Reply to Reviewer #3

The original comment by the reviewer is in black, while our replies are in green. Text directly copied from the original submission is in purple to help facilitate referencing the original submission.

General Comments: This paper presents a good overview of the importance of ice charts for not only navigation guidance, but as a relevant archive for climate monitoring and use as a validation variable. This is a very nice paper that provides some good and novel approaches to evaluate how to provide some level of uncertainty in ice charts, particularly with the Kappa statistic and Krippendorff's alpha applications. Additionally, this is hopefully the start of a trend in producing much needed literature that focuses on the role of ice charts in research and development for environmental monitoring. This includes the challenges with how different data sources should be handled and the variation between information input, including the subjective nature of ice analysts. This topic has always been of great interest to the research community that depend on ice charts as a source for validation and to provide initial conditions in models. With the onset of large volumes of data and predicted increased activity in the Arctic ,the ice services are preparing for an integration of automated operational products to assist them in providing accurate information for end-users. For this reason, it is crucial that papers focusing on this topic can clearly communicate 1) why this is important in the operational sense, 2) fundamental challenges in automating sea ice information for operations (which has been going on for the last 30 years) and 3) apply metrics relevant to evaluate automated techniques to the level that can be useful for operations. This paper touches on these three points and opens up more opportunities to explore this topic further in future research.

We agree with many of the points the reviewer has made above regarding the relevancy of this area of research and current challenges that we face.

There are some overall suggestions that would be good to include which would improve the paper. The geophysical limitations with monitoring different sea ice types and concentration relative to the season was mentioned in the conclusion but should be stated earlier in the paper because this is one of the main challenges in sea ice automation. There was no mention of the difference of automation with passive microwave and SAR and seasonal/ regional limitations. Often this has been confused in the research community and should be clear that both sensors will be limited by the same environmental conditions (i.e. wind (noise) and melt). This is will help to clarify the sentence in P3 L29 regarding the melt season caveat.

We would like to thank the reviewer for making overarching suggestions for placing this paper within the context of automated detection algorithms (including current difficulties and why this has not already been implemented across Ice Services). Our original submission discussed this much more briefly.

We have modified the paper to address the difficulty of monitoring different types of sea ice earlier, as suggested. This helped to explain how the polygons were delineated for this study as well. We address this point further below when addressing the specific comments by the reviewer.

Title should reflect that this comparison is only a case study using HH polarization in SAR. The current title and the inclusion of passive microwave suggests automation is being applied to various sensors and polarizations. Current classification applications that ice services are evaluating are primarily focused on dual-pol SAR and experimenting with full polarimetry as well.
The title of the paper has been modified to show that only HH was used in this study.

P2, L3. By stating the analysts visually segment the pixels suggests there is some sort of segmentation application applied here. I am assuming this is referring to how analysts make the decision on where ice and open water is located in the image? If so, state that the ice analyst is able to determine these areas and manually delineates them with polygons. Then you can omit the statement ("this segmentation is not drawn)".
The sentences in question: "Analysts at the CIS delineate boundaries around regions with similar ice conditions, for navigational purposes (we refer to these as polygons in the following). Next, they visually segment the pixels inside the polygon into ice or water pixels (this segmentation is not drawn). The analyst then assigns an estimated concentration value for the polygon using the visual segmentation".

Correct… We have changed the sentence to "Analysts at the CIS identify areas with similar ice conditions and open water for navigational purposes, then manually delineate them with polygons. The analyst then assigns an estimated concentration value for the polygon."

P2, L6. Here would be a good place to refer to the ice chart manual, MANICE, put out by CIS or the WMO. You can add that reference after this sentence. Please review: https://www.ec.gc.ca/glaces-ice/?lang=En&n=2CE448E2-1&offset=8&toc=show
A citation has been added for MANICE when describing types of ice charts produced at the Canadian Ice Service.

P2, L28. It would be good to provide a sentence or two on why automating ice information has been a challenge for ice services in the past. We are using the same types of sensors that have been available since passive microwave has been available and with the beginning of the use of SAR in the 1990's. Ice services continue to rely on manually drawn charts because automation for sea ice has significant limitations at the marginal ice zone, coastlines, first year ice types and ice edges for spring and summer sea ice, where we see the greatest amount of traffic in the Arctic.

Section 2.1 was greatly revised to address and expand on these issues. Please refer to the revised document for this section.

P3, L20-25. Would be good to include when CIS started to use SAR and the amount of SAR vs. PMR that is currently used today. From the understanding in ice services, CIS charts primarily consist of SAR and other high resolution data and only use PMR sparingly because it is unable to detect sea ice features and coastal zones well.

We have added this paragraph to section 2.1. Specifically, we added the following paragraph:

"The CIS relied on RADARSAT-1, a SAR sensor, for ice charting beginning in 1996 until its decommissioning in 2013. The Canadian Ice Service currently relies predominantly on

RADARSAT-2, but will start to use the RADARSAT Constellation Mission (RCM) operationally, following its recent launch in 2019. In the 2017 calendar year, the Canadian Ice Service received approximately 45 000 SAR scenes between Sentinel-1 and RADARSAT-2, and another 85 000 scenes from various satellites including GOES, MODIS, AMSR, and VIIRS. The lower number of SAR scenes reflects the fact that RADARSAT scenes are geographically targeted acquisitions ordered by the CIS, while GOES, MODIS, AMSR and VIIRS are publicly available swaths acquired for general use. The latter are less targeted for CIS Operations, but useful as a secondary, supplemental data source."

