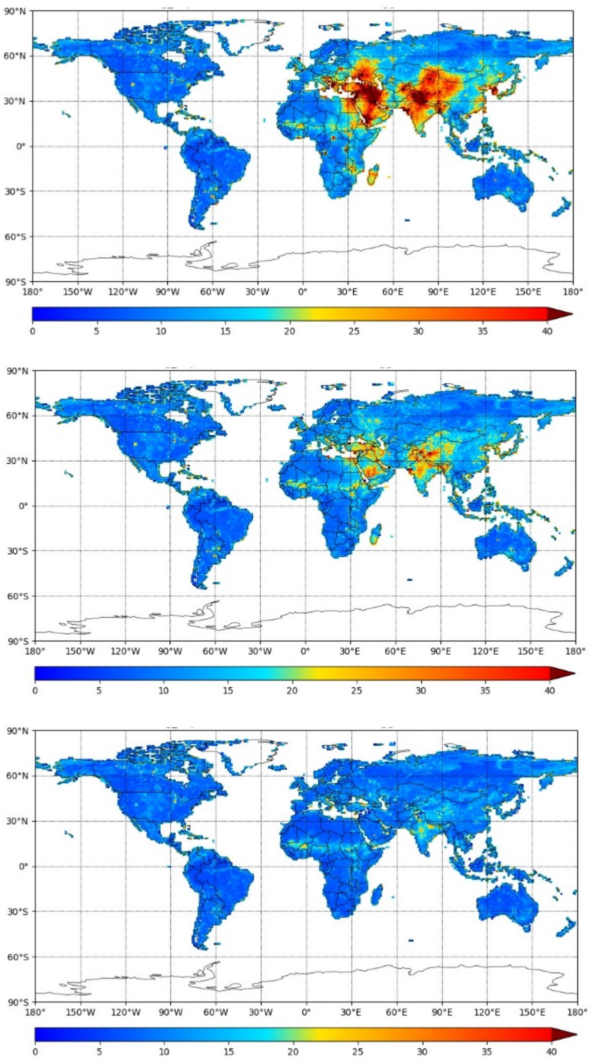


**Figure 1.** Standard deviation of the First Guess Departure with no RFI screening (top), the current RFI screening method implemented in SMOS (middle), and the GRDS screening method (bottom).

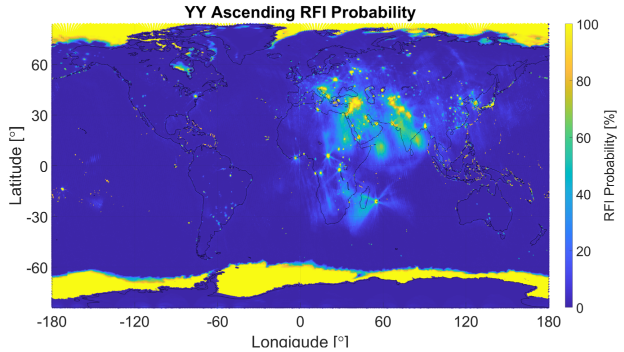
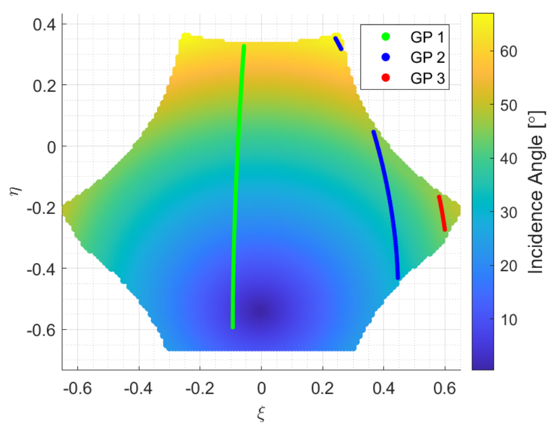
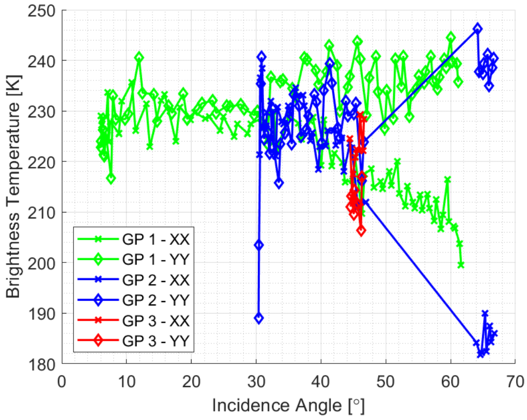



