Response to the reviewer 2 comments on the manuscript TC-2021-185,

Thank You for reviewing our manuscript and for the constructive comments, We have tried to take all the reviewer comments into account where applicable and updated the manuscript accordingly. In the following are our detailed response to the comments (in blue).

As two reviewers suggest, we decided to use the baseline-D data and recalculate everything. We'll also provide two versions of the algorithm: one trained to be compatible with the modeled SIT and another with the CS-2 SIT. Recomputation of everything and updating all the tables and remaking most of the figures will require a significant amount of work and time, so we'll request the editors enough time to perform this work in addition to our other duties.

The work presents a data fusion example, to combine CryoSat-2 altimeter data with Sentinel-1 SAR image data, for extending sea ice thickness estimates in the Barents and Kara seas. Overall it is very well written, logical, and well supported, and of interest to the Cryosphere community. I have listed many minor comments and inconsistencies that I think would make the paper easier to read, and one major comment regarding the main scientic assumption of the work. However, I only ask for more explanation about the assumptions and limitations going into this work. I agree with the authors that this will often produce valuable results. I will recommend major revisions, mostly because of the number of inconsistencies listed.

Major Comment:

The main scientific mechanism is assuming a relation between radar brightness and sea ice thickness. That is, inferring that similar sea ice surface properties (which a SAR is sensitive too) corresponds to the same thickness. Although this may often be true, it is not necessarily true, unambiguously, and you probably cannot tell when from the SAR alone. I think the sections explaining this may need some expansion and clarification. Consider:

• Page 6, Line 171: This sentence implies a direct relationship between level ice thickness and brightness, and that it is consistent with ice from different growth regions (i.e. land-fast ice being connected to level drifting ice).

This means that the thickness is mapped from similar reference areas. We think in addition to just brightness the ice analysts also use image texture as a measure of similarity, even though not explicitly mentioned here. Also other satellite data, such as microwave radiometer data, than just SAR are used by the ice analysts. The ice analysts also utilize their earlier knowledge of what certain kind of ice (with some typical thickness values or ranges of values) looks like in satellite imagery. We have updated the text to take this into account.

• I thought is was well established that the radar brightness saturates rapidly for level ice greater than about 30-50cm. This means that SAR does not really "see" level sea ice thickness over a certain thickness.

It is true that C-band SAR does not penetrate very deep into sea ice. Ice thickness estimates can only be derived indirectly based on the surface backscattering. One typical assumption is that thicker and older sea ice has gone through more deformations has a rougher surface affecting to the backscattering. Naturally this assumption does not always hold. However, here we only try to locate similar segments within a restricted distance and time and interpolate segment SIT based on the SIT of the spatially and temporally neighboring similar segments.

• Level ice brightness is also produced by different (wavelength scale) surface roughness properties, which may be different for different growth areas, different snow packs, salinity, different moisture histories and many things.

Yes, and the ice analysts also utilize the local variations (texture) in their analysis and base their classification on their experience.

• Brightness gets much more complicated for deformed ice, with both wavelength scale and large-scale roughness affecting backscatter. And rougher, does not necessarily mean thicker, although it often means brighter.

As mentioned in the response to the previous item, backscattering magnitude is not the only thing ice analysts use for the analysis. They include information from different sources and their experience. They also know the history, both short term (from hours to a few days) and long term (previous years) and thus know the ice situation during the previous days in the area and also the typical ice conditions during the earlier winters. All this knowledge and experience are used in the ice charting process.

• Did the fast-ice equivalent thicknesses data include deformation (e.g. ridges and rubble), or was it only level fast-ice regions?

See the response to the following item.

• Is this land-fast ice regression only used in the AARI ice charts, or is this also used in your work?

It is used in AARI ice charts, but it should be emphasized that it is only one methodological part of the ice analysis. And it is not so well-documented. We know the process of making ice charts at FMI and they base their analysis on satellite data from multiple satellite data sources. They make the image analysis based on the brightness and the texture, using reference ice thickness data measured recently at some points. They also use their experience from the past and know well what different types of ice in the imagery look like in different areas of the sea area to be analyzed. The process of ice charting is similar everywhere and using reference SIT of similar-looking image areas is just one method used in the analysis.

• I understand that the overlapping CryoSat-2 observations are used to estimate thickness for some of the 15 segments from the mean-shift algorithm.

