Reviewer 1, Anonymous Referee

Review of "Sensitivity of CryoSat-2 Arctic sea-ice volume trends on radar waveform interpretation" by Ricker et al.

This study concerns the use of new CryoSat-2 radar altimetry data to calculate trends in Arctic sea ice thickness and volume, and their associated uncertainties. The authors assess a range of the leading-edge retracker thresholds as applied to CryoSat-2 radar waveforms and investigate the resulting range in both sea ice thickness, and volume, that arise with respect to a particular threshold setting. The goal of the study is to "isolate and quantify the effect of SAR waveform interpretation from other uncertainties in the freeboard to thickness conversion", and is the first study to directly address this issue. The authors find that while the absolute magnitude of the CryoSat-2 thickness estimates is impacted by the choice of threshold setting, the overall trends in sea ice volume over the CryoSat-2 measurement period remain consistent, regardless of the retracker threshold setting. The authors conclude that the seasonal evolution of snow on sea ice should be considered when choosing an appropriate retracker threshold.

I recommend this paper for publication after the authors address the following comments, as well as the more detailed minor revisions listed below.
Some of the manuscript's authors have provided a public release of CryoSat-2 sea ice thickness products through the Alfred Wegener Institute "meereisportal". However the online dataset is not explicitly mentioned in this manuscript. Can the authors clarify if any/one of the algorithms assessed here is used in the generation of the online sea ice thickness product? Many other investigators are currently using the publicly available datasets such that it would be extremely helpful to be able to point to this (or another?) published paper to reference the online data, and to understand its accuracy and limitations. If methodologies differ significantly, that might also be addressed here.

We thank the reviewer for his comments. We added a citation of the "meereisportal" where we directly link to the products that are provided there. We basically used the same processing chain, but recently implemented some improvements. We will therefore update the data that they coincide with the results of this paper.

Can the authors describe in more detail the choice of waveform fitting routine (grey line in Fig. 3) utilized for the lead waveforms? The authors state that it is a "result of linear interpolation and smoothing". Please provide more sufficient details on this approach. The waveform fit (or 'model', e.g. grey line, Fig 3b) appears to be at least a full range-bin wider (a half range bin on the leading- and trailing-edges) than the waveform itself. This misfitting will cause a bias in the estimated lead/sea surface height elevation. I suspect if the lead waveform fitting routine were to work more efficiently, the difference between the elevation tracking pointing associated with each of the TFMRA thresholds would be very close, and differences would be negligible.

There has been a bug in the plotting routine for Fig. 3 regarding the normalization. This is fixed now and shows that the interpolated and smoothed waveform matches the original waveform at the leading edge. We also added some more explanation to the text.

P1839 L4-5: Can the authors explain how the open water/lead elevations are interpolated – i.e. what interpolation scheme is used, and what is the result of the low-pass 25 km filter? Does the routine result in one SSA estimate per 25 km grid cell or results sampled at a higher resolution along-track?

We retrieve the SSA track-wise from linear interpolation between the detected leads. When we observe a dense cluster of leads, we obtain significant noise on the SSA doing the linear interpolation (due to the speckle noise that is affecting each measurement). Therefore we apply a running mean with 25 km width to reduce the noise on the SSA. We also added some more explanation to the text.

The thickness results over multi-year ice are confounding, particularly those results presented in Figure 8 and Table 3. Why is MYI sea ice thickness thicker for the TFMRA50 threshold than for the TFMRA40 threshold, particularly in March 2013? This is counterintuitive based on the waveform examples provided in Fig. 3 and the threshold re-tracking points on the leading edge.Moreover, the title rows in Table 3 are confusing, e.g. the second row of “FI(m)” seems redundant and misaligned with the results. Please revise the layout of the table so that the results are more clearly presented.

