Reply to Reviewer #1

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.

This paper assesses the accuracy of CIS products and the agreement between analysts in the generation of the CIS ice charts. This type of study is relevant to the field of sea ice mapping since CIS charts can be used as training/validation data for developing automated methods. Thus, it is important to assess the accuracy of these products and understand how they are generated, their strengths and weaknesses.

This paper presents good statistical analyses of the gathered information and is a good start to assess the accuracy of CIS charts. Nonetheless, I would recommend more caution in the discussions section since with more analysts, there could be more of them that disagree with the general consensus than just one out of eight.

We agree. In designing this study, we structured the test such that more analysts can participate and/or more images and polygons could be added to the test dataset, which would help with narrowing the confidence intervals of our estimates.

Does the methodology implemented here reflect how the CIS products are generated? Since CIS offers daily products, they surely refer to previous day analysis to determine ice concentration. This would greatly impact the general agreement between analysts. This should be discussed in section 8.

The methodology used here differs from the method that is used to generate CIS charts. For example, analysts normally have the previous day's chart as a reference. In this study, analysts were asked to estimate ice concentration without access to any other analyst's estimation.

Furthermore, we have added the following text to section 8 of the paper.

"Analysts typically start with the most recent chart when producing a new ice chart. This is done to ensure consistency and continuity between ice charts, and prevent fluctuations and variability in how polygons are drawn, or the information that they contain due to variability in analyst interpretation of SAR. In the past, analysts have carried forward the previous day's ice concentration for a given polygon, unless the analyst estimated what they felt was a significant difference in ice concentration compared to the previous chart. Nowadays, analysts are assigned specific areas to produce charts for over time to ensure better consistency. Therefore, ice concentrations in charts may exhibit higher agreement between analysts' estimates than they do without a reference chart. In a future study, analysts could be given an ice concentration estimate and then asked if they agreed or disagreed with the estimate to see if the result changes based on what the estimate given was."

Another analysis that I should be added to this study, and the authors partially addressed it in section 8, is how the agreements vary with ice types and not only ice concentrations. Ice type is also a significant variable for shipping and weather/climatology modeling. I suggest adding details

in section 3 about the distribution of ice concentrations and ice types for all the different polygons to show how this study reflects the ice conditions seen in Canadian waters.
We propose adding this text to section 3.

"The images used for this study were selectively picked to be areas with well-defined floes with high contrast against the black (water) background. SAR image quality varies from image to image, and even within image. Likewise, the structure of sea ice in Canadian waters can vary greatly, with brash and rubble ice along the East Coast and well-defined floes in the Beaufort Sea. Ice without well-defined spatial structure may not be captured due to the resolution of the sensor. For example, first year ice can appear similar to open water, making it difficult to determine its edges. Brash ice is composed of small pieces of ice (less than 2 meters in diameter) that cannot be resolved at the resolution of the (SAR) sensor. Furthermore, segmentation of sea ice in visually ambiguous conditions (i.e. first year ice during the melt season; brash ice; etc.) by automated algorithms is still sub-optimal. As a result, we did not present analysts with ice conditions that would have been difficult to automatically segment. The sea ice types used in the samples of this study are not representative of all sea ice conditions typically found in Canadian Service Ice Charts. This study quantifies the accuracy of sea ice concentration estimates under the best case scenario of well-defined floes in very clear SAR images. It is expected that accuracy would decrease under brash ice conditions and/or poor image quality."

I think in a future study, the authors should look at how MAGIC compared to CIS charts generated operationally to assess the agreement between both products with a higher statistical significance than just imagery with low contrast. The authors should elaborate a bit more in the discussion section how MAGIC performs in low contrast or different ice conditions and how this could impact results of this study.
We restricted our analysis to only instances with high contrast to measure analyst ability to estimate sea ice concentration as our primary focus, rather than assessing the performance of MAGIC. However, we used the results of the analysis to make some general comments about MAGIC. We agree that comparison against MAGIC for a future study would be beneficial to compare how MAGIC compared to sea ice charts in general, without being validated by any analysts.

P.4. , L10. Add "(floe size)" after "predominant form (F) of ice"
We have changed the document as suggested.

P.16, L33. I imagine that 7/10 is considered 1/10 overestimation in this case? Please clarify.
That's the correct interpretation. The sentence has now been changed to "For example, if a polygon had the modes 5/10 and 6/10, then an ice concentration estimate of 4/10 was considered 1/10 under-estimation away from 5/10 and 7/10 was considered 1/10 over-estimation."