The meanshift algorithm provides 15 segment categories, but the number of segments for one SAR mosaic is much more. The segments just belong to one of these categories (indicating that they have similar average sigma0 for both HH and HV channels). We have tried to make the text more clear to better indicate this.

• You don't show a figure of these thickness versus segment properties, for the overlapping categories, to assess accuracy. Perhaps it is difficult, since you have 15 features. But some sort of regression or least squares fit line, in some dimensions may be meaningful to portray the fitting accuracy.

Actually we fit the differences of the texture features and the SIT difference. We have now included a figure of the linear combination of feature differences vs. SIT difference. There is quite large deviation but the linear combination is better than single feature differences. The correlation between the linear combination of feature differences and SIT difference is about 0.5. For single feature differences it is less than 0.3 for each.

• Your T-distance weighted combination to get SIT for non-represented categories may work for interpolation, but how does it extrapolate? Or, is it effectively a multivariate regression?

I am not saying that this cannot be done, as it may be better than nothing, but please make some small comment about this assumption or limitations. This is your main mechanism for extrapolating the SIT measurements away from your CryoSat-2 tracks.

We used the term extrapolation from a “geographical” point of view, meaning that the segments which are not geographically between any segments with CS-2 SIT assigned are also assigned a SIT value. Actually this is also rather interpolation and we have dropped “extrapolation” from the manuscript and use only “interpolation” which is technically more correct.

Minor Comments:

1. The second sentence of the Abstract might be better at the end of the Abstract, maybe the second last sentence, just before "Our results are directly applicable...". Then the Abstract follows the more logical layout of the paper, and the result comes after the lines about what you actually did and the comparisons. The Abstract is not so long that this delay should cause any loss of emphasis.

We have moved the sentence as suggested.

2. Section 2, page 3, line 81: It may be interesting to confirm that the "two years" of data, "2016 and 2017", are in fact calendar years, and not two whole seasons. This is, in fact, one whole season and two half seasons. Does this influence any of your statistics?

Yes, 2016 and 2017 calendar year were used in this study. As one Jan-April and one Oct-Dec period were used for both training and testing data sets we don't think there is significant difference compared to using two seasons. In any case this just a preliminary study and for to develop a reliable operational SIT estimation system training data over several seasons would be preferable.

3. Page 3, Line 89: the word "figure" is missing at the end of the line. I think it should be "figure 3", and not 2, as well.

Corrected.

4. Page 4, Line 95: "not be used classify..." should be "not be used to classify..:".

Corrected.

5. Page 5, Line129: It says "...by known snow depth...". How do you know, or where do these values come from? I guess they come from the Warren climatology paper that is mentioned in the next paragraph. Instead of saying "known" at this early sentence, could you perhaps say "climatological", or, say "climatologically derived" or something similar. That will connect better to the later sentence with reference.

Yes, changed this to "assuming known" and mentioning "climatologically derived".

6. Page 5, Equation (1): Symbols for h_i and h_s are not defined. Although I can guess them, you should really define them for completeness.

Explanations added.

7. Page 9, Line 254: The abbreviation "MS" at the end of the line is not yet defined. Please add it to the first instance of using the MeanShift algorithm, if that is what it represents.

Corrected.

8. Page 9, Lines 264-266: May want to include "interpolation" somewhere here, to connect it to words in your Abstract.

We have replaced one "assigned" with "interpolated" in the sentence.

9. Page 9 and 10: Feature List.
• You may want to remind readers that the two backscatter coefficients are in decibels [dB].
• The symbols for coefficient of variation F_v is written as C_v later in the text, Page 11, Equation (7).
• The ordering of many features in this list is different from the order they appear later in the text. It would help to be more consistent.

Corrected and re-ordered the list.

10. Page 10, Equation (3): Why is there a square on the log term? I thought Entropy was essentially p.log(p).

Two does not stand for square here in indicates two-based logarithm, changed to subscript.

11. Page 11, Lines 315-318: Could you summarise what these "edge and corners" actually are here? So that we do not have to read the refs. to find out.

We have added a short explanation.

12. Page 11, Line 323: It looks like the figures should be 4 and 5, and not 3 and 4.

Figure referencing corrected.

13. Page 11, Line 325: Does the L1 have a more expanded name or reference? I know its fairly well known in mathematics, but this audience may appreciate just knowing that it is a mathematical vector measure.

Added that L1 difference means the absolute difference.