Only for the TFMRA40 retrieval we applied a penetration correction term to the freeboard, since from the comparison with the laser altimetry data we observed that for the 40 % threshold, the radar is not penetrating the snow completely. In contrast we assumed a full penetration of the snow for the TFMRA50 and TFMRA80 retracker in consistence with Laxon et al. (2013) and Kurtz et al. (2014). Assuming only partial penetration leads to a lower ice thickness compared to the assumption that the freeboard represents the ice freeboard. We acknowledge that these assumptions were confusing for the reader. Therefore we now consider the freeboard without any corrections and also compare the freeboard retrievals of the 50 and 80 % threshold with the laser altimetry data.

P1842 L8-9: What is the “ICESat/GLAS” surface-type mask? Why is it applied to the CryoSat-2 data in this study?

We used this mask to exclude the Canadian Archipelago but additionally excluded the Baffin bay and now also the area between Svalbard and Sewernaja Semlja, since there are only few or no measurements that are considered for the W99 fit (see Warren et al. (1999)). Only snow depth data from regions where the W99 fit bases on measurements should be used.

P1842 L13-14: What motivated the choice of a boxcar average to interpolate data across the 2 degree polar hole in the CryoSat-2 data? I don’t understand the result of a boxcar interpolation scheme in this case. Is the approach actually a linear interpolation, since you only have one data point across either side of the polar hole for a given CryoSat orbit? Have you checked the results against other interpolation schemes – e.g. a weighted average?

We former used a linear interpolation to close the gap at the pole. We then used a boxcar average to smooth the interpolated area. This is now obsolete since we excluded the sea-ice volume considerations to focus more on the difference of the freeboard retrievals from different thresholds.

P1846 L3-5: the authors state that the ALS system provides “high-precision and high-resolution measurements”, but then go on to state that “the accuracy for the range measurements is about a few cm”. The specific details of the lidar system – elevation accuracy, precision, and measurement resolution, should all be included here so that the CryoSat-2 elevation measurements can be placed in the context of lidar data.

The precision of the laser measurement itself is very high indeed (in the range of mm). But the positioning of the aircraft that is needed for data processing introduces some uncertainties that increase with the flight distance. More details can be found in *Airborne lidar measurements for Cryosat validation* by Forseberg et al. (2002) and in the ARCGICE final reprot *Combination of Spaceborne, Airborne and In-Situ Gravity Measurements in Support of Arctic Sea Ice Thickness Mapping* by Forseberg et al. (2007).

Figures 6 and 7 appear to suggest that the long repeat period of the CryoSat-2 orbit, with non-uniform monthly subsampling of the Arctic basin, partially contributes to the radar freeboard uncertainty term. An assessment of radar freeboard uncertainty indicates a peculiar pattern wherein uncertainties align along lines of longitude (e.g. Fig 6b), a result that is not easy to explain by the physical nature of the sea ice pack. The fraction of detected leads also indicates a similar pattern. Although the authors describe these patterns in the conclusion section, a re-evaluation of your gridding routine/25 km grid cell size to account for the non-uniform monthly sub-sampling of the Arctic Ocean due to the CryoSat-2 orbit appears necessary. Since the goal of the study is to "isolate and quantify the effect of SAR waveform interpretation from other uncertainties in the freeboard to thickness conversion" the magnitude of uncertainty associated with orbit subsampling should be addressed with respect to the uncertainties arising from retracking thresholds, and ideally should be removed before considering the volume trends.

The orbit subsampling and the procedure of gridding is definitely an item for future investigations, but not in the scope of this paper. Furthermore this is rather a question of how to interpret the monthly average. These uncertainties do not result directly from the CS-2 measurements or the freeboard-to-thickness conversion.

Section 5: How does the resolution of the ASCAT backscatter map relate to the ice- type mask utilized in this analysis? Figure 6 would indicate that the higher backscatter associated with MYI features are not accurately represented by the "MYI mask", especially over the specific features that the authors identify. How do the authors justify the use of the coarser resolution mask rather than ASCAT backscatter to delineate ice type? Indeed, a comparison of the FYI and MYI ice thickness results using both methods would be insightful.

