Reviewer 2, Christian Haas

The authors did a great job describing the nature of CryoSat waveforms and the resultant model and retracking procedure. However, I am not able to verify the detailed correctness of the model. It may also be worth to even more extensively describe the consequences of the applied simplifications and assumptions, e.g. if generally level ice was assumed, or what the consequences of mixed ice/lead surfaces are. The comparison with the threshold-derived freeboard estimates is solid and identified differences between the methods and their results are well explained. However, comparison of results with airborne laser altimetry is questionable particularly due to assumptions about the accuracy of snow thickness retrievals from airborne snow radar measurements and the assumption that CryoSat radar returns originate from the snow/ice interface only. With the expected uncertainties resulting from these two error sources the good (average?) agreement between waveform-retracking and laser altimetry is almost unbelievable. In fact, the discussion of poorer correlation coefficients in the end of the manuscript raises some questions about what conclusions can be drawn from the good average agreement.

Although I recommend to publish the paper with minor revisions, I am very concerned about the quantitative comparisons of thickness results. I would strongly encourage the authors to focus on freeboard retrievals only, and to omit any quantitative discussion of ice thickness results. It is clear that larger freeboard resulting from one data set would result in larger thicknesses if the same assumptions would be made otherwise. I think it would be sufficient to leave it at that, although the author's attempt to point out the importance of using the same assumptions about snow and ice properties is of course commendable. It is also clear that clever assumptions of ice and snow densities and thickness can be made to yield better agreement between freeboard and thickness retrievals from different data sources. But it is somewhat a different topic and problem than the description of the new method. And many other authors have published sensitivity studies of the importance of various snow and ice properties on freeboard-to-thickness conversions.

We thank the reviewer (Christian) for his comments, we realize there were many unclear aspects in the original manuscript which needed to be clarified and addressing these issues will strengthen the results, The comparison between the airborne data is meant to provide independent validation of the improvements which are made from the waveform fitting method, which proceeds largely from theoretical modeling of the CryoSat-2 return. The focus is on showing how improvements can be made, particularly with respect to the RMS and mean differences. It should be emphasized that this study does not provide strict error bounds on the retrieval method. What is needed for that is a detailed assessment of the errors in the airborne data through comparison with in-situ data (the overflights of CryoVEx survey areas by IceBridge would be a good place to start). Once errors in the airborne data are better characterized, a more detailed study could be used to place error bounds on the CryoSat-2 freeboard and thickness retrievals. Towards that end, we have removed use of the word "thickness" from the title as more work needs to be done to ensure accurate freeboard to thickness conversions from CryoSat-2 data alone. We have left the comparison with the thickness data in the manuscript because we feel it better emphasizes one of the primary goals of

CryoSat-2 which is to measure sea ice thickness and volume. Also, this shows that if consistent physical assumptions are used in the retrieval process, then a better merging of the the sea ice thickness data record from both radar and laser altimetry can be attained.

Several other aspects should be clarified as outlined below.

Specific comments

Title and abstract: Please consider to remove “thickness” from the title and abstract. The abstract should be rewritten to be clearer on the used or developed methods and to better represent the results. E.g. there is confusion about empirical versus numerical model, although this becomes clear in the text, where additionally physical model is used. I don’t think the comments about consistent assumptions in the abstract are easily comprehensible for non-experts. And I could not find any reference in the text to the exact values of differences between CryoSat and OIB stated in the abstract? How were the values in the abstract obtained?

The word “thickness” has been removed from the title as indeed sea ice thickness retrievals are not the main focus of the paper. We have kept the thickness aspect in the abstract as we wish to emphasize that the main goal of the paper is to show that if consistent physical assumptions are used in the retrieval process, then merging of the the sea ice thickness data record from both radar and laser altimetry can be realized.

The abstract has been rewritten to remove confusion about the model, we now refer to it as a physical model in order to avoid mixing terms.

The values in the abstract as to the exact differences between CryoSat-2 and OIB are the mean values from Table 2. These values have now been placed into the text to state how they were computed.

P 725 top: this discussion is important but could be kept much more general, as the individual values aren’t supported by references and show relatively high variability in nature. Many other authors have already presented sensibility studies in this regard.

The individual values reported here are taken directly from the studies which have been referenced. A statement has been added that there is uncertainty in the sea ice thickness due to the parameterization of sea ice density in different studies as well natural variability has been added in, this has implications for estimates of sea ice volume trends and a reference to a recent study by Zygmuntowsak et al., 2014 which discusses this has been added in.

