General comments

The paper is interesting and introduces an empirical algorithm for detecting and monitoring the ice slab regions across the entire Greenland by using Sentinel-1 products. The text is well written and easy to understand. It opens by introducing the scientific problem, the physical mechanism upon which rely the algorithm (interaction between electromagnetic waves and the ice sheet), the regions excluded from the analysis to limit the noise in the problem, and the algorithm itself, along with the methodology for setting up the thresholds. Then it keep on with a suitable description of the results along with a fair discussion about these achievements and the uncertainties in the process. A comparison with previous mapping done with SMAP is also provided. I haven't found any major issue for the paper publication however I feel that some points should be improved before to proceed.

Thank you for taking the time to review and provide constructive comments on our manuscript!

Specific comments

[R1] *The description of the "ten-fold cross-validation scheme" is quite convoluted and not straightforward to understand. It has to be improved.*

We have thoroughly revised Section 3.3 in the manuscript in response to this comment and suggestions from other reviewers. The new text is provided below in the blue font, along with an updated Figure 4.

Section 3.3 Threshold Optimization and Uncertainty Analysis

Sections 3.1 and 3.2 describe how we form σ^0_{xpol} and σ^0_{HV} mosaics over high-melt regions where ice slab formation might be possible. To then map ice slab extent, we need to choose backscatter thresholds that can delineate regions with ice slabs from regions without ice slabs. We also wish to assess uncertainty by quantifying the range of plausible S-1 inferred ice slab extents that would be consistent with the OIB airborne ice-penetrating radar observations. We approach this problem in two steps. First, we use all available OIB ice slab detections to find the optimal backscatter thresholds that produce the best ice-sheet-wide agreement between the S-1 inferred ice slab extent and the OIB ice slab extent. By applying these optimal thresholds to the backscatter mosaics, we produce a map of the most likely ice slab extent across the ice sheet. Then, to assess uncertainty, we use a 10-fold cross validation scheme where we generate 10 new sets of thresholds, each optimized using only a small subset of the OIB data. From the results of these ten trials, we use the backscatter thresholds that produce the largest total ice slab area to define the maximum plausible ice slab extent, and the thresholds that produce smallest total ice slab area to define the minimum plausible ice slab extent. Together, this quantifies the range of plausible S-1 inferred ice slab extents that are still a good fit to the OIB observations. Below, we describe how we optimize these thresholds in detail.

3.3.1 Most Likely Ice Slab Extent

We use a training data set built from the Jullien et al. (2023) high-end estimate of ice slab extent derived from OIB flight lines surveyed in March-May 2017. (This high-end estimate corresponds to the maximum likely refrozen ice content given the observed ice-penetrating radar signal strength.) For each flight line that passes through an ice slab area, we extract the portion of the flight line that overflies the ice slabs, as well as an additional 50 km buffer that extends inland of the upper limit of the ice slabs. We discretize these lines into points every 50 m and assign each point a value of 1 if an ice slab was detected in the OIB data at that location or 0 if no ice slab was detected. These observations are then used to optimize the backscatter thresholds.

We use a brute force search to find optimal values of α and β that maximize the agreement between the upper elevation limit of the ice slabs as detected by airborne ice-penetrating radar, and the upper limit of the ice slabs as estimated by S-1. Areas where $\sigma^0_{HV} < \alpha$ and $\sigma^0_{xpol} < \beta$ are taken to be ice slabs. We then test all combinations of thresholds for -7.12 dB $< \beta <$ -2.37 dB and -13.6 dB $< \alpha <$-2.1 dB, calculate the F1 score for each combination, and choose the threshold values that give the highest F1 score. The F1 score is a measure of the accuracy of a binary classification and is calculated following Equation (1).

$$F1 = \frac{2 * true\ positive}{2 * true\ positive + false\ positive + false\ negative}$$

