This manuscript introduces a new method for the ice mapping based on dual-polarized Sentinel-1 SAR data during summer season. The proposed method was developed from a conditional random field based on mixed statistical distribution. The results indicate the potential to derive reliable ice extent operationally. Unfortunately the author's use of English is very poor, and the meaning was ambiguous and confusing in most cases. Missing methodological details, incorrect use of models, and the large number of typo and formatting errors do not make an impression of a self-contained manuscript. The authors should not submit a manuscript which is not ready for submission. I would recommend a rejection of this paper, but I think that the author could have a chance of publishing the results of their study if they prepare the manuscript better next time.

We are grateful to the reviewer for the constructive comments on our manuscript (tc-2021-85) entitled "Sea ice and water classification on dual-polarized Sentinel-1 imagery during melting season". We have addressed all the comments. Our point-by-point responses are attached below in blue, while the original Reviewers' comments are in black. Thank you again for valuable comments on our manuscript. We will go through it and revise the parts of the manuscript following the reviewer's suggestions.

General Comments:

1. The description of the methodology is so poorly structured, which makes the logic of the research very confusing and hard to follow. One example is that the description of data preprocessing and training samples selection should not be introduced in the section of methodology. I am pretty sure a lot of efforts are still needed for improving the general structure of the paper.

Response: Thanks for your comments and suggestions, we will rewrite the manuscript to make it more understandable, but we think the training sample selection and preprocessing are also the key steps of the sea ice and water classification procedure, and so they actually fit into the methodology section.

However, we will add a clearer description of the training procedure, the definition of research area and the data used, you can see it in the response to reviewer 1 as well.

2. The training samples was selected from the MET Norway ice chart. The MET ice chart is a weekly product and it inevitably has a time lag with SAR data. The change of sea ice is fast in melting period. How do you make sure the samples you choose are correct?

Response: The MET Norway ice charts are a daily averaged product, but are only available from Monday to Friday, and it can also be used for daily sea ice classification. We select the samples using a combination of MET Norway charts and visual inspection. The MET Norway ice charts are currently the only sea ice product that can be obtained with a temporal resolution of one day. The reference (Zakhvatkina, 2017) also use the MET Norway ice chart product for training and verification of sea ice classification. Therefore, we chose the MET Norway ice charts in the paper for sample selection and validation. In order to improve the accuracy of sample selection, we have also included visual inspection. Besides, we have also analyzed in the paper that due to the difference between the SAR data acquisition time and the MET Norway ice chart acquisition time, the drift and freeze-thaw changes of the sea ice also affect the classification accuracy.

3. In the step of incidence angle correction, the authors used an incorrect sea ice scattering model. In equation (1), the backscattering of sea ice is described as the function of nadir backscattering and $\cos^n(\theta_i)$. When the radar echo is incident vertically, the scattering mechanism of sea ice is specular scattering which is completely different from the scattering mechanism of SAR. Therefore, the used approach is illogical and unphysical.

Response: You are right. The equation (1) is removed from the manuscript, we only use the measured backscatter value. Usually the backscatter depends on the incidence angle, and for Sentinel-1 SAR images with an incidence angle larger than 15°, the backscatter decreases with increasing incidence angle. The scattering mechanism of sea ice includes volume scattering (relevant for multi-year ice). We refer to a publication from TGRS.

W. Lang, P. Zhang, J. Wu, Y. Shen and X. Yang, "Incidence Angle Correction of SAR Sea Ice Data Based on Locally Linear Mapping," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 6, pp. 3188-3199, June 2016, doi: 10.1109/TGRS.2015.2513159

4. The mean-shift method is critical to the proposed classification method. But the principle of mean-shift algorithm and the parameter setting for unsupervised segmentation should be introduced.

Response: The mean-shift method is a classical segmentation method in image processing. We did not introduce the details of the mean shift algorithm in this paper too much, but gave the mean-shift parameters setting. You can find detailed information on the mean-shift method in the reference, and it will be included in the next version. The principle of the mean-shift algorithm is to first define an offset value of the backscattering coefficient, and define the point where the difference between the backscatter coefficient and the current pixel point meets the offset value as the same clustering unit. In this paper, we give the parameter settings for mean-shift over-segmentation. For example, if the number of pixels within one superpixel is greater than 5000, and considering the heterogeneity of superpixels, In our manuscript, we define sub-superpixels of smaller units for each superpixel. We calculate about 100,000 sub-superpixels among them, and the average size of one sub-superpixel is 24 pixels.

Lang F, Yang J, Yan S, et al. Superpixel segmentation of polarimetric synthetic aperture radar (sar) images based on generalized mean shift[J]. Remote Sensing, 2018, 10(10): 1592.
Ming D, Ci T, Cai H, et al. Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm[J]. IEEE Geoscience and Remote Sensing Letters, 2012, 9(5): 813-817.

