

## ***Interactive comment on “Classification of Sea Ice Types in Sentinel-1 SAR images” by Jeong-Won Park et al.***

**Anonymous Referee #2**

Received and published: 16 July 2019

In this paper the authors propose a classification method for determining ice types in Sentinel-1 SAR images. The structure and methodology are similar to those found in other studies, and overall the paper reads reasonably well. The study makes good use of recent published work by the authors on denoising Sentinel-1 SAR images, and improvements to the calculation of texture information in these images.

Specific comments:

To the best of my knowledge, this is the first study to examine the classification of Sentinel-1 SAR images for determination of sea ice types. These images are of great interest to the scientific and operational community. As the authors point out, the images are noisy, and the residual noise after the ESA correction is still significant. Certainly, the ability to classify ice types from such noisy images is of great use. However,

C1

I have difficulty following some of the claims made, in particular in the abstract and introduction. Primarily, I am not certain if it is clear to the authors that operational ice charts are generated manually, and contain significant bias and other possible errors of subjectivity. It is a little difficult to find information about this online, but the studies by Partington et al. (2003) and in the text by Johannessen et al. (2006) clearly state that the preparation of NIC charts (former reference) and AARI charts (latter reference) is through manual inspection of various sources of satellite imagery and other sources of data. Other studies (such as J. Karvonen, 2015) look at the accuracy of manual analyses by ice analysts. Training using a large volume of these charts would reduce operator-to-operator bias, but not the overall bias these charts are believed to contain since they are produced in the interest of marine safety. Based on this, the claim in the abstract and elsewhere that the use of ice charts allows training/testing data ‘void of biased subjective decisions’ should be revised.

The ‘novelty’ of using ice charts in this way as training data should be clarified. These charts are fairly similar to the training data that was used for the sea ice type classification study by Zakhvatkina [2013], where homogeneous areas identified by trained ice analysts are used. Image analysis charts, which are very similar to daily ice charts with the exceptions that they are based only on the SAR imagery, are used directly as training data in the study by Wang et al. [2017]. In that study the ice concentration information was used directly in the same manner as ice type in the present study (the available charts were mapped to the SAR image latitude and longitude), however it was ice concentration information that was used, not ice type. These similarities should be discussed.

Random forest classifiers are very popular at this time, and have been shown to be useful in many studies. To better motivate the present study, I suggest the authors compare their method to a multi-class random forest. In particular, the reference given for choosing the one-vs-all classification scheme as compared to multiclass problem is not closely related to the problem at hand. Did the authors try the multiclass method?

C2

Given that the motivation here is for operational implementation, it would be of interest to know if the multiclass method performs similarly, and the computation time difference between the binary one-vs-all method and multiclass method. I have a similar question regarding the use of all Haralick texture features. How long did it take to calculate these features over the 64 grey levels used here? Are all features needed, or is it not relevant (in the sense that the additional time required and change in accuracy is not significant). If the main contribution is to be the classifier itself, then a more careful examination of the method should be carried out. It would also be very interesting to see how the denoising methods they have developed lead to improved ice type classification. I am not sure if that would be difficult. Without this information, others are likely to attempt ice type classification without following rigorous denoising procedures. With this information, this piece of work could be a much stronger contribution to the sea ice community.

In the end, it is found that the classification accuracies are higher when considering only three classes, first-year ice, multi-year ice and open water. Could the authors add a little discussion to the conclusions as to if they envision a three-class or five-class operational implementation? If it is three-class, would they recommend using ice types from another sensor as training data? Some discussion on how the method is expected to work for other times of year should also be included.

page 6 - line 10 - Can the authors explain what they mean here by a 'sparse dataset' and why a dataset used for ice/water and ice types from SAR imagery would be considered a 'sparse dataset'? I am not sure I follow this line of reasoning.

page 6 - line 32 - Why is the 'Richard's curve' chosen over a typical curve fit? Do the authors obtain more robust or interpretable results using this method? Please provide more context.

page 7 - If I understand correctly, the authors manually selected 57 image (or do the authors mean scenes here?) for training and testing from a set consisting of 958 im-

C3

ages (or again is this scenes)? Can they say something about these 57? Are they from similar geographic regions? times of year? specific features? Going through 958 images manually to choose a training data set is not automated. Using an ice type product generated in an automated manner from another sensor (for example open water/FYI/MYI from passive microwave data or scatterometer data) could provide an automated workflow.

page 8 and Figure 6 - What method was used to determine the feature importance score and why was this method chosen? How is this score calculated?