Figure 3 shows this optimization trade space with the optimal threshold combination shown in the white dot. We find that using both σ^0_{xpol} and σ^0_{HV} thresholds together leads to modestly better agreement with the OIB detections, compared to using only σ^0_{xpol}. When only σ^0_{xpol} is used to delineate the upper elevation limit of the ice slabs, this F1 score is 0.787, compared to an F1 score of 0.811 when both backscatter thresholds are used. When using only σ^0_{HV} to delineate the upper elevation limit of the ice slabs, the F1 score is only 0.674, so it is clear that σ^0_{xpol} provides additional information that significantly improves the delineation of the upper boundary.
Initial analysis of the backscatter mosaics suggests that σ^0_{xpol} does not display any distinct change in behavior associated with the lower boundary (see Figure 1a), so we optimize a separate threshold, $\sigma^0_{HV} > \phi$, to delineate the lower elevation limit of the ice slabs. We optimize ϕ following the same method as described above, but using a new version of the OIB training dataset that covers the ice slab region and a 50 km buffer down-flow into the ablation zone. Altogether, the area defined by $\sigma^0_{xpol} < \beta$ and $\phi < \sigma^0_{HV} < \alpha$ is our most likely estimate of the spatial extent of ice slabs across the ice sheet.

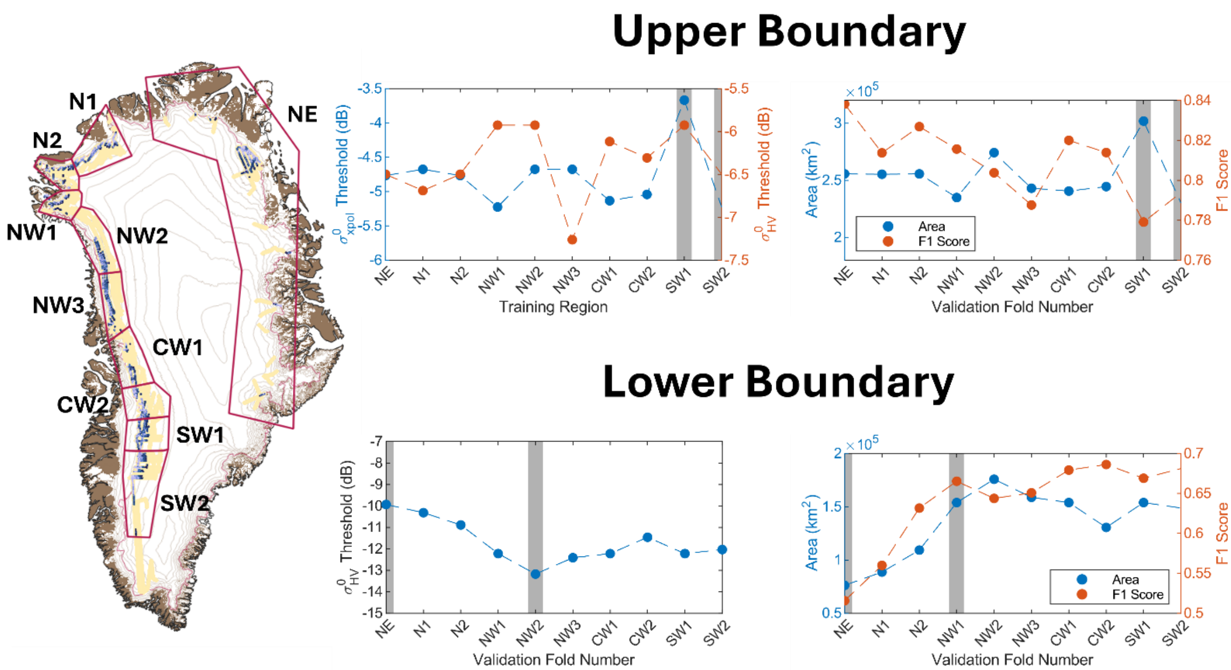
3.3.2 Maximum and Minimum Ice Slab Extent

To quantify uncertainty in this most likely estimate of ice slab extent, we use a 10-fold cross-validation scheme. We divide our training dataset into 10 subsets, each containing OIB ice slab detections from a different region of the ice sheet (see Figure 4). For each of the ten regions, we use a brute force search to find the values of α, β, and ϕ that produce the best agreement between OIB ice slab detections and S-1 inferred ice slab extent in that region. We then apply those local thresholds to the entire ice sheet and calculate the F1 score by comparing the S-1 ice slab mapping to the ~90% of the OIB observations that were not used to choose α, β, and ϕ in that trial. As with the most likely ice slab extent, we calculate separate F1 scores for the upper and lower limits of the ice slabs. From the results of these ten trials, we use the backscatter thresholds that produce the largest total ice slab area to define the maximum plausible ice slab extent, and the thresholds that produce smallest total ice slab area to define the minimum plausible ice slab extent.