5. What I am most dissatisfied with is the use of distribution models. The distribution model of Gamma, Weibull and Alpha-stable is based on the statistical characteristics of pixels. However, the distribution model was for “sub-superpixel” (patches derived from unsupervised segmentation method) not for pixels. I don’t think these distribution models could be adaptable to image patches.

Response: we are sorry for the confusion. We think the reviewer has misunderstood this part. We will improve the sentence in the manuscript. In fact, the statistical distribution is fitted to the distribution of all the pixels within one sub-superpixel. Meanwhile, the statistical distribution is used as an input feature for super-pixels or sub-superpixels to calculate their potentials.

6. There are many SAR sea ice classification methods, taking these methods as baselines and comparing them with your method is necessary for validating the effect of your method. Moreover, the authors claimed that the advantage of proposed method is to identify sea ice in melting season. So you should give more examples to prove that the developed method can solve the problem of sea ice classification in summer.

Response: In this paper, the statistical distribution based CRF is proposed for sea ice and water classification. In order to verify the effectiveness of the algorithm proposed in this paper, the SVM algorithm (Zakhvatkina et al, 2017) is used as the comparison method in this manuscript. In their study, Zakhvatkina et al. (2017) use Radarsat-2 data to achieve a good detection effect (~90%), but the classification accuracy is poor (~75%) in summer. This paper focuses on the sea ice classification in summer, and the result shows that the method in this paper has better seasonal adaptability than the SVM method, and the classification accuracy in summer can reach about 90%. Compared with ASI sea ice concentration in summer (figure 13 and 14), our method provides satisfying results. To our knowledge, SAR images have so far been a focus in winter, now we are using SAR images to solve the problem of sea ice classification in summer.

7. According to the results of Table 4, the classification accuracy depends on used reference incidence angle. In equation (2), $\cos^n(\theta_{ref})$ is a constant value. I don't understand why the variation of constant value has an impact on classification accuracy.

Response: Thanks for your comments. Varing the reference angle will change the backscatter value, which will also have an influence on the MSTA-CRF training, and the classification result will also be different. Referring to table 4, the selection of the incidence angle has little effect on the classification accuracy. However we have to select one reference angle for classification.

In the incidence angle correction, we transform the backscattering coefficient of the entire image in the Sentinel-1 data to the reference incident angle θ_{ref}, and it will be used as the input data for the sea ice classification experiment. In order to verify the optimal θ_{ref}, we have designed the experiment of reference angle selection. In Table 4, the θ_{ref} is set in the range of 20-40 degrees, and the CV is used as the criterion. The result shows that the optimal reference angle is 23°. Therefore, all the classification experiments in this manuscript have corrected the backscatter of the original image to 23° in order to obtain the final classification result.

8. How to determine the parameters used in the proposed method (e.g. n and weight coefficients) is not clarified. Many details are not clear and need further explanation.

Response: We will try our best to describe this in more detail. The parameter estimation of the mixed statistical distribution adopts the EM (Expectation Maximization) method, which can be found in the paper by Tadjudin (2000). The general idea is to first calculate the distribution parameters of each statistical model, and then estimate the weight parameters of each distribution in the mixed distribution.

Tadjudin, S., and Landgrebe, D. A.: Robust parameter estimation for mixture model, IEEE T. Geosci. Remote, 38, 439–445, https://doi.org/10.1109/36.823939, 2000.

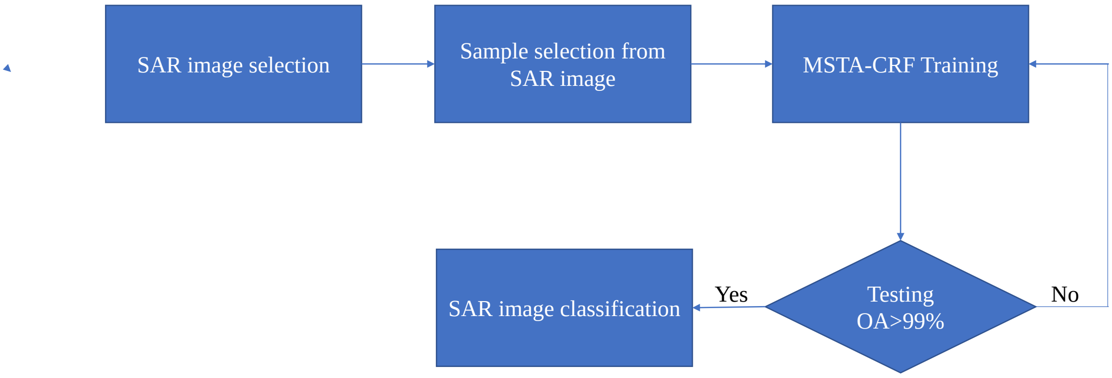
9. As the stated by the author, the accuracy of classification was validated by all the training data (see Page 10 Line 6). This is obviously incorrect. I am very confused about the sentence "If the overall accuracy (OA) is lower than 99%, we add 100 patches (50 for ice and 50 for water) from the rest of the training dataset to train the revised model, ……". I'm not sure of your reasons for doing this?