We use the OSI SAF ice-type product for discrimination between FYI and MYI. It also uses the ASCAT backscatter as input data. The "MYI mask" might have been confusing. We actually interpolate the OSI SAF ice type product on each CS-2 track. When we grid the data, we also grid the ice type for each CS-2 measurement and obtain a monthly mean ice type, related to the CS-2 product. The term "mask" might be confusing here since we only use the gridded ice type to visualize the ice type on the maps. However, we observe that the shape of the CryoSat-2 waveforms also indicates the ice type and we might use this information in the future for data processing.

Minor revisions:
Re-check English usage and grammar throughout P1832 L2: "last decades" – clarify which decades P1832 L13: change to "on the order of"
P1833 L8: change "multi-seasonal" to "multi-year"
P1834 L1: What is the reference for "large footprint of the order of 10km and an orbit coverage limited to 82.5 oN"? Check this statement. The literature suggests a surface footprint size approximately 2–10 km in diameter on sea ice (e.g. Connor et al., 2009 and references therein). ERS-1, -2 and Envisat have coverage to a latitudinal limit of 81.5 oN.
P1834 L25: Change to "Laxon et al. (2013) used a leading-edge threshold retracker, while in a recent study. . ."
P1837 L4: change "exemplary" to "example" or "typical" P1838 L13: Change "refused" to "discarded" or "removed"

P1839 L18: change "inside the snow layer" to "through the snow layer"
P1841 L6: include the word "modified" before "Warren snow climatology"
P1841 L21: Insert "Consistent with the approach of Laxon et al. (2013)" before "we use ice densities of . . ."
Table 2: Rather than the term "variable", the range of values considered for each parameter would be much more helpful. Please revise.

The minor revisions have been considered and were incorporated into the manuscript.

Reviewer 2, Christian Haas

Review of “Sensitivity of CryoSat-2 Arctic sea-ice volume trends on radar waveform interpretation” by Ricker et al.

This is an interesting paper describing the processing and validation of CryoSat ice thickness retrievals and application to three years of Arctic data. It includes a sensitivity study of the impact of certain retrieval parameters on derived ice volume, as well as a thorough error analysis and budget. This is an important paper given that the authors are leaders in the processing of CryoSat data and that little has been published about the usage of CryoSat data so far. The authors include the innovative consideration of uncertain penetration depth of radar waves into the snow. Results should definitely be published. However, I am afraid that the paper is overloaded and therefore certain aspects are not sufficiently presented or hard to understand or confusing respectively. I strongly suggest to consider to split the paper into several manuscripts and to more carefully focus on the key aspects in each of them:

- Study of the impact of different retracker thresholds
- Validation with airborne data
- Error analysis
- - Application to Arctic ice volume trends

There are key concerns with all these four topics which should be carefully addressed. These are outlined below. In addition to these major revisions, there are numerous language issues with the paper. I will send an annotated manuscript directly to the authors where these are marked.

We thank the reviewer for his comments. We followed the reviewer’s suggestion to focus on the freeboard retrieval and to point out the differences between the applied retracker thresholds. Therefore we excluded the consideration of sea-ice volume and trends. But we keep the validation part because we think it is an important supplement to point out the differences between the thresholds. We also keep the error analysis to present the uncertainties due to the choice of the threshold in the context of other uncertainties that arise from freeboard processing and the freeboard-to-thickness conversion.

Major concerns:

1. Retracker thresholds

The authors should carefully discuss the relation of their results to the results of Kurtz et al. 2014 in The Cryosphere Discussion. That paper shows that the appropriate threshold may be very close to the waveform maximum and should even be above 80%. Why are the results in the present manuscript different? Other studies have applied different thresholds to measurements over ice and leads, respectively (Laxon, Giles, et al.). Why has this approach not been considered (and possibly rejected) in the present manuscript?

We acknowledge that the 80 % threshold is only simulating the method of Kurtz et al. and significantly differs from our method. They used a forward model and fitted the model to the waveform, which might explain the deviating results. Nevertheless we want to simulate the range of retracking approaches that are used in literature to point out potential uncertainties for the freeboard and thickness retrieval.
We use the same threshold for sea ice and leads to avoid potential biases due to the usage of different retrackers.