Figure 4 shows the results of this cross-validation. We find that across the 10 validation trials, F1 scores for the upper elevation limit of the ice slabs vary from 0.78-0.84, with no clear spatial trend. Since the F1 score for the most likely ice slab extent is 0.811, this suggest that values of α and β chosen based on data from one region of the ice sheet generalize well to other regions. Indeed, these thresholds vary by only $\sim \pm 1\ dB$ across all regions of the ice sheet. Therefore, the algorithm is reasonably spatially robust.

We do find a clear spatial trend in the generalizability of ϕ between regions. In particular, when ϕ is derived only using data from regions NE and N1, the resulting S-1 inferred ice slab extent in Northwest and Southwest Greenland agrees poorly with the OIB observations. However, conversely, the value of ϕ estimated using only data from the Northwest and Southwest does apply well to the North and Northeast. We suggest several reasons for this behavior. First, the North and Northeast regions have the least number of ice slab detections, so thresholds derived from data in those regions may be overfit to conditions that are not representative of larger areas. Second, we see steeper gradients in backscatter as a function of elevation in the North compared to the Northwest and Southwest. This suggests that small variations in ϕ would lead to large changes in ice slab area in the Northwest and Southwest, but small changes in ice slab area in the North and Northeast. As a result, the agreement between the OIB observations and S-1 detections is much more sensitive to errors in ϕ in the Northwest and Southwest than in the North and Northeast.

Revised Figure 4



Minor points

[R2] *In several points of the text: I would refrain from the use of statements like "excellent agreement" and prefer something more mild as "fine agreement".*

We will soften these statements or reword them in response to both this review and comments from other the other reviewers. For example, in the abstract, we have made the following changes:

[Original] "The S-1 inferred ice slab extent is in excellent agreement with ice-penetrating radar ice slab detections from spring 2017."

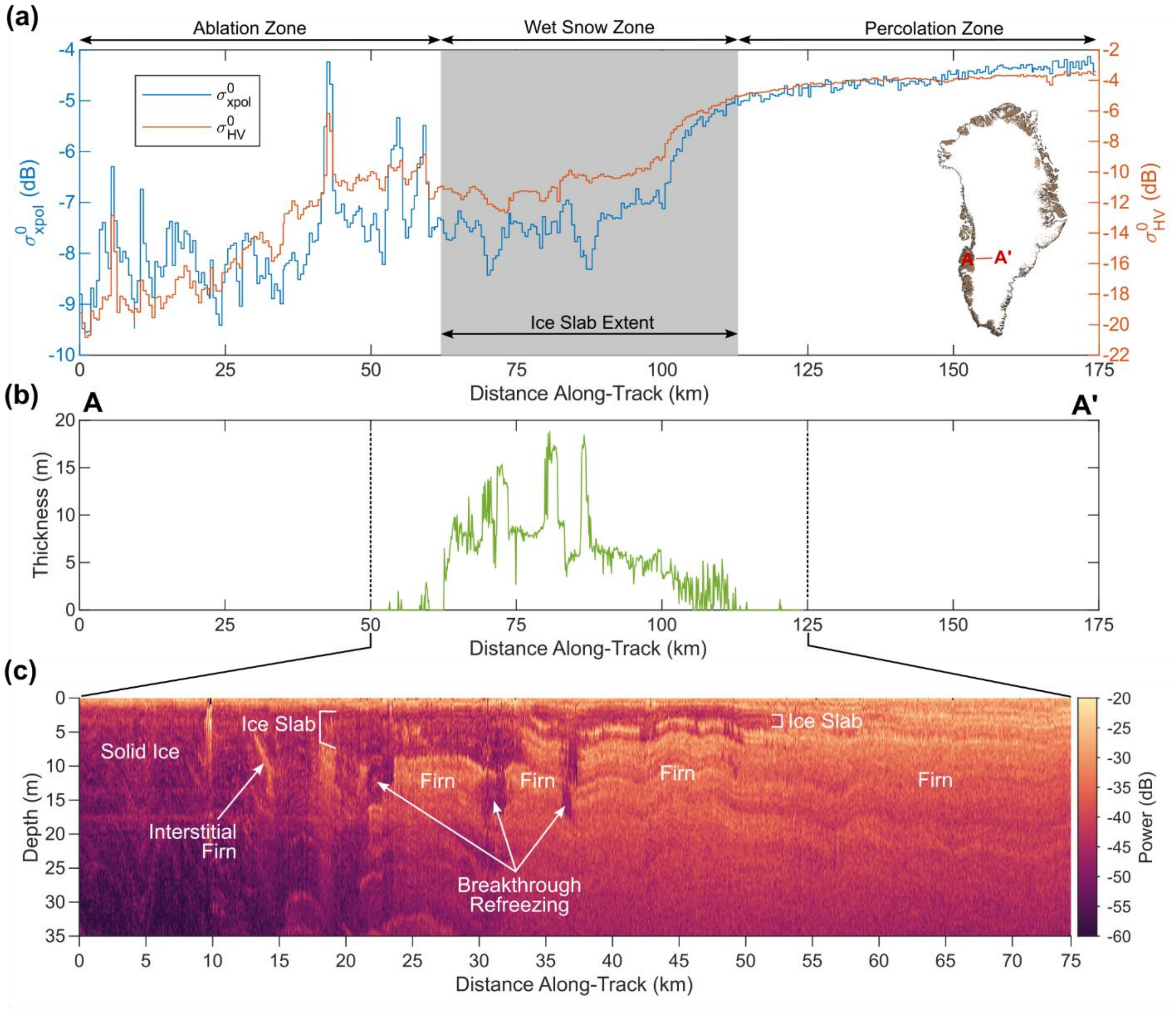
[Revised] "Our results show that there is a sufficiently strong relation between C-band backscatter and the ice content of the upper ~7 meters of the firn column to enable ice slab mapping with S-1."

