Our comments

Editor:

Dear Dr. Hansen and co-authors,

I have received two reviews based on your revised version of the manuscript "Estimation of degree of sea ice ridging in the Bay of Bothnia based on geolocated photon heights from ICESat-2." Both reviewers agree that manuscript has been substantially improved but still recommend a few more minor revisions. Reviewer #1 has several specific comments outlined in their referee report. Reviewer #2 still has a concern about the FIS DIR validation dataset and has provided several suggestions to the authors for alleviating this concern.

Please submit a revised version of your manuscript addressing the specific concerns of Reviewer 1 and validation concern raised by Reviewer 2.

Thank you for submitting your work to The Cryosphere.

Steve

We wish to thank both of the reviewers for their sharp eyes and invaluable comments, which have improved the manuscript. We have taken their comments into consideration, and hope that the Editor finds our revisions satisfactory.

In particular, we have:

- **Updated the figures to alleviate the concerns of both Reviewer#1 and #2, which includes updated figures on the distributions (correcting the units), enlarging the SAR intercomparison figures, and improving the photon profiles as proposed by Reviewer report#1. Furthermore, we have amended the color scales according to the comments of Reviewer report#2.**
- **Corrected the text where new edits are proposed by the reviewers.**
- **Small corrections in the text for clarification or consistency (e.g. writing data as plural instead of singular form, as given by the guidelines).**

Sincerely, on behalf on the co-authors,
Renée Mie Fredensborg Hansen

Referee report #1 (Reviewer #2):

General comments
The manuscript has been greatly improved. Comparison with SAR imagery increases confidence in the value of IS2 measurements. Analysis of uncertainties and error sources is detailed. Nevertheless, feasibility of using FIS DIR for training the classification algorithm is still questionable.

We thank the reviewer for the positive feedback, and for the detailed review, which have greatly improved the manuscript.

I'll try to explain my concern from a different angle. In Section 2.4 you mention that clear separation of elevation anomalies (EA) into three groups appears only when you use P90 and you even use only 5% of anomalies with the highest values for computing the thresholds. But it means that only 5% of the classified IS2 DIR correctly correspond to FIS DIR. The rest, 95% of EA cannot be classified in accordance with FIS DIR by definition.

To illustrate I use your plot with histograms of EA separated by DIR level. I draw a sum of histograms as a thick black line – that's a distribution of a priori unknown EA. It is not accurate as I draw it by hands, but it gives a realistic representation. I add dashed lines at location of your thresholds. I mark with green, blue, red patches the sections of EA histograms that are correctly classified. If you compare the area of these patches with the rest of the histogram, you can also see that it constitutes approximately 5%. The rest of EA (area under the thick black curve) are classified incorrectly or not classified.

The fact that less than 5% IS2 DIR correspond to FIS DIR must be clearly presented in the manuscript.

We thank the reviewer for this in-depth description of their concern. Indeed, it means that less than 5% IS2 DIR corresponds to FIS DIR. We must also point out that it is not expected that the entire distribution would overall be very different, since ridges (elevation anomalies) are expected to be in the higher end of the distribution, and the distribution also includes elevation anomalies covering level ice (which will be the more dominant ice type). We will clearly state this in the manuscript for clarification.

We have included the following sentence in the end of Section 2.4:
"We note, that the intervals are based on the 5% highest elevation anomalies (using 95th percentile data), as we assume the highest elevation anomalies will include information on the ridges. Were one to use all of the elevation anomalies, it would also include elevation anomalies from the level ice (as seen by the overlapping distributions in Fig.3a)."

In my opinion we can confidently conclude that the available FIS DIR cannot be used for training or testing of DIR classification from IS2 data. Therefore, either the title should be changed as proposed earlier to "Comparison of IS2 elevation anomalies with FIS DIR…", or a clear statement that more data (e.g. more FIS DIR observations or quantitative SAR data) is needed for developing a reliable DIR level classification algorithm should be added to conclusions.

We are happy to include a clear statement in the conclusions that this is required. We have included the following sentence in the conclusions:

"We note, that this methodology of estimating IS2 DIR is based on the 5% highest elevation anomalies, and that while the IS2 DIR follows the deformation information seen in the SAR frames (qualitative assessment), the IS2 DIR only follows FIS DIR for some cases. To develop a reliable DIR level classification algorithm from IS2, more FIS DIR observations or quantitative SAR data should be included. "

Detailed comments

Line 180. Higher reliability of the classification is not proven. Only the classification of the 5% of points is more reliable. Classification of the rest of the data (extrapolation ability) may become less reliable when you decrease the amount of data. Either remove "reliable" or prove that it is more reliable by splitting the training dataset in two random parts, training the classifier on one part of data and applying it to another part of data.