**Figure 2.** RFI probability for YY polarization and ascending passes for July 2019.

The snapshots are taken consecutively and are highly overlapped. Thus, any ground point is observed multiple times with different incidence angles. XX and YY polarization snapshots are alternately taken, which means that a given ground point is not observed simultaneously with the same incidence angle at both polarizations. Also, the observations' incidence angles are irregular (not quantized), not even for the same ground point for different orbits. The track along time of three ground points is shown in Figure 3~~Figure 3~~ a), and the observed brightness temperatures at XX and YY polarizations are shown in Figure 3~~Figure 3~~ b). Observe that the ground points that are farther from the center have a smaller number of observations within that one orbit. Note also that the second ground point (GP 2 in Figure 3~~Figure 3~~) has a gap between 45-65º of incidence angle, and is observed with higher incidence angles than the other two ground points. That means that not all ground points are observed in the full range of incidence angles, so the missing ones need to be filled. In order to regularize this, the authors opted to collocate and average the XX and YY brightness temperatures in incidence angle bins of 10º (0-10º, 10º-20º, …) for each polarization. If a bin has been filled, a valid data flag for that particular bin is raised. Then, the empty bins will be filled with the average brightness temperature of that particular polarization. In the extreme case that there is no information for one polarization, all bins will be set to the average brightness temperature of the other polarization.




(a)




(b)

**Figure 3.** a) Incidence angle in the SMOS Snapshot, with 3 ground points being observed by the instrument, and b) brightness temperature of the three ground points at XX and YY polarizations for one orbit.

Another problem to overcome is that the SMOS instrument has a bias which is present in all snapshots and is scene dependent [3]. Figure 4~~Figure 4~~ shows the estimated instrument bias over the Amazonian Forest for YY polarization. Note that the color scale has been limited to ±5 K for optimizing its visualization, but its peaks reach up to ±15 K. A ground pixel observed following the trajectory of the solid red arrow will have a positive bias of 10-15 K at the beginning of the trajectory, whereas a ground pixel following the dashed red line trajectory will have a bias with similar amplitude but opposite sign. As a consequence, stripes of false negatives were observed in the preliminary results. To remove these stripes, the biases had to be taken into account. To do so, the xi and eta coordinates of the observations were counted in bins of 0.1 discrete cosine units from -0.65 to 0.65 in the xi case, and -0.7 to 0.4 in the eta case and added as input features, thus allowing the neural network to "know" the trajectory of the ground point along the snapshot.

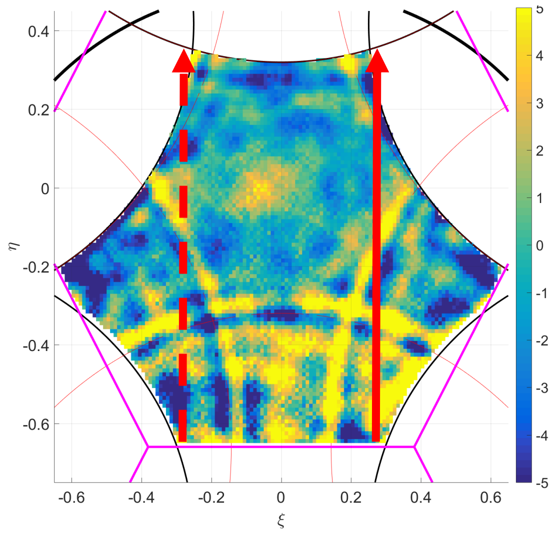



**Figure 4** SMOS Estimated bias for YY polarization over the Amazonian Forest. The color scale has been limited to ±5 K for optimizing its visualization.

Once all these difficulties have been overcome, the sea ice prediction neural network is fed with a total of 55 features: the collocated and averaged brightness temperatures (7 features per polarization), a vector indicating if the data is valid (7 features per polarization), the vectors indicating the xi-eta path of the observations (13 features for xi and 11 for eta), the water fraction (1 feature), and the total average brightness temperature (1 feature per polarization).