Technical comments

1) abstract - overall accuracies vs. overall accuracy - Be consistent in your use of plurals here

2) abstract - In what way would this work support automated ice charting? Were the authors thinking that fewer operational (manual) charts would need to be produced? Clarification of this point would be helpful.

3) page 1 - line 12 - 'In most of the previous works...'... please provide a few references in this sentence to the works you have in mind.

4) page 1 - lines 12-15 and lines 20. I reiterate my earlier point. Ice charts are generated manually by trained analysts. Although they are available in the public domain, and this means using these charts directly relieves the individual designing the classification algorithm from the 'laborious' and possibly biased process of manually choosing training and testing data, it does not enable an automated workflow.

5) page 1 - line 20 - Again, ice charts are generated by humans. They contain human error. They are often produced under a strong time constraint, and in the interest of marine safety (the latter point meaning they likely contain bias to ensure safety).

6) page 3 - line 25 - 'ice edge determined from AMSR-E' - an ice edge cannot be determined from AMSR-E without using an algorithm. Which algorithm was used?

C4

Please revise.

7) page 3 - lines 28-29 - I don't know what the authors mean by 'has a precision of decimals'.

8) page 3 - line 26 - Similar comment regarding the ice edge determined from SAR - a methodology must have been used to get this ice edge. Was it visual inspection, or another method? Please revise.

9) page 4 - line 3 - better than what?

10) page 5 - line 2 - wording is not specific - many of the previously developed methods - methods for what? These references are a mixture of ice/water and ice type classification studies. These two tasks are different from the perspective of a computer algorithm. Also, a reference to Shokr [1991] should be included.

11) page 5 - lines 9-11 - I don't understand this sentence, what is averaged for multiple distances, and what is the normalized GLCM?

12) page 5 - line 5 - direction? or should this be orientation?

13) page 5 - line 12 - The term spatial resolution is not clear. Some authors consider this the scales that are resolved. It may be better to state that the spacing between the GLCM texture feature windows is 1km? (or please reword if I am not interpreting this point correctly).

14) page 5 - It would be nice to have the Haralick features listed in a table, and to provide a brief rationale for including all of them in the study. Information as to how long it took to calculate these features using 64 grey levels for their set of imagery is also important.

15) page 5 - the number of Haralick features is referred to inconsistently as 13 on line 7 and 26 on line 25. The 26 is likely just accounting for the two polarizations, but the two should be referred to in a consistent manner. Similarly on page 7 lines 22-23, please

C5

use either 'Haralick texture features' or 'texture features' consistently when describing the three classifiers.

16) page 6 - line 16 'they are' - what are 'they'? Is this the number of operations?

17) page 7 - If a binary ice/water classifier is 'simple' (line 6), why are the authors starting with ice type classification? I suggest this be reworded.

18) page 8 - lines 20-21 - The sentence starting with, 'Since the training and test datasets were extracted from the same...' I find a little out of place. With this placement, it seems like it is trying to account for the results from FC2 and FC3. It might be better to start this one with 'When the evaluation is carried out with the 2018 data, the training and test datasets....'

19) page 9 - line 32 - 'capturing' should be 'to capture'

20) page 9 - line 32 - 'more details' - as compared to what?

21) Figure 3 - Could the authors provide some information in the text as to what the map of partial concentration is (top right). Is this the partial concentration of the dominant ice type for the given polygon?

22) Figures 3,7,8 and 9 should have geolocation data provided.

23) all numbers less than ten should be written out in words, eg., 3 -> three

#### References

A comparison between high-resolution EO-based and ice analyst-assigned sea ice concentrations, Juha Karvonen and others, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(4):1-9, 2015.

Evaluation of second-order texture parameters for sea ice classification from radar images, Mohammed E. Shokr, Journal of Geophysical Research, 96(C6),10,625-10640, 1991.

C6

Late twentieth century Northern Hemisphere sea-ice record from the U.S. National Ice Center ice charts, Kim Partington, Tom Flynn, Doug Lamb, Cheryl Bertoia and Kyle Dedrick, *Journal of Geophysical Research*, 108(C11), doi:10.1029/2002JC001623, 2003.

Remote sensing of sea ice in the Northern Sea route: Studies and applications, Ola M. Johannessen and others, Springer Science and Business Media, 472 pages, 2006.

Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network, Lei Wang, K Andrea Scott and David A Clausi, *Remote Sensing*, 9(5), doi:10.3390/rs9050408, 2017.

---

Interactive comment on The Cryosphere Discuss., <https://doi.org/10.5194/tc-2019-127>, 2019.