Author Responses to RC2: Comment on tc-2022-8'', Anonymous Referee 2, 19 Apr 2022

Citation: https://doi.org/10.5194/tc-2022-8-RC2

The Author Responses are marked with bold text.

1 General comments by RC2

Review of: Visual Interpretation of Synthetic Aperture Radar Sea Ice Imagery by Expert and Novice Analysts: An Eye Tracking Study. Alexandru Gegiuc et al.

The paper presents the results of an experiment in the use eye tracking technology to "identify elements behind uncertainties typically introduced during the process of sea ice charting using satellite synthetic aperture radar (SAR) imagery" by comparing the efforts of experts to novices.

Unfortunately – almost all of the papers insights into the problem could have been written down at the outset. The paper demonstrates the obvious – that SAR images of sea ice are complex and difficult to interpret particularly in areas that contain the signatures of several different phenomena. And that even expert analysis can take different approaches and produces inconsistent results. In fact the Novice analysis is almost completely superfluous – you could easily delete it from the paper focus on the expert analysis (which you do for pages 13 to 22 anyway) and which are the basis for almost all the conclusions (except that Novices are bad at interpreting SAR sea ice imagery). .

A.R.: We will shorten the novice analysis part but still see it compared to expert analysis as an important piece of information. For example in considering automated SAR image algorithm development it is important to know how the experienced ice analysts perform and to avoid novice behavior.

The expert analysis is used to conclude that "eye movement data in further studies to deepen this kind of knowledge and to understand the uncertainties introduced" [in ice charting]. And while that seems to be a reasonable and noteworthy result- there seems to be no path by which these results can be effectively translated into practice ("link between the observer and the automated method has not yet been established"). Analysts looking at imagery is slow, inefficient and (as the authors report – inconsistent) and is not a viable way to process the ever increasing amount of available SAR sea ice data. It would seem that establishing this link is one of the most important thing to do. (rather than compare expert and novice with the obvious conclusion)

It was frustrating that after reading through 26 pages – with lots of detail on pages 11 to 24 only for the authors decide to tell me on page 26 that the "main findings are qualitative in nature" (so why all the detail) and "more SAR data and more

ice analysts are needed for a comprehensive study". Really - it would have been nice to know all that up front (I would have skipped the details).

A.R.: We will improve the structure in the revised version and try to focus on the most relevant issues.

The paper in fact immediately raised a red flag when I discovered it relied on imagery from 2010 to 2012 – i.e. imagery 10 to 12 years old as its primary focus. The study area in the Baltic Sea has approximately 2000 passes from the Sentinel-1 spacecraft between 2014 and 2022. It is particularly odd to focus on old imagery given the references on page 3 to… "The basis for the Baltic Sea ice charting at FIS is daily SAR mosaic…. The mosaic is updated once per day, typically in the morning, to include most recent available SAR" So why was RS2 and not S1 used for the study?

A.R.: It is not relevant which data sets are used. The RS-2 and S-1 C-band SAR imagery are very similar. For this study some images with different and interesting ice conditions were and easy access selected. They just happened to be older data. We do have a lot of SAR data but there were other reasons limiting the amount of data to be used (limited time of ice analysts and the setup and maintenance of the eye tracking system during the experiments). This has also been a learning process for us and we hope that we'll be able to set up a more sophisticated eye tracking experiment connected to the actual ice charting process to collect more representative data sets. However, this will require significant additional resources (human resources and funding for hardware and salaries).

Please go back and decide what paper you want to write Novice v Expert SAR Ice Analysis (nothing to see here / expected results) or What Can the Eye Tracking of Experts Tell Us about How to Improve Sea Ice Charting (potentially really useful) – in particular please tie it to a potential way to quantify error in charting or improve ice classification. But certainly don't spend pages on details and then tell me at the end they don't matter. (and that you need a bigger study to get any useful answers)

A.R.: The main idea is related to collecting information on how the ice analysis is performed by an experienced ice analyst and to gain information to possibly improve ice charting and automated sea ice SAR analysis. We'll also try to improve the comparison between novices and experienced ice analyst to identify where the experienced analysts behave differently. This information is very valuable information for developing automated algorithms and also to be able to guide the novices to perform better.

2 Specific comments by RC2:

Specific Issues:

A.R.: We will address the specific issues in the revised version of the manuscript.