In addition to these concerns, it is unclear why the authors chose to modify the retrievals with the 40% threshold retracker with estimates of penetration depth and snow thickness. Shouldn’t the same assumptions be made for all retrackers such that the results are really comparable?
In addition I find the treatment and interpretation of the 40% threshold very problematic.

Indeed, this part was confusing for the reader. We wanted to point out that a penetration correction is needed, supported from the results of the validation measurements with airborne laser altimetry. At the same time we wanted to simulate the results by Laxon et al. and Kurtz et al. since they assumed that the main scattering horizon is located at the ice freeboard.
We now included the freeboard retrievals from all thresholds into the comparison with the airborne laser altimetry. Furthermore we discarded the penetration correction and focus on the uncorrected radar freeboard.

The authors seem to confuse penetration issues with the effects of different retrackers. While it is true that these can have the same apparent effects, I don't think that thresholds should be adjusted to compensate for the (unknown) penetration uncertainty. Instead, a physical understanding of the meaning of different thresholds should be developed. Penetration is a physical fact depending on snow properties, and does not change if thresholds are changed. The comparison of the effect of different thresholds is interesting and important on its own, given the spread of thresholds used/suggested by the various existing studies. It does not need to be blurred by modifications based on assumptions on penetration.

We acknowledge that our usage of terms for discussing these different effects was confusing for the reader. We improved that section and try to distinguish between effects of physical penetration, surface roughness and the choice of the threshold/retracking method on the leading edge. Nevertheless we want to point out that we treat these uncertainties as one since we are not able to quantify and isolate these different effects.

2. Validation with airborne laser data.
This is a great data set and it is nice to see that validation has been attempted. However, why does the validation only include the 40% threshold retrievals? Please include the retrievals with the other thresholds as well to better see the impacts and their magnitudes of the different parameters on the agreement with the airborne data. Or do a separate validation study with more extensive comparisons.

We included now the retrievals from all thresholds into the comparison with the airborne laser data (see above).

3. Error analysis
I commend the authors for their most extensive error analysis. However, I think the study of penetration effects and overall error sensitivity should be significantly expanded and should include explicit sensitivity studies for the individual, novel parameters considered, e.g. penetration depth, level 1 b and SSA uncertainties, similar to what has been done by Giles, Kwok, and others. The authors may also refer to a paper by Tonboe et al (Canadian Journal of Remote Sensing 2010) which includes a description of the penetration issues and the scattering horizon, and sensitivity studies.
In addition I wonder if it is correct to consider the uncertainties due to unknown snow thickness and snow and ice densities as random (P. 1844, L19ff, and Eq. 12). They should rather contribute a constant bias (or vary within the scales of the natural regional and spatial variability), and will therefore not become smaller if averaged over longer times or over more measurements in a given grid cell?

We followed this suggestion and carefully separated the uncertainty budget into a random and systematic part and show the impact of the individual parameters on the freeboard / thickness retrieval. We also state numbers for individual contributions to the freeboard and thickness uncertainty budget. Indeed the systematic uncertainties dominate the budget and can significantly affect the freeboard and thickness retrievals.

4. Ice volume trends

With the high uncertainties of the effects of snow and ice densities, snow thickness, and penetration depth, and ice type the conversion of freeboard to thickness adds an additional layer of complexity to the interpretation of the results. I strongly suggest to restrict the sensitivity study to maps if freeboard and its variability over the three year study period. This will show more clearly where anomalous patterns appear and their interpretation will be easier because they haven't been convolved with assumptions about snow thickness variability and ice types. It is trivial to see that higher freeboard in one year should correspond to thicker ice in general.
Some of the results, particularly the freeboard increase between March 2013 and November 2013 seem surprising as mentioned by the authors. They hypothesize that this is due to a thicker and wetter snow cover than usual. Although this could be one possible explanation indeed, I wonder how likely such an event is in reality and if there is any other evidence from (in-situ?) observations? Snow thicknesses during the CryoVEx 2014 campaign in the Beaufort Sea and north of Greenland were quite normal, although admittedly regionally very limited.

We followed this suggestion and exclude the sea-ice volume part, since this needs a more detailed consideration.