To speed up the neural network training, Principal Component Analysis (PCA) is used for dimension reduction. First, each feature is debiased and normalized by its standard deviation. Then, PCA is applied to the complete training set, and the number of dimensions used is chosen to preserve the 99% of the variance: the number of features is reduced from 55 down to 29.

The neural networks aim to first estimate the percentage of sea ice concentration (SIC), and then to catalogue the pixels as function of this estimation. For this purpose, the dataset has been chosen in order to maximize the number of datapoints with sea ice concentration higher than 5% (pure water) and lower than 95% (pure sea ice) according to sea ice coverage daily maps from AMSR2 [4]. AMSR2 sea-ice data products were chosen over SMOS for being available over both poles and because the product is more consolidated.

The first dataset originally consisted on six days of SMOS products during the thawing peak season, three in each hemisphere. The chosen dates are 06/01/15 to 08/01/15 for the austral summer, and 04/07/15, 09/07/15 and 10/07/15 for the boreal summer. Figure 5~~Figure 5~~ show large areas partially covered by sea ice on 06/01/15 and 09/07/15. The histogram of the ground truth sea ice concentration in Figure 6~~Figure 6~~ shows a high representation of sea and sea ice pixels. This dataset is separated in 3 sub-datasets: training (70%), test (15%), and validation (15%). A second dataset with similar characteristics but from different dates

(07/01/16 and 18/07/16) is generated only to validate the networks when some geophysical parameters change along time, such as surface roughness, salinity, air temperature, etc.

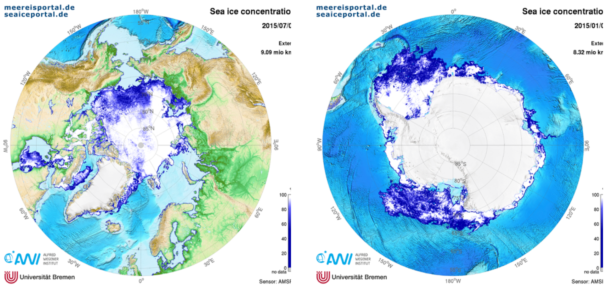



**Figure 5.** Sea ice coverage during 09/07/15 and 06/01/15. Note the dark blue areas indicating neither pure ice or water.

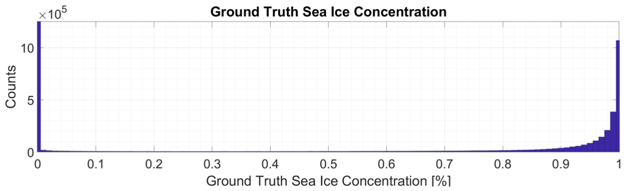



**Figure 6.** Histogram of the sea ice concentration from the first dataset based on AMSR2 ground truth. Note that the sea pixels (SIC = 0%) exceed the scale with a value of 139e5, an order of magnitude over the second highest represented value of the plot, pure sea ice pixels.

In order to determine the optimum size of the dataset, the performance of the ANN training, test, and validation datasets are studied for different dataset lengths by removing random entries of the total dataset. The performance is evaluated as the mean squared error between the sea ice coverage prediction and the sea ice coverage ground truth.

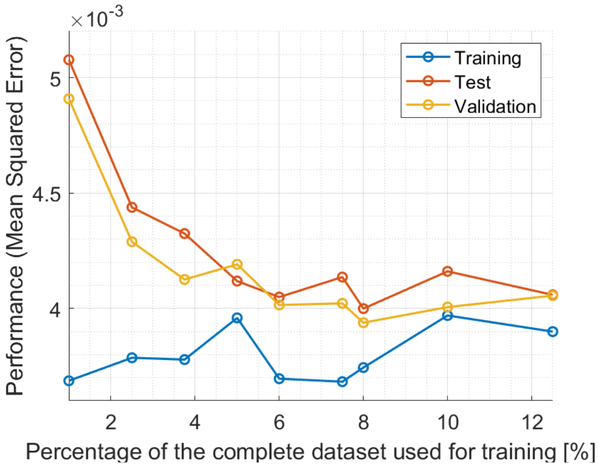



**Figure 7** Performance of the training, test and validation datasets as percentage of the original dataset size for the ANN.

Figure 7~~Figure 7~~ shows the performance of the training, test and validation sub-datasets for different dataset sizes (defined as the percentage of the original dataset). From the figure, using more than 10% of the original data does not seem to increase the system performance.