Abstract: "We also show that the experts are able to correctly map large sea ice covered areas only by looking at the SAR images" I am not quite sure how they reached this conclusion give that the Balitc Sea is a rather small mostly enclosed marginal sea.

Section 2.1 “We used five RADARSAT-2 (RS-2) ScanSAR Wide images covering different regions of Baltic sea across three different winter seasons”. Why was it important to use 3 seasons – all imagery from February. The sea ice conditions might or might not be similar.. (or was that the point?)

“In FIS, the original SAR images are typically reduced in size for easier manipulation and saving disk space. This reduces the amount of detail available for analysis. The 100 m resolution of the SAR imagery used by the FIS analysts is lower than the original resolution. Here we used the same down-scaled resolution for the RS-2 SAR images” This is one of those paragraphs that make me wonder if this paper has been sitting on a shelf for a decade and someone just dusted it off. Are the authors really having disk space limitation issues? A ScanSAr tiff is about 300 MB. They never specify the original resolution (its 50 m for ScanSAR – meaning at most you saved at most factor of 4 by moving to 100 m). What other preparation of the imagery were undertaken? Are they calibrated to produce Normalized Radar Cross Section so that the SAR signatures across years can be properly inter-compared or was it just the digital number?

Section 2.2 “Two novices (N1 and N2) with little or no familiarity with classification of SAR sea ice imagery participated to the study” and Section 3.1 “We instructed the participants to look at the selected images and interpret the content verbally....When looking at the SAR images, the participants had the task to describe their content freely by identifying sea ice types and features and classify them as they would in a typical ice charting routine”

So this would appear to be contradictory – The authors expected the Novices to verbally describe sea ice types and features – when they had “little or no familiarity with classification of SAR sea ice”

Section 3.1: “while the SAR images were opened and viewed with an image viewing program (Irfan View) which allowed users to freely change the scale or pan the viewed images.” Why this program – were the participants familiar with it? Was the SAR imagery presented in isolation? I.E was there a map reference for location? Reference scale for NRCS gray scale values or a map scale to indicate the size of the image and features?

Figure 3: Why are the land areas presented as both white and black. Is this display representative of how the images were displayed to the participants – without scales or map references?

Section 3.3 “We divided the gaze data into segments that correspond to the scanning phase and the analysis phase” Is this a standard practice when performing eye tracking work?

Section 3.3 “We computed the average dwell time, fixation duration mean, standard deviation, and fixation density (number of fixations per ice area).” Is this standard reporting metrics when performing eye tracking work? If so what kinds of information do these metrics convey?

Section 4. The title is Visual Interpretation of Synthetic Aperture Radar Sea Ice Imagery by Expert and Novice Analysts. Yet section 4 is devoted to the experts and there appears to be no corresponding section for the Novices. As stated above – the authors could delete the Novice analysis entirely and not lose much)

Section 4.1 “The first difference between experts (E) and novices (N) was noticed right away, based on their reactions when they were shown the first SAR image.” ... and the first difference is associated with.....? They never explicitly call out a second (subsequent) difference..

Section 4.1 “Even if novices recognized fast (in the first five seconds) that the image shown is a SAR image” Even if!!!? Did either novice need more than 5 seconds to determine they were looking at a SAR image given that the other images were of easy to recognize natural images (human face, a flower, a fish, a cat and a bird) displayed on entirely different system (Tobii Studio vs Irfan View) - Seriously

Section 4.2.2: “This is an interesting result, showing that the experts need only few seconds (5-14) of fixation time to be able to classify an ice area. Also it underlines a difference in style of analysis. E2 spent in average about five times longer analyzing an image and about three times longer analyzing an ice area than E1. These differences could be explained by the inequality between the two experts in terms of years of experience and of training.” I’m not sure its really all that interesting. Is there an alternative explanation? I don’t not believe that the years of experience (10 v 25) can account for that amount of difference. After 10 years of experience a good analyst has seen pretty much everything. More likely this difference is explained by “personal styles of analysis” and not expertise.

Section 5: Limitations and Future work “The aim of this study was to act as a proof of concept study… The sample number in our study is low, and thus, the main findings are qualitative in nature..” All of this should have been stated at the beginning of the paper. Citation: https://doi.org/10.5194/tc-2022-8-RC2