P3, L30-31. Should provide a statement that the current state of automation for ice services will rely on the two channels (dual-pol) until a compact polarimetry is available. Therefore, would be good to provide an explanation as to why HH only used in this study when the HH/HV has been available with RS2 since 2007.
Regarding dual-polarization, we added the following in section 2.1

"Automation of sea ice classification algorithms currently use dual-polarization imagery, but will use compact polarimetry as it becomes available."

With respect to why only HH was used, we added the following text in section 3.2.

"Only the HH band was used for both segmentation and visual interpretation in this study. Typically, ice charting is done with HH as a primary polarization, and HV is only used to distinguish ambiguous ice types. However, the sample polygons used in this study focused on examples with minimal ambiguity."

P4. Figure 1 is misplaced in the manuscript and should be located after its mention in the text.
All figures have been moved to the end of the document (as it was supposed to be according to the template).

P4 L2-6 . Include reference to MANICE manual and also the Dedrick paper (K.R. Dedrick, K. Partington, M. Van Woert, C.A. Bertoia & D. Benner (2001) U.S. National/Naval Ice Center Digital Sea ice Data and Climatology, Canadian Journal of Remote Sensing, 27:5, 457-475.
Reference has been added to the paper.

P4 L9 Include reference to WMO manuals 259: JCOMM Expert Team on Sea Ice (2009) WMO Sea-ice Nomenclature, WMO/OMM/ĐŠĐIJĐ̄d - No.259 Suppl.No.5. Lin-guistic equivalents. Geneva, Switzerland, JCOMM Expert Team on Sea Ice, 23pp.(WMO No, 259, Suppl. 5). http://hdl.handle.net/11329/113.
We have added this reference to the paper.

P5, L11. What does "greatest intersecting overlap" refer to? Please provide a more clear explanation.
The sentence in question. "The polygon sizes were compared to polygon sizes from the published operational daily charts and image analyses that used the same RADARSAT images. The sample polygon sizes were compared to the sizes of polygons from published charts with the greatest intersecting overlap."

We tried to find a polygon from the published ice charts that corresponded to our samples. Since our sample polygons sometimes spatially overlapped more than one polygon, we took the one with the greatest intersecting overlap.

We have changed the phrasing in the paper to be:
"Since the polygons were delineated differently, sometimes the sample polygon would spatially intersect with two or more polygons; making it difficult to directly compare the sizes of polygons. We addressed this by identifying the polygon with the greatest spatial intersection with the sample polygon, and comparing the two areas."

P5, L12-16. Where is the comparison done for the Wilcoxon-Mann-Whitney test? Is this a general comparison that had been done before with ice charts or is the image analysis referring to the new polygon generated by an analyst for comparison? Also, and what is "image analysis" referring to? Is this automated image analysis. It appears that you are referring to a previous comparison that had been conducted because you specify the new polygons in this study in the following sentence in P5 L13.
The sentences in question: "The polygon sizes were compared to polygon sizes from the published operational daily charts and image analyses that used the same RADARSAT images."

We believe there is some confusion as to the difference between image analysis charts, daily charts, and the samples generated for our study. We have since added a new citation for MANICE, which explains what an image analysis is.

We found polygons from published image analysis charts and daily charts, then compared the sizes of those polygons to the polygon samples in our study to assess if the sample polygons' sizes greatly different from the sizes of polygons from image analyses and dailies.

We have changed the text as follows. "The polygon sizes were compared to polygon sizes from two types of published operational charts: daily charts and image analyses. The image analyses and daily charts used the same RADARSAT images that were used to delineate polygons used in this study. Of interest was determining if there were differences in the size of polygons drawn for this study and the sizes of polygons in published charts, since polygon sizes could impact analyst ability to estimate ice concentration."

P5, L25-27. Already stated in previous section that only HH is used in this study. Should instead provide reasoning as to why only HH is used when it is first mentioned in P3 L30-31. Again, why is HH only used? You state to "ensure only difference in ice concentration estimates between individuals were restricted to only interpretation of the segmentation, rather than interpretations of

the multiple polarizations normally available," however, how does the interpretation of only using one polarization differ from multiple polarizations regarding introducing any bias in the analyst interpretation?
We have revised section 3.2 to explain why only HH was used.

"Only the HH band was used for both segmentation and visual interpretation in this study. Typically, ice charting is done with HH as a primary polarization, and HV is only used to distinguish ambiguous ice types. However, the sample polygons used in this study focused on examples with minimal ambiguity."

Since we selected only samples with high separability between ice and open water, there was very little ambiguity. HV is typically used for differentiating sea ice in difficult conditions, which was not the case in our study. Therefore, only HH was used.