[R3] *lines 65-66: the sentence "...the depth-integrated surface echo measured by the instrument contains information about the near-surface structure" is usually true, however C-band SAR data can be affected also by phenomena originating deep in the ice. For instance subglacial Vostok Lake,*

Antarctica, clearly visible in the Radarsat image map of Antarctica. I would say something like "...contains information mainly about the near-surface structure".
Good point – we have edited this sentence as suggested.

[R4] *line 78: delete the s at "cms". It is a SI symbol and doesn't require the s for the plural.*
Edited in text as suggested.

[R5] *Figure 1: the image of Greenland with the A-A' transect is very useful but must be better highlighted. As it is put, it gets unnoticed since the reader attention goes immediately to the top or bottom panel which are full of colors. Instead the map should be seen first. An option could be moving the top panel legend northwest, and then replace it with the Greenland map. Anyhow any different solution is fine.*
We have revised this figure as suggested and moved the map to the top panel. The revised figure is below:



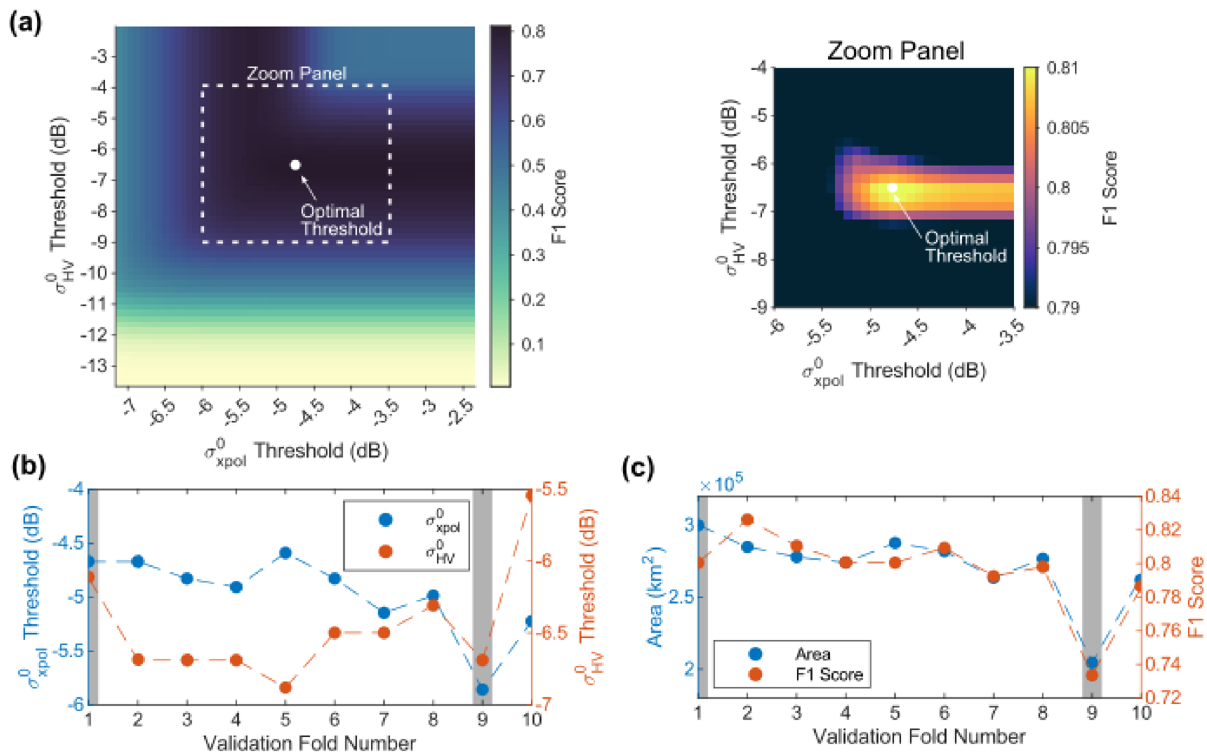
[R6] *Line 120: what "Agency" means? Also at line 138.*
Thanks for noting this! This is an error in the display of a citation to an ESA technical document and has been fixed in the manuscript.