More minor but important comments and suggestions:
Title: change "volume trends" to "freeboard trends" and limit scope of paper on free- board which is the most directly retrievable parameter and less error prone than ice thickness and volume?

We changed the title to 'Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation'. We keep our consideration on sea ice thickness retrievals, even though we rather focus on the uncertainties that arise due to the conversion from freeboard. We think it is important to state thickness uncertainties because sea-ice thickness (also the provided AWI product) is widely used in the sea-ice community, probably more than freeboard.

Abstract: shorten and remove unnecessary introductory parts. L19-21: Unclear without further explanation. Use term waveform. L1-3: better point out this key result. Include validation results/mentioning should you decide to keep the validation section. Last lines: unclear.

Fixed.

Introduction:

Usage of term "scattering horizon" in L11-12 and elsewhere in the paper: Clearly define the scattering horizon, and use the term properly. Clearly distinguish it from the apparent effects of different thresholds. For example, on page 1853, L1-2, retrackers do NOT penetrate, and L12, the main scattering horizon is NOT penetrating - how can it??

Here we meant the location of the scattering horizon below the snow surface. We changed that.

P1834, L20: do you mean the RANGE TO the scattering horizon?

Yes, exactly.

P1835 L 6 and elsewhere? It is commonly assumed that CryoSat cannot be used in the warm season. Therefore moist snow is less likely. Mention that stratified, high density snow can cause these effects.

There is evidence that during May and also at the end of April already melting occurs in some parts of the Arctic Ocean. Also melting and refreezing in October might be a potential process.

Data and Methodology
P1836, L15: Which data products and versions did you use exactly?

We used the Level1b, Baseline B data from ESA. We included that.

P1837 L 4 and elsewhere: Replace exemplary with typical or example or example of

Done.

L 23: mention that surface type "lead" typically does not represent a single, large lead, but a mixed sea ice surface including a few leads within the footprint.

Done.

P1838, L9 and elsewhere: What does SSD stand for? Could it be SD? Explain what standard deviation you mean exactly, and of what variable.

SSD stands for *stack standard deviation* and is a measure of the variation in surface backscatter with incidence angle of the different beam echoes that are stacked. Included.

q. 2&3: can you show waveform example that explain the differences between Eq 2 and 3? In Fig. 3?

We marked the range bins in Fig. 3 that are considered for the *left* and *right* peakiness.

Eq. 8: Although this is only a mathematical exercise, it would be better to use ice freeboard and not snow freeboard for the thickness conversion. Ice thickness depends less critically on uncertainties of snow thickness if ice freeboard is used than if snow freeboard is used. One of the (many?!) advantages of radar altimetry. . .

Since we do not apply a correction term to the freeboard anymore we also consider the ice freeboard for the conversion. Although in this case the ice thickness depend less critical on snow depth uncertainties directly it does indirectly due to the freeboard that is affected by the snow depth uncertainties (penetration, lower wave propagation).

P1841 L19: how are these values identical or different from the "unified" ones from Kurtz et al. 2014?

The snow densities from the W99 climatology vary between 290 kg/m$^3$ (November) and 320 kg/m$^3$ (March) approximately. Kurtz et al. used a constant value of 320 kg/m$^3$.

P1842 top: Why not used modal thickness instead or in addition to arithmetic mean? May be more representative?

We changed our processing and now use a weighted arithmetic mean where we use the random uncertainties to weight the individual measurements and to have less impact from measurement with a high uncertainty of the sea-surface height. One could also look at the modal thickness in future studies.

L8-9: good idea to exclude Baffin Bay. Probably there are no in-situ data to support W99 anyways?

Considering the Warren et al. Paper we actually further have to exclude the region between Svalbard and Sewernaja Semlja, since there are only few or no measurements that are considered for the fit. This has been applied to the thickness retrieval.

L12-14: usage of boxcar average is unclear? Do you do this in 2D, or for radial lines through the North Pole, or what?

We finally just used a linear interpolation and then smoothed the gab with a boxcar average. Nevertheless this is not done anymore since we do not consider the sea-ice volume anymore.