P7, L3-4. We should assume the analyst understands the user interface before doing the assessment. Unless there is something that could be shown with these two polygons that demonstrated the analysts understood the user interface, this disclaimer does not need to be here.
Although we agree, we felt this was worth addressing for the Operational folks who would question whether some of the results were due to lack of understanding. Also, we felt it was worth mentioning because the analysts were initially confused by the way the data was being presented to them. This may be useful information for others trying to replicate a study of this type.

P7, L15-16.
One of the major challenges with automating products for operational ice charts are due to the large differences of surface appearances based on regional and seasonal variability. It was noted in the abstract and conclusion but since this is a common problem it would be better to state it in the beginning paragraphs or something that provides information on how monitoring sea ice in these areas vary with respect to region (fast ice vs. drifint ice) and season, particularly with the melt and Summer season. It will help the reader to have context as to its difficulty and why this data hasn't previously been automated for ice services. Additionally, a table listing the images and dates should be included somewhere so that the reader can get an idea of the types of ice conditions that were being assessed in this study.
With respect to the first part of this comment, 2.1 has been revised to contain information about automated classification of sea ice in satellite imagery, with some examples of difficult conditions for automatic classification.

A table has been added to the appendix listing the image acquisition file names and date of acquisition.

P19 L1-6. This paragraph describes the types of images and criteria that were selected for this study. The description of the area selection with regards to the contrast and floe size would be best placed in section 3.1 in order to help set the stage for the study. This section in the conclusion can reiterate and summarize this again and continue to expand on it more detail as you have in the later part of the paragraph.

Agreed; this paragraph was moved to 3.1. Judging by the reviewers' comments, there was definitely a lack of clarity in how the polygons were drawn so it makes little sense that this explanation was given at the end, in the conclusions.

Technical Comments

P5 L9. "…polygons created by the analyst were compared to the corresponding areas from the published operational charts that used the same Radarsat images."
This is better phrasing but the wording was already addressed/changed with a previous revision: "Since the polygons were delineated differently, sometimes the sample polygon would spatially intersect with two or more polygons; making it difficult to directly compare the sizes of polygons. We addressed this by identifying the polygon with the greatest spatial intersection with the sample polygon, and comparing the two areas."

P13, L4. "The first objective of this study was to compare analyst…"
Original sentence: "This part of the study focuses on the first objective of this study, which was to compare analyst…"
We agree with the correction.

P13, L6-7. The sentence "Segmentation is not necessarily…" is a very strong statement and could be refuted in some ways without any resources to provide support. You can replace it with something that describes it similar to the following justification: Segmentation papers tend to explore very limited samples of satellite data which they do very well but there are not many papers that apply the same types of techniques across a wider spatial and temporal scales. Whereas ice charts have been produced on a consistent basis for more than 40 years by a wide range of different agencies. Though we know there are differences among ice charts, they overall agree where and what types of ice are present within a given area, with small variations."
We replaced the text with your suggestion accordingly. "Automated sea ice classification algorithms often use sea ice charts as a truth dataset for verification since they cover large geographic areas and have been produced for many years by many Ice Services. Furthermore, while there are differences among ice charts, they generally agree with respect to types of ice present and where they occur."

P14, L2. Replace "…ranged…" with "…the accepted segmentation results varied between analysts."
The original sentence in question: "Furthermore, analysts ranged in their level of acceptance of the segmentation results."
We agree with the correction.

P14, L3. Rephrase to state "In 36.8% of total polygons, the analysts were unanimous in agreement with the outcome of the automatic segmentation."
The original sentence in question: "In 36.8% of the total polygons, all analysts unanimously stated they accepted the segmentation results."
Changed; we agree that the reviewer #3's phrasing was superior.

P15 L1-3. "An overall agreement between analyst estimation and segmentation results are shown along the diagonal line, where the proximity of entries outside the line represent the extent to which analysts are over or underestimating ice concentration (Figure 7). There was an overestimation of ice concentration from the analysts with respect to MAGIC."
The original sentence in question. "Perfect agreement between analyst estimation and the segmentation results lie along the diagonal; entries below (above) the diagonal show over (under) estimation by analysts: the analysts tend to over-estimate the ice category with respect to MAGIC."
Changed; we agree that the reviewer #3's phrasing was superior.

P15, L4-6. The first sentence is redundant and could flow with the following sentence by combining them and not repeating the same results to state: "Over-estimation of low ice concentration (i.e. 2/10 to 4/10) resulted in an increase in the number of polygons with high ice concentration (9/10 to 10/10)."
Changed to the reviewer's suggestion.

P15, L7-8. Delete "In the cases where some analysts accepted the segmentation results, while others did not, we only considered the responses where it was valid." Already stated in the first sentence in L7.
Changed to the reviewer's suggestion.

P15, L9. Delete the parentheses because this is an individual sentence: "Figure 10 shows the combined responses from all participants in this study. (Individual responses are shown in Figure 11)."
Changed; we agree.

P17, L27. Sentence does not need to be put in parentheses.
Changed, agreed.

Misplaced throughout the manuscript and should be more closely aligned with the text throughout the document. They should be placed after the mention in the text rather than before, or after subsequent figure references.
All figures were moved to the end of the document (as it was supposed to be according to the template).