To evaluate the system performance, the predicted sea ice concentration is first cropped below 0 and over 1, and then is compared to AMSR2 [4], and SMOS L4 [5] SIC. The metrics used are the Root Mean Squared Error (RMSE) and the $R^2$ parameter. Then, a threshold determines if the pixel is catalogued as sea or sea ice. This value is set to 15%, the typical value used for measuring sea ice extent [4]. All pixels in between ±40º latitude are set to sea since no sea ice is expected at such low latitudes. The classified pixels are then compared to AMSR2 (SIC>15%), SMOS L3 [6] (Sea Ice Thickness (SIT) > 0), and SMOS L4 (SIC>15%). Note that SMOS sea ice products are only available for the northern hemisphere, so all data out of it will be discarded for the comparison, also in the case of AMSR2, in order to conduct the comparison in fair conditions. In the case of sea ice classification, the common precision, recall, accuracy and F1 score metrics will be evaluated [7]:

$$precision = \frac{tp}{tp + fp} \qquad (1).$$

$$recall = \frac{tp}{tp + fn} \qquad (2).$$

$$accuracy = \frac{tp + tn}{tp + fp + tn + fn} \qquad (3).$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \qquad (4).$$

Where $tp$, $fp$, $tn$, and $fn$ are true positives, false positives, true negatives and false negatives, respectively.

Finally, the RFI probability map obtained during the GRDS validation campaign with ECMWF in Figure 2~~Figure 2~~ is recomputed using the obtained neural network that re-classifies the thresholds used in RFI detection for Sea-Ice pixels.

## 3 Results

Figure 8~~Figure 8~~ shows a scatter plot between the predicted SIC and the estimated SIC from AMSR2, and the resulting metrics can be found in Table I~~Table I~~. Even though the data points in scatter plot are quite dispersed, the $R^2$ metric is high. This is caused by the high density of points in (0,0) and (1,1); that is, the error estimating the SIC of pure sea ice and pure sea pixels is low. The obtained $R^2$ are over 0.96, and the RMSE below 6.5% in the three sub-datasets (training, test, and validation) may also be caused by the high concentration of pure sea and pure sea ice pixels.

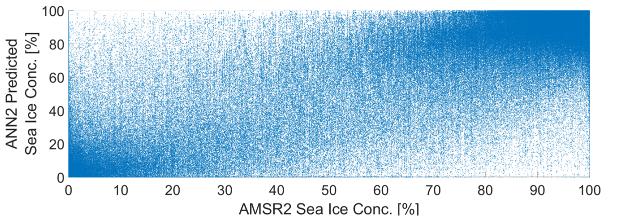



**Figure 8.** Scatter plot between the sea ice concetration predicted sea ice concentration and the AMSR2 data.

**Table I.** Obtained metrics after training the sea ice concentration prediction neural networks.

| Dataset | RMSE | $R^2$ |
|---|---|---|
| Training | 5.93% | 0.971 |
| Test | 6.05% | 0.97 |
| Validation | 6.01% | 0.97 |
| All | 5.96% | 0.971 |

When comparing the system performance on the second dataset to AMSR2 ground truth in Table II~~Table II~~, both the RMSE and the $R^2$ (6.1% compared to 6%, and 0.96 compared to 0.97, respectively) remain similar to the ones obtained in the first dataset. When using SMOS L4 as ground truth, the RMSE increases from 6% up to 9.8%, and the $R^2$ slightly decreases from 0.97 down to 0.95.

**Table II**. Obtained metrics for the sea ice concentration prediction using the second validation dataset.

| Ground truth | RMSE | $R^2$ |
|---|---|---|
| AMSR2 | 6.09% | 0.959 |
| SMOS L4 | 9.84% | 0.948 |

Then, the sea ice concentration estimation is compared to a threshold of 15% and compared to the ground truth of AMSR2 and SMOS L3 and L4 and is shown in Table III~~Table III~~. The SMOS precisions are higher compared with respect to AMSR2, but the Recall is lower. The resulting F1 score is higher with SMOS than with AMSR2. The accuracy is higher in the AMSR2 case, which is consistent with the fact that the neural networks has been trained using data from this sensor.

