Reply to Reviewer #2

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.

The paper presents a comparison of ice analyst assessments of sea ice concentration values at the Canadian Ice Service, both between analysts and against an automated retrieval method. It is well written and I see only some minor issues that can be fixed or explained better. The study is limited to Canadian Ice Service analysts, and finds very good agreement between them. With more analysts, including from other ice services, it is likely the range of opinions will increase. However running this type of exercise is time intensive, so only a limited number of such studies have been per-formed (i.e. Karvonen et al, 2015). Other studies have tried examining the areas where ice charts overlap, for example a comparison between Norwegian and Danish ice charts at https://www.ecmwf.int/sites/default/files/elibrary/2018/17968-satellite-observations-sea-ice-concentration-and-drift.pdf(p.14-15). More of this type of paper evaluating ice chart data are needed to provide quantitative measures of uncertainty that can be used when ice charts are used as training data for automated classification algorithm developments and as climate records.
We were quite fortunate to have eight analysts participate in this study and agree that there is difficulty in finding operational analysts available to participate in research studies, which is likely why there are few studies in this area. We also agree that more papers quantifying uncertainty in ice charts would be useful and are hopeful to continue to contribute to this area of research.

P2, L2. Analysts at the CIS delineate boundaries around regions with similar ice conditions. This raises the question, what was the rationale for the example polygon areas in Figures 3 and 4 as these contain dissimilar ice conditions?
We used different criteria to delineate sample polygons than the criteria normally used to delineate polygons in regular ice charts. Emphasis was placed on identifying areas with fractional ice cover, since there is little value in evaluating analysts' ability to estimate 0/10 or 10/10 ice concentration. Secondly, polygons were not drawn with the purpose of classifying the entire image; instead, polygons were only drawn in areas with fractional ice cover and high contrast between ice and open water to maximize the automated algorithm's ability to clearly demarcate ice from open water.

P2 L28-34. This uncertainty is not limited to ice charts, automatic satellite derived sea ice concentration products also have varying levels of uncertainty.
The specific lines referred to are:
"The uncertainty of sea ice concentration estimates can result in downstream uncertainties for applications that rely on sea ice charts. For example, sea ice concentration estimates from Canadian Ice Service charts are used as a data source for input to initialize sea ice models (Smith et al., 2015; Lemieux et al, 2015). The error in the initial condition of sea ice concentration estimates can propagate and grow with time, and impact the accuracy of prediction from numerical models (Parkinson et al., 2001). Uncertainty of ice concentration estimates could also impact the accuracy of climatology studies of ice concentration derived from operational ice charts, although that has not been investigated."

Agreed; satellite derived products also have uncertainty. We tried to minimize the uncertainty from automated satellite derived products in this study by having analysts validate the output from the segmentation.

P3 L31 and P5 L25
How does limiting the analysts to only HH affect their choice of concentration? As noted at L29, the addition of HV has improved distinguishing between ice and water, and Canadian analysts have had access to this form of RADARSAT-2 for over a decade. One would imagine limiting to HH-only would be generating additional analyst uncertainty. It would be interesting in a further study to provide one set of images as HH, and another as HV.
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.

We have added the following text to the paper to clarify why only HH was used: "Only the HH band was used for both segmentation and visual interpretation in this study. HH was the only polarization available in samples drawn in RADARSAT-1 imagery, and therefore, was the only band used for both segmentation and visual interpretation. HH was the only polarization used for samples drawn in RADARSAT-2 imagery as well, to be consistent with the polarizations available for all images. 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, L8. The egg code is a World Meteorological Organization international standard. Reference should be made to WMO publication 259, Sea Ice Nomenclature.
Agreed. The WMO publication has been referenced in the paper.

P6 Figure 3
The example provided shows an area containing both dark areas of open water and lighter areas of concentrated sea ice. The open water areas are larger and contiguous enough that most ice analysts would delineate them as separate areas, and some explanation is needed as to why these should be grouped with large areas of concentrated ice. MAGIC looks to have done a good job of marking these open water areas.
The purpose of the study was to assess the agreement in ice concentration estimates of analysts with a polygon. The purpose of the study was not to assess the way that polygons were drawn (although we would like to pursue this in a future study!)

P8 Figure 4. Similar question, why does the polygon contain areas of areas of very open drift ice that are continuous with adjacent areas outside the polygon?
Homogeneity of ice was not a requirement in our study. Our only requirement was defining areas of high contrast between ice and open water to ensure there was little ambiguity in differentiating between them.

P13 L11-P14 L3. 181 disagreements, or almost a third, is a large proportion. MAGIC provided lower estimates in 175 of cases, but analysts tend to err on the conservative side, and if the mix of ice concentrations shown in the example image polygons is representative of the dataset, it is expected that there would be more disagreement, as shown by only 36.8% of examples having unanimous agreement between analysts.
It was stated in the abstract that "true accuracy is expected to be lower than what is found in this study," because we restricted our samples to only those cases where ice and water had high separability. Satellite images often have areas of ambiguity in separating ice from open water, where we expect MAGIC to have greater difficulty. Therefore, we expect true accuracy to be lower than what we found in this study.

P18 L31-32. A 84% at plus/minus one-tenth accuracy shows very good analyst skill, most ice services only work in concentration ranges.
Reply to reviewer: I have now added this information to the paper. "Many chart products by Ice Services report ice concentration in ranges, rather than specific tenths; therefore, 84% shows good analyst skill."

P19 L22-23. Agree, the definition of the polygon is critical. The examples provided appear somewhat ambiguous so some explanation as to why the areas shown in the examples are as they are would be useful.
We used different criteria to delineate sample polygons than the criteria normally used to delineate polygons in regular ice charts. Emphasis was placed on identifying areas with fractional ice cover, since there is little value in evaluating analysts' ability to estimate 0/10 or 10/10 ice concentration. Secondly, polygons were not drawn with the purpose of classifying the entire image; instead, polygons were only drawn in areas with fractional ice cover and high contrast between ice and open water to maximize the automated algorithm's ability to clearly demarcate ice from open water.

P19, L35 – P20, L1. The ice charts are for maritime safety so will err on the cautious side.
Reply to reviewer: Agreed, and the sentence is modified as follows. "This could result in ships being prevented from going into areas that they normally would be able to enter; however, Ice Service charts are produced for maritime safety, and therefore, err on the cautious side."

Technical Corrections

Figure 2. Difficult to view at this size and could be larger.
We have resized the figure to be larger.

Figures 2 and 4. The red and green lines are hard to see and could pose an issue for color-blind readers.
We have changed the colour to purple and yellow.