The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC1
https://doi.org/10.5194/tc-2022-8-RC1, 2022

Comment on tc-2022-8

Anonymous Referee #1

Referee comment on "Visual Interpretation of Synthetic Aperture Radar Sea Ice Imagery by Expert and Novice Analysts: An Eye Tracking Study" by Alexandru Gegiuc et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-8-RC1, 2022

After going through the first part of this paper I realized that there were already some fundamental flaws in the experiment. There was a lot of very good and detailed information regarding the contrast between expert and non-expert analysts, as well as differences within these groups. However, the overall objective of this paper and the relevance of these comparisons was not clear in the beginning, nor was it throughout the manuscript.

Regarding the experiment design, in this study (and Karvonen et al 2015) the analysts are only shown a series of satellite images on Irfan view from different snapshots during the winter season. However, this did not include analysts having any of the ancillary information that typically goes into an analysis, such as: knowledge of prior sea ice conditions (large-scale and regional), alternative ground-truth/observations (if used on these days), and prevailing weather conditions. The continuity that analysts have about environmental conditions is the core of the local knowledge that ice services rely for their routine chart production. The use of unusual hardware and software that puts the expert and non-expert analysts in unnatural situations to do an ice chart the way they would normally be done. The expert analysts are having to navigate a new setup that changes the way they habitually put together information for an ice chart analysis. Generally expert analysts have different systems between one another that allow them to be efficient to mentally collate all the information as they generate an analysis. For example, some will spend more time in the beginning reviewing information from the previous day(s) and others will focus on setting up the GIS layout. How this information is transferred between their own understanding and integrated in the analysis can vary greatly between analysts. Additionally, using a graphics viewer, Irfan Viewer, that is not the standard GIS that allows you to use familiar tools that expert analysts would be used to, that includes being able to access multiple sources of information and overlay them on one another, and have the same resolution they're accustomed, introduces a significant error in the design of these comparisons. If this were to be applied to analysts from other ice services we would expect to see a much greater spread. Fig. 7 shows these clear differences between the unnatural analysis using solely the image, and the actual ice chart which has much more information content.

Regarding the issue of using this method to measure uncertainties of human perception, rather than trying to pick out a signal from noisy and subjective data, such as human bias, intercomparison studies between sea ice features in different products are a more accurate assessment of subjectivity and experiments are easier to carry out, especially when there's such a critical need to evaluate multiple skill sets of ice analysts from various international agencies. Studies like the Karvonen et al, 2015 and Cheng et al, 2020, do these types of comparisons on heavily processed datasets, where drawing the delineations between areas by the analyst is not required. Given that reference data is useful for quantifying the ability of automation to capture the variability in sea ice, an independent variable in which to compare a real vs controlled situation is necessary when it comes to human subjectivity. There does not seem to be anything in the paper that supports the assumption that the areas where one tends to focus should correlate with lower confidence, thus higher uncertainty. Please refer to my previous comment above describing how analysts use information. The conclusion that analysts use more cognitive effort in areas where there is more uncertainty is also not convincing given the spread of the expertise included in the small sample size of this preliminary experiment. This same expert analyst from one ice service would not be expected to be as proficient in understanding ice conditions in different areas covered by other ice services due to different environmental conditions, such oceanographic and weather conditions, nor would they have expertise on the regional variability of the ice.

Despite the small sample size and relatively new approach that this proposed methodology could add in understanding uncertainties introduced by ice analysts, the initial outcomes from this case study does not add any value to the work already being done to resolve the issues of subjectivity in ice charts. Though the authors state in conclusions (5.1) this was proof of concept and they recognize a larger sample if needed, this current experiment design is not a reasonable method and complicates the evaluation process further because there are more variables that need to be taken into account regarding the expertise of the user and the amount of information that is available to them. This method would be especially challenging during the melt and summer seasons where the spread is going to significantly vary due to the geophysical limitations with the satellite sensors. Therefore, the continuity of the analyst needs to be taken into account, similar to weather forecasters, and the amount of time it takes for them to understand the situation should not be as significant a factor as how close the analysis is to actual environmental conditions.

Last, this method is not easily feasible, economically or timewise, to use with ice analysts. Though the cost of the eye-tracking software is a factor, the usefulness is more related to the amount of time ice analysts are able to spare outside of operations to provide feedback towards these types of intercomparison studies. This approach is far more cumbersome to implement and open to further interpretation, rather than developing a more scientific metric-based evaluation to analyze uncertainties with subjectivity in ice charts.

