

Interactive comment on “Inter-comparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge” by Ron Kwok et al.

Anonymous Referee #1 ((Referee comments are in italics))

General comments:

In the study "Inter-comparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge", snow depth retrievals from five different algorithms are compared with each other, with in-situ snow depth measurements from two field surveys, a modified Warren climatology, and snow depths derived from ERA-Interim products. Most of the retrievals seem to reproduce the expected spatial snow depth patterns, certainly the gradient between first-year and multiyear ice. Nevertheless, differences between the products exist and biases can be significant. The authors state that the aim of this study is not to select the best algorithm, but to give an informative overview of existing retrieval algorithms, also serving to develop the next-generation retrieval algorithm.

The manuscript fills an important gap in providing a comparative overview of the different Operation IceBridge snow depth retrievals that have been presented over the past years. The scientific analysis seems to be rigorous and the manuscript is well structured. Therefore, I do not really have any major objections. But I think the paper can be improved in terms of figure quality, and also some clarifications and more details are needed in some parts of the manuscript (see specific comments). For example, it is not always clear over which period the different retrievals exist. Sometimes, one retrieval is missing in the comparisons and figures. I guess this is because an algorithm was not applied during the time when the in-situ measurements took place? I would suggest to include an additional figure or table, showing the intervals of the 5 retrievals and the overlap with the in-situ measurements.

Moreover, some figures lack a complete description, which sometimes makes it time-consuming to understand their content (see specific comments). I also wonder if the authors can provide some concluding remarks regarding the characteristics of the individual products in the conclusion section. To be more specific, GSFC-NK seems to reject much more echoes than other algorithms. Moreover some retrievals rather seem to over- or underestimate snow depths depending on how the interfaces are picked. Although I acknowledge that the aim of the paper is not to select the best algorithm, I think a statement about the characteristics of the different retrievals should be added in the conclusion.

Thank you for your suggestions.

In the revised manuscript, we now include a table that shows the availability of retrievals for each algorithm at the time the inter-comparisons were carried out. We also note that there is only one standard product (archived at NSIDC) provided by the OIB project, and other snow depth estimates used here (Wavelet, SRLD, JPL) were contributed by scientists interested in the retrieval process. The contributed retrievals and the results here will be used to guide the next step in the development of a community retrieval approach. At that time, there will be a detailed description of the new approach used in the processing of the snow radar data.

Specific Comments:

P5L7: How does the automated snow depth probe work? Is there any reference to the instrument? I also couldn't find much information in Strum et al. (2006).

A magnaprobe is an automated snow depth probe, approximately ~1.5 m in length, connected to a GPS and data logger in a backpack. At the end of the steel probe is an adjustable white basket, which sits on

top of the snow surface as the probe enters the snow pack. The depth is recorded as the distance between the tip of the snow probe and the white basket at the surface. GPS coordinates of each snow depth measurement are simultaneously recorded in the data logger. We now provide a citation to a document that describes the patented technology (Sturm and Holmgren, 1999).

P7: What about the method and analysis in Holt et al. (2015)? Why is it not considered here? At least, I believe, results should be discussed and mentioned briefly in the introduction and/or discussion section.

The authors of that algorithm elected not to participate in the inter-comparison project. We have added a citation to the work of Holt et al. in the text.

P10L27: Why were the wavelet retrievals not available for the assessment? This question/comment aims to the one in the general comments. I think it would be useful to clarify the availability of the different retrievals, may be by including another figure or table.
We added a table that shows the availability of retrievals from each algorithm for the intercomparisons period.

P15L7: “work was not to the select...”: delete “the”
Done.

P17L28: “Large interannual variability of retrievals from a given algorithm suggests issues in algorithmic robustness in adapting to changes in radar data quality.” Could these signals not be (partly) real? Why should the interannual variability suggested by W99 be the reference?

Yes, not all of the retrievals are in error - it is only that the basin-scale averages may be biased by those retrievals affected by system artifacts and changes in radar data quality.

We expect extreme deviations from modW99 and ERAI-sf to be suspect given:

1. the expected interannual variability of climatological snow depth of ~6 cm.
2. the relative agreement between modW99 and ERAI-sf over the period.

Figure 2: a) Why is NSIDC missing here?

The reason is that the standard NSIDC products are not sampled at the correct rate for these plots.

Figure 4: It would be helpful to add a legend in the figure explaining red and black dots rather than only mentioning it in the caption. And, there is a typo in the last sentence of the caption: “...are showN...”.

Legend added to the figure and typo corrected.

Figure 5: I guess the error bars represent the standard deviation at certain cm bins? It is not really stated, neither in the caption. This should be added.

Added.

Figure 6: Why are there gaps in the histogram in d)?

The gaps are likely introduced by discretization of the snow depth estimates.

Figure 9: a) Some points are beyond the axis, which looks a bit odd. Moreover, the number in the first column in the insets is not explained. I guess it is the number of measurements?

The quantities in the insets are now explained (they are the mean and standard deviation of the differences, the correlation of snow depth between the two tracks, and the number of samples).

Figure 11: Why is the wavelet retrieval missing here (counts also for Figures 7, 8, 9 and 12)?

The wavelet retrievals were not available at the time of this work.

b) I do not really understand the maps. Are the shown tracks the outbound and return tracks, but shifted to make differences visible? I would suggest to show just one map with the location of the tracks without color scale, and then add xy along track plots with snow depths along outbound and return tracks.

The tracks are not shifted in these figures. The only repeated outbound and return tracks (i.e., near coincident tracks) are shown in Figure 9. This is clarified in the caption of Figure 7.