Response: The review is right. We only use a small aspect of samples in the training dataset and the rest for testing. Considering the completeness of the training samples and the problem of overfitting, we only used part of the training sample set for model training, and the remaining samples are used for verification. During the training procedure, we first randomly select 100 samples from each category (sea ice and water) in the training dataset to obtain an initial model, and then use this model on the remaining samples to verify its accuracy. If the accuracy is lower than 99%, 50 samples are added for each class to update the model, and the added training samples are removed from the test samples until the final classification accuracy on the test data is better than 99 %. We repeat the training procedure ten times and find that when the training samples reaches 1000 the accuracy is over 99%. We finally selected 1000 samples for model training (500 for each category), which accounts for 10.25% of the entire training dataset. The table and the corresponding flowchart of the training procedure are listed below and will be included in the next version of the manuscript.

Step 1	SAR image selection One SAR image on each day from June to Sept in 2015-2018 is randomly selected to construct the training data set, finally we get 488 images.
Step 2	Training and testing data set construction: 10 patches (samples) for each category (ice and water) with the size of 64*64 pixels are randomly selected from the 488 SAR image using MET Norway ice charts, then we get 9760 patches for constructing the training data set.
Step 3	MSTA-CRF training: 100 patches for each category are selected for training the MSTA-CRF model, and the rest are used as testing samples to decide by the overall accuracy whether the training will be repeated.

Step 4	Testing: If the overall accuracy on the testing samples is larger than 99%, then we get the final MSTA-CRF model, otherwise 100 patches (50 for each category) will be added to retrain the MSTA-CRF model, and the newly selected 100 patches will be removed from the testing samples.
Step 5	SAR image classification: Repeat step 3 until we train a satisfied model, and the newly trained model will be used for sea ice and water classification on all the SAR images.

We also give the flowchart of the training procedure in the following figure.



Minor Comments:

Page 2, line 5: “search-and rescue” --> “search-and-rescue”.

Response: corrected.

Page 2, line 6: “ERS-1/-2, RADARSAT-1/-2, Sentinel-1A/-1B” --> “ERS-1/2, RADARSAT-1/2, Sentinel-1A/B”.

Response: corrected.

Page 2, line 14: “introduced” --> “have”.

Response: corrected.

Page 2, line 15: “channel” --> “polarization”. Please replace “channel” with “polarization” in the full text.

Response: We have checked the manuscript and replaced “channel” with “polarization”.

Page 2, line 18: “for improved” --> “for improving”.

Response: corrected.

Page 2, line 27: here Radarsat-2 is “RS-2”, but its abbreviation is “RS2” in line 6.

Response: We have corrected this and finally use RS2 as the abbreviation.

Page 2, line 28: “-3” --> “Sentinel-3”.

Response: corrected.

Page 2, line 29: “with low resolution passive microwave form low resolution microwave from AMSR2” reformulate this sentence.

Response: corrected. The operational Ice Service at the Norwegian Meteorological Institute (MET Norway) uses high-resolution SAR data from Sentinel-1, RS2 and COSMO SkyMed, in combination with optical imaging from Sentinel-2 and Sentinel-3, NASA Suomi NPP VIIRS, NOAA AVHRR for visual interpretation, with low-resolution passive microwave from AMSR2 used as a last resort if no other data is available.

Page 2, line 32: “As the backscattering is usually affected by ocean waves propagating into the ice area, ….” for thin sea ice, the backscattering coefficient could be affected by wave. But for thick sea ice, the effect of waves on backscattering is very low.

Response: We agree with the reviewer’s comment. During the melting season, the backscattering coefficient could be affected by ocean, the sea ice classification accuracy may be affected by using the textual feature based SVM method. Our study is to propose a sea ice classification algorithm for the melting season.

Page 3, line 7: “SVM realize” --> “realizes”.

Response: corrected.

Page 3, line 7: “by training the kernel with the transformation into high dimensional space,” reformulate this sentence.

Response: corrected. The SVM method uses the kernel function to project the feature into a high-dimensional space.

Page 3, line 9: “Textual feature based neural network methods also shows” --> “show”.

Response: corrected.

Page 3, line 10: “Murashkin et al. (2018) use” --> “used”.

Response: corrected.

Page 3, line 11: “th MIZ” --> “the MIZ”.

Response: corrected.

I stop here with my comments and I think I almost had comments in every single sentence. There are a lot of grammatical issues but also, more seriously, inaccurate statements.

Response: We have gone through the manuscript and revised all the text, and removed ambiguous passages. We have also taken advice from native speakers.