2.1.2 Snow layer corrections: This section is critical for a thorough discussion of penetration depth, and needs much better introductions of the issues of wave propagation speed in snow and how this relates to the assumed snow thicknesses and penetration.
A conscious adjustment of thresholds can only be done if penetration depth is really known? Requires knowledge of regional variability as well?

This part was indeed critical. We estimated a constant penetration from the comparison with the airborne laser altimetry data. Since the penetration in conjunction with surface roughness effects seems to be very variable this only a coarse approach. Nevertheless the approach of introducing a correction term due to changing snow and surface properties to reduce the observed bias seems valuable to us.

Eq. 6: where does the factor of 0.28 come from and how was it estimated?

The value of 0.28 m has been the mean difference between the snow freeboard from the validation data and the coincident CS-2 measurements from April 2011. Therefore we assumed that this is the apparent penetration of the radar pulse. However, since the we do not now the real physical penetration this value is rather uncertain and can also result from the choice of the retracker as written above.

P1840, L19: check usage of “those”. You mean values outside this interval are discarded.

Fixed. Yes, values outside this interval are discarded. We further extended the interval to allow also negative values, considering the range of the speckle noise.

P1843: explain what speckle noise is. Check numbering of eq. 6 and 8. Reference Kurtz et al for interpolation errors.

For the speckle noise we refer to Laxon et al. (2013). A short explanation is also included now.

Over what spatial range is the mean SSA determined?

The SSA is is determind by linear interpolation between the detected leads. Afterwards we smooth the SSA with a running mean of 25 km.

P1844: explain that sigma_p is defined through Eq. 6 and 7?

We excluded the penetration term in our analysis. This error is now considered as a conjunction of uncertainties of the choice of the retracker (threshold), the surface roughness and the radar penetration.

P1846: Snow freeboard from ALS: Should you not use the same leads as for CryoSat to have the same SSA? If you chose them from the ALS data differently the derived SSAs may be different?

We wanted to avoid introducing potential biases into the CryoSat-2 freebard. We found a bias of a few cm between the ALS SSH and the CryoSat-2 SSH, so we decided not to combine both products.

P1847 L16-20: refer to CryoVEx data report (Willat and Haas) and include include results from 50 and 80% threshold retrackers.

Done.

Results

P1848 top: We know that the OSI-SAF ice type classification is uncertain at the ice type boundaries. Why are the errors along this boundary not larger, as CryoSat may experience some MYI in the FYI classified region, and vice versa?

With introducing systematically uncertainties this is fixed (See figure 13).

L13: be careful with roughness effects on backscatter? The impact of volume scattering of MYI on high backscatter are just as high.

Yes, it's finally a matter of dielectric properties of the snow and the surface roughness that both change with the age of the ice. We included that.

P1849 top: use "north of Canada" instead of "Canadian Archipelago"

Fixed.

Why do you assume that the 50% and 80% thresholds represent the same scattering horizon? The waveform amplitudes are clearly different and so must be the retrieved height. The issue as discussed above is that penetration is a physical fact which does NOT depend on a certain retracker in any way.

See above. We tried to make the same assumptions like in literature (Laxon et al. (2013) and Kurtz et al. (2014)) to simulate their results.

Discussion – use sub-headings for better structuring contents.

Fixed.

P1851 L 2: it is known that there can be quite deformed ice in this region downstream of the East Siberian Islands. This could cause the same effects on backscatter and freeboard.

Yes, but ASCAT should be rather sensitive to small-scale roughness and not to deformed ice? So this would rather suggest MYI for this area (which can be deformed as well)?

P1853: Mention Willat results much earlier when introducing penetration
Last paragraph: How do results compare with other studies?

Regarding the spatial distribution of the freeboard we find similar results to Laxon et al. (2013), which is also supported by the backscatter map. In contrast we find different spatial patterns compared to Kurtz et al. (2014), in particular the gradient between FYI and MYI is less present in the results of Kurtz et al. (2014).

Conclusions

L10-16: mention some numbers to summarize results. Be more specific about key results in general.

Fixed.