**Table III**. Obtained metrics for the sea ice classification prediction using the second validation dataset (Sea: SIC<15% or SIT=0; Sea Ice: SIC>15% or SIT>0)

| *Ground truth* | *Pre.* | *Rec.* | *Acc.* | *F1* |
|---|---|---|---|---|
| AMSR2 | 98.51% | 99.47% | 98.99% | 98.99% |
| SMOS L3 | 99.94% | 96.60% | 98.24% | 98.24% |
| SMOS L4 | 99.75% | 95.33% | 97.49% | 97.49% |

Figure 9~~Figure 9~~ shows the recomputed RFI probability map. The yellow stripes near the poles are not present anymore. The transition between sea and sea ice is subtle near the poles, but can be seen more easily in Hudson Bay and near the coast in the Labrador Sea. Some false detections are now seen below latitudes of -40º.

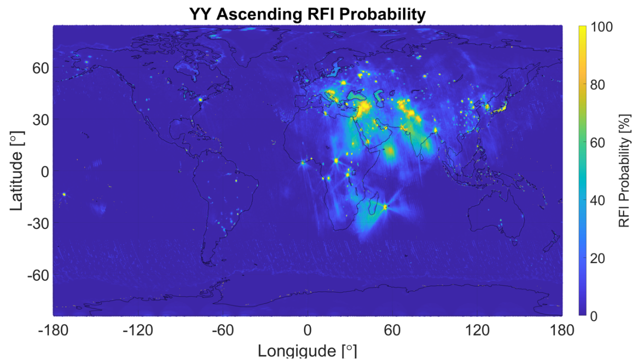



**Figure 9.** RFI probability for YY polarization and ascending passes for July 2019 after sea ice detection. Only sea pixels have been analyzed.

A second neural network was also studied. Instead of using the collocate brightness temperatures, the collocated brightness temperature polarization ratio $T_{pr}$ [8] were used:

$$T_{pr}[n] = \frac{T_{YY}[n] - T_{XX}[n]}{T_{YY}[n] + T_{XX}[n]} \quad (5).$$

Where $T_{XX}$ and $T_{YY}$ stand for the brightness temperature at XX and YY polarizations respectively, and $n$ is the index that refer to the bin number (n=1 refers to 0-10º, n=2 refers to 10-20º, …). If $T_{XX}$ or $T_{YY}$ are missing for a given incidence angle bin, the average value for the whole polarization is used. In this neural network the vector indicating if an incidence angle bin has valid data is computed as the AND logic operator between the valid flag of that particular bin for both polarizations; that is, $T_{pr}[n]$ is valid only if the $T_{XX}[n]$ and $T_{YY}[n]$ are valid. The results of this neural network are not shown since the results did not improve the presented ones.

## 4 Discussion

The scatter plot in Figure 8~~Figure 8~~ shows that $R^2$ metrics obtained in all the experiment might be artificially high due the high number of pure sea (SIC<5%) and pure sea ice (SIC>95%) pixels in the dataset. Nevertheless, the percentage of intermediate sea ice pixels (5%<SIC<95%) is low in the SMOS data: they can be observed in the thawing season, and still the percentage of intermediate sea ice pixels is small compared with respect the number of sea or sea ice pixels in those products. The obtained $R^2$ values are therefore evaluating properly the system performance taking into account how the data is distributed. Note that if a new sea ice concentration system is trained using a more uniform sea ice concentration distribution, the optimum number of samples to train the system shown in Figure 7~~Figure 7~~ may change.

The metrics obtained in Table II~~Table II~~ are slightly worse than the ones obtained during the training process as expected. The cause could be that the second dataset is from different days compared to the first one, and therefore many hidden variables might be playing a role in the results, such as the air temperature, sea surface temperature, the surface roughness caused by the wind speed, etc.

When comparing the metrics of AMSR2 with respect to SMOS L3 and L4 in Table III~~Table III~~ it can be observed that for a given neural network, the accuracy and F1 score decrease when comparing to AMSR2 than when comparing to SMOS, which is consistent with the fact that the neural networks have been trained with the former. Besides, the precision when comparing to SMOS increases but the recall decreases, which is caused by a decrease of false positives and an increase of false negatives. The main factor causing this might be the fact that AMSR2 sea ice concentration maps are generated with 89 GHz data, and SMOS radiometer works at 1.4 GHz: AMSR2 is able to observe thinner ice than SMOS, and also the pixel size is different.