[R7] *Line 127: I notice that speckle filtering is not considered in the processing chain while it is a fundamental step for SAR processing at high resolution. Is there a justification for not applying it?*
We mitigate speckle through multi-looking in both space and time. Data are first multi-looked to 500 m resolution. In each pixel, we then estimate a linear relationship between incidence angle and backscatter based on data points from all images collected between 2016-10-01 to 2017-04-30. Finally, we calculate the theoretical average backscatter at 35-degree incidence angle from this relationship. This has the effect of averaging all backscatter values in a given pixel across that whole time period. Therefore, in each pixel in our mosaics, we have typically averaged together at least 200 separate measurements. The averaging in time is necessary not just for speckle reduction, but to smooth out temporal variations in backscatter due to random snowfall events, wind scouring, etc. However, we agree that in future work, it could be interesting to explore more sophisticated speckle filtering techniques to see if we could produce a robust map of ice slab extent at higher resolution than 500 m. When revising the manuscript, we will include a clarifying discussion of this averaging approach when we discuss the resolution of the mosaics.

[R8] *Figure 2, caption, third line: "We excluded all the regions outside…" I suggest using a positive sentence as in the main body of the paper. "We considered only the regions…" is easier to understand.*
Edited in the text as suggested.

[R9] *Figure 2, caption: The disclaimer "Contains modified Copernicus Sentinel data 2016-2017, processed by ESA." makes the text heavy and is not informative at all. I suggest putting the disclaimer in the References and leave here a citation (or write a footnote). The same apply to Figure 1 caption.*
The ESA terms of use require this statement and our interpretation of the license is that it should accompany the relevant image. However, to de-emphasize this non-scientific information, we will move it from the beginning to the end of the figure captions.

[R10] *Figure 3, top panel: the colormap of the image makes it difficult to be read given it compress the almost entire information in dark similar colors (F1>0.2). A different colormap should be used.*
Unfortunately, this is less an issue with the colormap, and more with the fact that many of the F1 values are with about 0.05 of the optimal F1 value. If we use the full range of the colorbar to show these values, then values from ~0-0.7 will be highly saturated instead. We have added a zoom-in panel to the region around the optimal threshold to better highlight the small changes in F1 value around the optimal point. The revised figure is shown below:



[R11] *Line 332 and on: actually high-resolution data from PALSAR sensor (either onboard ALOS and ALOS-2) are already available without the need of waiting for NISAR or ROSE-L. Perhaps the issues are the full coverage of GrIS, the data of acquisition or the price of the products. I think the sentence can be better formulated.*

Thanks, this is a reasonable point. The main issue is indeed that ALOS-2 products are not freely available for scientific research and the ALOS-1 coverage of the GrIS is fairly patchy in most years. In particular, this makes it more difficult to average out the effects of speckle and temporal variations in backscatter to get a reliable mosaic, since there is typically only one acquisition over a given area per year, even in years with good coverage. However, we will amend this discussion to note that the ALOS-1 data is available and likely sufficient for an initial proof-of-concept testing of an L-band algorithm.