The following has been removed: "..., ~~and thus a threshold-based classification is more reliable~~".

Figure 6.
The intention behind the figure is very good but it needs to be improved.
SAR images are too small for visual analysis, profile plots are too busy to distinguish anything. Please enlarge SAR images and reduce size of profiles and remove excessive data from profiles (e.g. present only one beam on a plot) and decrease the range of Y-axis down to [-0.5, 1] in order to stretch the profiles vertically. Titles on planes f and h should be corrected to 27 March 2019.

We do not believe that only one photon/elevation anomaly profile of one beam will be sufficient, when you are looking at all three beams on the map, and since they are separated by ~3km, they are observing different surfaces. We have changed the set-up of the figures, to enlarge the SAR images (4 subplots in one, new Figure 6), and added a new figure (new Figure 7) on the photon/elevation anomaly profiles, with 3 subplots per photon profile (P1-P4) to include each individual subbeam. We have also combined the HH polarization figures (Figure 7 and 8) into one figure (new Figure 8). We shall shrink the Y-axis a little, but will keep it at [-0.5, 1.5] since we also aim to look at the photon profiles used to generate the elevation anomalies and they extend from -1.5 to 3 m (with 3 of the 4 photon profiles exceeding the Y-axis suggested by the reviewer).

Line 470. The argument that "IS2 seems to carry similar information as the S1 images, and the FIS DIR are based partly on icebreaker observations and partly on SAR imagery" is obviously weak and cannot be used to justify usage of FIS DIRs to train or validate algorithms. As mentioned above, only 5% of the classified anomalies correspond to FIS DIR, making it unusable.

We thank the reviewer for this comment. It is however not uncommon to train algorithms with ice charts (e.g., Gegiuc et al. 2018), so usage of FIS DIR to train with has been done before -- albeit not on IS2. We have revised this sentence, to clarify that more data is required (or potentially use other data sources such as Sentinel-1), to allow for a more robust classification in the future.

The sentence now says:
“However, since the classification of this study is based only on the 5% highest elevation anomalies, either more FIS DIR data or quantitative S1 data should be included and used for training, to ensure a robust classification. “

Referee report #2 (Reviewer#1):

1 General Comments

1.1 Synopsis of Changes

The authors have introduced some more quantitative aspects to their analysis and have evaluated against SAR data, which looks to me to be well done but this should be corroborated by Reviewer #2. New sections have been introduced on ridge anisotropy and cloud-cover in response to my previous comments, which I find to be satisfactory. However several small issues have been introduced within the figures. As with my first review, I recommend this manuscript for publication pending the minor modifications suggested below.

We thank the reviewer for the positive feedback. Indeed, some small issues has been introduced in the figures, hence based on these comments of this review (and the comments of referee report#1), we have amended the figures accordingly.

2 Specific Comments

2.1 Figures

Fig. 3: PDFs have units that are the inverse of the x axis, such that the area under the curves equal to one and are dimensionless. So the y axis shouldn't be % if the x axis is in m.

Indeed! We thank the reviewer for their sharp eye and for making us aware of this mistake. This has been corrected.

Figure 5: Colorbar ticks should be coincident with color transitions - this issue was present in the originally submitted manuscript but I didn't pick it up. This should be xed prior to publication.

Thank you, we have aligned the colorbar ticks with the color transitions. We have also amended Table 3 and the text to fit with the values now aligned with the color transitions.

Figure 10: The lon/lat ticks here are so small as to be almost unreadable. Please enlarge.

Thank you, we have now enlarged it.

Figure 4: Not a problem that `up' on your maps aren't orientated North, but it is fairly standard to add a small annotation in the corner to indicate to the reader the North direction (when it is not upward). This is particularly important when your grid lines are faint and other maps (e.g. Fig 10) are North-orientated. Apologies again for not picking this up on the first review.

