

Reply to Anonymous Referee #1 from 21 Oct 2017

*Note: Author responses are in plain text following the original referee comment shown in italicized text.*

*The figures, in particular the captions need some work. My opinion is that figures and captions should be stand alone, such that a reader should be able to understand what each figure is without reading the text. In some cases adding a key or a description of symbols in the figure caption would achieve this goal. For example, Figure 3 has no key for the classified image.*

We have added a classification key to Figure 3. Thank you for the suggestion – we have also incorporated edits throughout to make image captions more complete and descriptive.

*Line 50. Missing a reference to Fetterer and Untersteiner (1998).*

We have included this reference as well as one to Arntsen et al. 2015.

*Line 63. When discussing alternative classification methods, it would be good to enumerate those methods applied to classification of sea ice. Furthermore, the authors give maximum likelihood classification as an example or an unsupervised algorithm. Maximum likelihood can be a supervised algorithm.*

Each reference on line 63 details a classification method applied to sea ice. We have edited line 98 to indicate that those references refer to a method applied to image processing generally, and not specifically to the classification of sea ice.

It is true that maximum likelihood classifiers can be supervised in some cases - we have revised this discussion of unsupervised classification algorithms to be more precise.

*Section 4.1. In the analysis of seasonal evolution of surface characteristics, my guess is that the field of view does not contain the same ice. The authors should make this clear or state why they think it is the same floe/ice. How fast is the ice likely to be moving at this location?*

You are correct – the ice seen at this location is not a single floe. This explains the sudden increase and then subsequent decrease in open water fraction in late June, as well as the completely ice-free water by August. We have included the following sentence in Section 4.2 to clarify that we are looking at Eulerian rather than Lagrangian view: “The site is Eulerian; it observes a single location in space rather than follow a single ice floe through its lifecycle as it drifts”.

*Line 433. I disagree with the statement that misclassification means that the algorithm fails to replicate human decision making. That might be the goal but one that is impossible to reach. To my mind, misclassification indicates that the algorithm doesn't give the same answer as a human would.*

We agree that we are not replicating the decision-making process, but rather the end result. We have revised this section to clarify our definition of misclassification. Line 474 was changed to: “The algorithm either assigned the same classification as a human would have, or it did not”, and line 478 has been rewritten to: “The first type of internal error is misclassification error, where the image classification algorithm fails to assign the same classification that a human expert would choose”.

*Line 531. This analysis is interesting but does it apply to the image used or to all images in general. How might the result change through time or with season?*

This analysis applies specifically to the image used, though there is nothing particularly unique about the image analyzed. We believe the results are applicable to all images in general, but a complete demonstration of that is better suited to its own study, and we suggest future work to expand this analysis to a general rule. The statistical methods that we use here should be independent of the seasonality of ice (so long as the metric you are investigating can be captured by the image scale, e.g. this works for melt pond fraction and not for ice fraction). At the core, we are just using a sample of some size as a means to estimate a population statistic.

We have added line 598 to detail this in the manuscript: “The statistical approach for determining image statistics should not depend on the seasonality of the image nor the type of image used so long as the total area observed is sufficiently greater than the variability in the surface feature being investigated.”

*Line 558. The Central limit Theorem is a mathematical theorem complete with proof. I wouldn't say that it can be tested. What you are doing here is evaluating if you can predict the regional/image mean from a set of smaller*

*samples/local means. One framework to evaluate this is hypothesis testing in which you pose the hypothesis that  $N$  sample means can predict/estimate the regional mean. This test applies the Central Limit Theorem but does not test it. This section needs to be reworked.*

*Line 559. The standard definition of the Central Limit Theorem is that independent variables can be added and normalized by  $(X - \mu)/(\sigma/\sqrt{n})$  to yield a normal distribution  $N(0,1)$ . Where  $X$  is the sample mean,  $\mu$  population mean and  $\sigma$  population standard deviation.*

These are good points. We have significantly revised this section. The takeaway message remains largely the same as in the original version, but both the text and methodology has been improved for new version. Instead of applying the Central Limit Theorem directly, we instead analyze the sample size required to estimate the regional mean, and address the difference in measuring from spatially correlated samples versus randomly selected samples.

*Figures: Check figure numbering*

Thanks for catching this – we have fixed the figure numbering issue.

*Figure 13. The cyan rectangle over the black dots make the dots look green (at least on my screen). The labels in the key need to match the description in the text. If some of the information is not discussed, I suggest removing it from the figure.*

We have fixed the coloration issue with the mean dots.