The current method does not support the outcome that “the long fixation duration are connected with larger uncertainties in the final ice charts” stated on P27 L5, as there are a number of other factors which can be affecting the analysts decision-making. It is important that these types of studies are being developed so we can understand the

human bias in ice charts and it is great to see these new and innovative approaches. However, the experiment in this methodology needs to be 1) redesigned to allow the analysts to include additional sources of information that they would regularly require for routine ice analysis, as well as 2) putting them in their normal working environment using the common systems that they are familiar with. This will allow them to use all necessary sources of information without compromising the functionality or spatial resolution in which they're familiar and will allow for more appropriate assessments on the subjective nature between expert and non-expert analysts.

Reference: Cheng, A., Casati, B., Tivy, A., Zagon, T., Lemieux, J. F., & Tremblay, L. B. (2020). Accuracy and inter-analyst agreement of visually estimated sea ice concentrations in Canadian Ice Service ice charts using single-polarization RADARSAT-2. The Cryosphere, 14(4), 1289-1310.

The following are specific comments from the first part of the paper:

P2 L9: use of term inconsistencies

P2 L10 Replace "miss-classification" with "misclassification".

P2 L10-12: wouldn't areas that require more cognitive effort be prone to less miss-classification? The following sentence then states that areas less restrictive to navigation are more flawed. Ice analysts spend more time on areas where high traffic areas are known to be, including areas that are more restrictive, as a safety precaution. If areas are less restrictive, ideally they would require less cognitive effort. These sentences contradict one another. Additionally, the combination of both sea ice regimes and level of regulation in a given area for ice charting has significant implications on how analysts focus on the attention to detail in a particular area. Sea ice operations in the Baltic and the Arctic are often confused so this should specify that this paper is focused on the Baltic.

P2 L12-13: What is not being highlighted is that experts are able to map large sea ice covered areas because they have continuity in observing how the ice is changing on a daily or weekly basis. This is very different from someone who understands how to interpret sea ice in SAR imagery and may be looking at it for the first time, without having knowledge of environmental conditions in the area. This statement is incredibly misleading.

P2 L15: What is the purpose of this paper? To use eye-tracking as a metric to calculate uncertainty? If so, this should be stated clearly.

P2 L14 Confusion of terminologies, “open ice” and “very open ice” refer to concentration and not to whether the ice thicknesses are mixed.

P2 L16: What is meant by "large areas?" Does this mean synoptic? If so, up-to-date information is required at meter scale resolution, especially for tactical navigation. For route planning, large-scale information is more useful. Depending on the area, navigators require both but the "typically over large areas" simplifies the needs of maritime users and their data requirement needs.

P2 L24: need to include the challenges that snow cover and melt have on the surface roughness because this is the key challenge in sea ice monitoring and one of the main reasons for ice charts continuing to be fully manual, as opposed to semi-automated.

P2 L26: Sentence needs to be revised.

P2 L31: Is there a metric used in this comparison?

P4 L7-8 MANICE gives only a brief outline of ice charting practices, specific to Canadian Ice Service, and more the type of information content to be found in ice charts.

P4 Table 1: New ice and level ice categories are typically not used in sea ice concentration analysis.

P4 L16-18 Check Zakhvatkina et al 2019 reference is an overview, maybe more just what AARI have been doing?

P5 L1: Omit "Even"

P5 L10: Omit "for long"

P6 L1: Revise, awkward. Suggestion: "The FIS ice analysts have experience with analysing

SAR images for drawing sea ice charts since....."

P6 L5: This does not need to be a separate sentence and Table 3 can just be referenced at the end of sentence from P6 L4.

P6 L7: Does this refer to Table 3 or Figure 3?

Pg6 L10: Specify original resolution

P6 L9-10 Specify the original resolution of RADARSAT-2 ScanSAR Wide. Depending on the processing it can be 100 or 50 m.

P7 L18 “an external monitor with a 22" diagonal size, similar to the ones used in the operational ice charting”. FMI typically uses a Wacom digitizing screen so that the analyst is looking directly at and drawing on the image being processed. Was this set up changed for this experiment?

P7 L27-28 “the SAR images were opened and viewed with an image viewing program (Irfan View)” This is again different from the ArcGIS software used by FIS ice analysts.

Pg10 L10: Replace "fore" with "for a"

P10 L28: Who does "he" refer to? E1 or E2? Probably the use of pronouns should be neutral throughout the paper to maintain neutrality in subjects.