Last, Figure 9~~Figure 9~~ shows a proper determination of sea ice. Nonetheless the sea-sea ice transition is still visible in the maps. The RFI probability trails over 40º latitude and below -40º latitude indicate a that a RFI detection technique, in this case the RFI intensity one, is flagging these pixels incorrectly. This technique flags those pixels over a maximum threshold and below a minimum threshold, and in the case of these false alarms the minimum threshold for sea ice is over the average sea ice brightness temperatures, so all sea pixels catalogued as sea ice are flagged for having a lower brightness temperature than the expected for a sea ice pixel. A refinement of the lower threshold of this particular technique is currently under study. Furthermore, the inner seas such as the Black and the Caspian seas have their RFI probability increased, probably because the presence of RFI contamination makes the neural network to catalogue them as sea ice, and therefore the RFI detection thresholds change. This is also currently being studied.

## 5 Conclusion

This work has shown that it is possible to detect sea ice using neural networks on SMOS L1C products. The SMOS data had to be collocated in incidence angle bins in order to be able to feed a neural network. Besides, information on how the data was being observed by the instrument was necessary to overcome the instrument biases. An artificial neural network was trained using the binned brightness temperatures. To train the neural network, a dataset was generated using data from the thawing season in order to maximize the ground pixels that were not pure sea or pure sea ice. Still, sea and sea ice pixels were over represented and the $R^2$ and RMSE metrics could be artificially high when evaluating the neural network performance. A second dataset from different dates was generated to validate the neural network when geophysical parameters changed, such as the air temperature, and showed slightly worse metrics as expected.

Last, the RFI probability map that leaded to this study was recomputed and showed that the sea ice regions were successfully detected, but showed some false alarms in the sea near the poles and in inner seas. Further work is required to reduce the false alarm rate, which could include adding more features to feed the neural network, such as the latitude, and taking into account the instrument footprint when training the neural network with ground truth with different spatial resolution.

## Acknowledgements

Sea ice concentration data were obtained from https://www.meereisportal.de (grant: REKLIM-2013-04). The production of SMOS sea ice thickness data was funded by the ESA project SMOS & CryoSat-2 Sea Ice Data Product Processing and Dissemination Service, and data were obtained from AWI.

## References

[1] de Rosnay, P., J. Muñoz-Sabater, C. Albergel, L. Isaksen: "SMOS brightness temperature forward modelling, bias correction and long-term monitoring at ECMWF", ESA /ESRIN contract 4000101703/10/NL/FF/fk WP1400, November 2018

[2] Y. H. Kerr, P. Waldteufel, J. P. Wigneron, J. Martinuzzi, J. Font and M. Berger, "Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission," in IEEE Transactions on

Geoscience and Remote Sensing, vol. 39, no. 8, pp. 1729-1735, Aug. 2001

[3] A. Khazâal, P. Richaume, F. Cabot, E. Anterrieu, A. Mialon and Y. H. Kerr, "Improving the Spatial Bias Correction Algorithm in SMOS Image Reconstruction Processor: Validation of Soil Moisture Retrievals With In Situ Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 1, pp. 277-290, Jan. 2019

[4] Spreen, G.; Kaleschke, L. and Heygster, G. (2008), Sea ice remote sensing using AMSR-E 89 GHz channels J. Geophys. Res.,vol. 113, C02S03,

[5] Ricker, R., Hendricks, S., Kaleschke, L., Tian-Kunze, X., King, J., and Haas, C.: A weekly Arctic sea-ice thickness data record from merged CryoSat-2 and SMOS satellite data, The Cryosphere, 11, 1607-1623, 2017

[6] Kaleschke, L. et al. (2016), SMOS sea ice product: Operational application and validation in the Barents Sea marginal ice zone, Remote Sensing of Environment, Volume 180, July 2016, Pages 264-273, ISSN 0034-4257

[7] Powers, David. (2008). Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. Mach. Learn. Technol. 2.

[8] Kimmo Rautiainen, Juha Lemmetyinen, Mike Schwank, Anna Kontu, Cécile B. Ménard, Christian Mätzler, Matthias Drusch, Andreas Wiesmann, Jaakko Ikonen, Jouni Pulliainen, "Detection of soil freezing from L-band passive microwave observations", Remote Sensing of Environment, Volume 147, 2014, Pages 206-218