We thank the reviewer for this comment. We have included a north arrow (either inside the map or just beside the maps. Furthermore, we have made sure that all the

maps with IS2/FIS and IS2/SAR frames use the same projection (Transverse Mercator) and that the latitude/longitude labels are larger now. The gridlines are still faint (since we do not want them to take up too much of the figure, especially over the SAR frames), but we have made the SAR figures larger, so that the cyan-colored gridlines should be easier to distinguish now as well.

Figures 2, 6, 7, 8: Much has been written about rainbow color maps (such as gist-rainbow used here) in the earth sciences, and they should be avoided (e.g. Borland and Taylor, 2007; Crameri et al., 2020). They have the effect of implying sharp transitions in the data where they do not exist, and are difficult to read for those who are colorblind (0.5% of women and 8% of men worldwide). Many good alternatives are available, see Light and Bartlein (2004); Stauer et al. (2015); Thyng et al. (2016)

Yes, indeed. We actually discussed this in the beginning of writing the manuscript. However, it is crucial that we use a color scale with as many color transitions as possible to easily discriminate between differences in the elevation anomalies. Which was the rationale for picking this colormap at first. We thank the reviewer for providing us with links to the various studies that have looked into such alternatives, and have decided to use 'magma' instead based on the studies the reviewer suggested. We hope this is satisfactory. Unfortunately, when using the sequential colormaps (and not miscellaneous), there are fewer options available with a large range of colours. While the 'magma' colorbar seems to have enough variety to show the differences of the elevation anomalies, it however also means that some of the areas with low elevation anomalies are almost black, which in turn makes it difficult to see them on the SAR images. We have highlighted areas where it is important to notice these low elevation anomalies (e.g., Figure 2, region 2A). However, the focus of our study is mostly on the higher elevation anomalies (the bright values) which are easy to distinguish using the 'magma' colormap.

2.2 Text

L137: You need units on your plus/minus figure (metres).

Thank you, we have now included this.

L211: Does the thicker snow on fast-ice smooth the surface more? It's not obvious to me that snow bedforms such as dunes and sastrugi would produce a smoother surface than a relatively level, bare FYI, particularly at the radar-wavelength scale. I think you should cite this or instead of `This is expected as' go for `we attribute this to...'.

We have rewritten the sentence to: "We attribute this to regions of fast ice primarily consisting of smooth level ice represented by small differences in the elevation anomalies, and the fact that level ice overall carries more snow than drift ice, which could smooth the surface even further."

L508: It's true that estimating ice type (FYI/MYI) is tricky, I think you're talking more generally about ice properties. Perhaps change type -> properties.

Thank you, this has been changed accordingly.

L509: Change cm to centimeter

Thank you, this has been corrected.

L510: It's a nice point that cm-scale roughness can confound radar estimates of large-scale roughness, but it should be cited.

The following references has been cited:

Manninen, A.: Multiscale Surface Roughness and Backscattering, Progress In Electromagnetics Research, 16, 175–203, https://doi.org/10.2528/PIER96060700, 1997.

L498: Should be near-real-time

Thank you, this has been corrected.

References

Borland, D. and Taylor, R. M.: Rainbow color map (still) considered harmful, IEEE Computer Graphics and Applications, 27, 14{17, https://doi.org/10.1109/MCG.2007.323435, 2007.

Crameri, F., Shephard, G. E., and Heron, P. J.: The misuse of colour in science communication, Nature Communications, 11, 1{10, https://doi.org/10.1038/s41467-020-19160-7, https://doi.org/10.1038/s41467-020-19160-7, 2020.

Light, A. and Bartlein, P. J.: The end of the rainbow? color schemes for improved data graphics, Eos, 85, https://doi.org/10.1029/2004EO400002, 2004.

Stauer, R., Mayr, G. J., Dabernig, M., and Zeileis, A.: Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations, Bulletin of the American Meteorological Society, 96, 203{216, https://doi.org/10.1175/BAMS-D-13-00155.1, 2015.

Thyng, K. M., Greene, C. A., Hetland, R. D., Zimmerle, H. M., and DiMarco, S. F.: True colors of oceanography: Guidelines for effective and accurate colormap selection, https://doi.org/10.5670/oceanog.2016.66,2016.

References

Gegiuc, A., Similä, M., Karvonen, J., Lensu, M., Mäkynen, M., and Vainio, J.: Estimation of degree of sea ice ridging based on dual-polarized C-band SAR data, The Cryosphere, 12, 343–364, https://doi.org/10.5194/tc-12-343-2018, 2018.