14. Page 11 and 12, Lines 330-345:
• It may be interesting to emphasise that this difference function T is a weighted sum of absolute differences, rather than a sum of squared errors, which is more common.
• Could you have used some sort of standard distance functions?
• Presumably these weights c_t; c_d; c_k, are to account for different scaling of each feature. Do you also include any priority or enhancement too? Could you have simply normalised all features beforehand and then used a standard Euclidian distance or something?
• You calculate the coefficients with a least-squares t, yet the values are absolute not squared.
• I understand that you will want to capture spatial, temporal, and spectral differences all together. Could you explain or motivate this choice of function better, as it is quite unusual.

The selection was based on the desire to want to include the segment similarity and the distance in space and time (in possible different conditions further away the assumption of similar segments containing similar sea ice is weaker than closer in space and time). Absolute difference was selected here mainly because it is computationally faster than squared (Euclidean distance). Also it is possible to perform a linear fit of sum of each individual difference, using Euclidean distance the fit should have been done for the squares of the differences and the weightting would have been different, intuitively it looks more logical to apply a linear fit to sum of differences than to sum of squares of diffrences.
The weights were defined based on the training data (2016 data) using non-negative linear LSQ regression. This way the weights correspond to the importance of each feature for explaining the SIT difference in the training data set. The weights also take into account the scale. Using scaled feature differences e.g. by scaling dF to dF'=(dF-dFmin)/(dFmax-dFmin) will not change the results.

15. Page 16, Line 494: "sqare kilometers" should be "square kilometers", and there should be a comma after Also at the beginning of the next sentence.

Corrected.

16. Page 17, Line 531: "metl-down phase" should be "melt-down phase".

Corrected.

17. Page 18, Line 538-539: Lots of commas missing here, perhaps five of them.

Checked and tried to correct this.

18. Page 27, Figure 1: Is there a systematic bias between ascending and descending

tracks. It looks like one set of tracks is mostly red coloured, while the other set are more yellow coloured. Or, is there some other factor like time, due to drift, involved? How is such bias handled in the processing, and do you mention this visible observation?

This is something we did not spot before. There should be no bias between the ascending and descending tracks, but looking at the figure it does look like there is. We shall look closer into this and if there is a real bias present, it shall be mentioned in the revised manuscript. We'll also use newer CS-2 data in the revised manuscript.

19. Page 29, Figure 4: Are these mosaics after incidence angle correction or before? There still seems to be brightness differences and edges visible. Please explain.

They are after the incidence angle correction. There always will exist some differences between SAR boundaries, in the open water areas these mainly depend on the wind/wave conditions. For sea ice the incidence angle correction is statistical and actually slightly varies for different ice types and also depending on the surface geometry w.r.t. the observing instrument.

20. Page 31, Figure 6: The pink boxes do not fit around the text on the PDF (or the text does not fit inside the boxes). Make sure the font sizes work in the final proofs.

Figure updated.

21. Page 31, Figure 7: It is quite natural to picture the 1:1 ideal line in such a figure, and then all the curves look quite poor. Admittedly, this could equally be reflecting a bias in the Model SIT values. At least the red, and two blue lines are pretty consistent. Could you please comment on this?

The mapping is based on the PIOMAS SIT because for thin ice ORAS5 does not provide reasonable SIT (only above 50 cm). This figure is to indicate how the mapping is related to CS-2 SIT and model SIT. It is evident that the model SIT and CS-2 SIT do not have very good correspondence but it is difficult to say which one is closer to the correct SIT, i.e.. they may represent different things, e.g. CS-2 measuring the maximum SIT and models giving averages over larger grid cells. In the revised manuscript we'll also provide a version which is better fit (trained using the training data set) to the CS-2 SIT.

22. Page 33, Figure 10: I notice that this figure, and the caption, only talks about values of the order of 50-100 cm and yet the colour scale goes up to 600 cm. I suppose that you deliberately made all colour scales cover the full range for more consistent interpretation between figures. I now found this mentioned in the main text at the top of page 16. However, it might be nice to mention this in the caption of Figure 10 too.

Now this figure has been updated with a different color scale (based on a comment of reviewer 1) and the colorbar has also been updated accordingly.

23. While finding the last comment, I suspect that the figure numbers are incorrect in many places, since you inserted a new figure 3. Please re-check all references to figures to make sure that the numbers are correct.

The figure numbers have been checked and updated.

Thank You!
Authors of the manuscript