P 725 L13 & 21: Clearly define and consistently use freeboard or surface elevation, or snow freeboard or ice freeboard etc…from the outset.

A sentence has been added which defines freeboard as the height of the sea ice layer above the sea surface within the context of this study.

P 726 top and L 8: briefly explain SAR, and say if stated footprint is for pulse-limited radar? Similarly, the discussion of SARIn seems to complicate the flow of the paper.

Why not ignore SARIn here and just discuss it in the discussion or conclusions?

A sentence describing the SAR processing which is used has been added in. The footprint size has been updated to state that it is pulse-limited at 1,650 m in the across track direction and pulse-Doppler-limited at 380 m in the along track direction.

We have added in the statement that the SARIn mode data were truncated and used in a manner which is consistent with the SAR mode data. The purpose was to state that the SAR mode data were used in the study because they cover a significant portion of the study area, but that differences in the observing modes were minimized to the best of our ability.

P726 bottom: What are the errors of these corrections, or do they just not matter because the same corrections are applied to the data before the different trackers are applied?

The errors of these corrections will vary significantly in space in time. For example, the errors in tides will be small in the central Arctic Ocean but larger in more shallow areas with complex bottom topography. In terms of comparing the ELTF and waveform fitting method, these errors indeed do not affect the comparison because they are applied independently of the tracking correction.

P727 middle: Why use 2013 quick look data if they are less accurate? Why not focus on 2011 and 2012 only? And why is the comparison of 2013 quick look data with CryoSat retrievals not worse than 2011 and 2012?
General comment: It is clear that the OIB data are well referenced to SSH and that accuracies depend on e.g. number of SSH tie points. However, what did you actually do with the CryoSat retrievals? I did not find a clear description about how SSH was re-constructed from the lead waveforms? What is the spacing of tie points in the CryoSat data?

Quick look data from 2013 have been included to broaden the temporal coverage of the comparison since final data from 2013 are not yet available. We have now included statistics on the expected additional error which occur because of the use of the quick look data. In terms of the better comparison between 2013 and the previous years, we do not know entirely why this is the case. It could be due to different sampling which included more first year ice areas during this campaign.

The description of how SSH was constructed from the lead elevations was not well written in the original paper. It has now been updated to clarify how this was done (see the points below).

P. 728 top: This model is a good first step towards better understanding and application of CryoSat waveforms. However, it is clear from the outset that the sea ice surface is not Gaussian (hence, e.g. the possibility to use open water tie points). Can you add a sentence at the outset that this is so (e.g. use a Wadhams reference). Of course you could well examine surface roughness and autocorrelation with the ATM data and use those results in future improvements of the model. What else is known about the shape of surface elevation autocorrelation functions?

A reference to a study by Rivas et al., 2006 has been added to address this point. This study analyzed the surface roughness from laser altimetry data and found that a Gaussian height distribution is mainly valid for smooth ice. For rough ice, an exponential autocorrelation and Lorentzian power spectral density more accurately characterizes the surface roughness. This has been added in after equation 3.

P. 729, L5: conducting is not a good word here; it is rather the dielectric properties. Why not say more generally that the ice is assumed to be opaque?

The statement has now been removed as it it somewhat redundant.

P729, L25: is sigma the rms roughness, more explicitly?

Sigma has now been explicitly defined as the standard deviation of the surface height.

P729, L27: I think it is known that the form is rather log-normal or exponential; but you have to start with the most simple assumptions.

Yes, we have started with a simple assumption and noted that more accurately characterizing the distribution could be used to improve the retrievals in a future study.

P 732 bottom and 733 top: This is confusing. Does the discussion so far assume rather level ice? (where the Gaussian distribution may be more realistic?). There will almost always be ridges within the footprint, except maybe on fast ice e.g. in McMurdo Sound.

The discussion to this point is general, it will apply to the extent that the autocorrelation model and height distribution are accurate with respect to the ice type. The assumptions the scattering model and a discussion of its validity has now been added in.

P. 734, L11: does the model only consider purely lead or ice surfaces within the foot-print? Does this mean that only the largest leads could be identified?

Yes, the model considers only a surface which consists purely of a lead or ice surface within the footprint, this has now been clarified in the text. In terms of the size of the lead which can be detected, this depends largely on the alpha term which geometrically limits the radar return to a small area closer to the nadir point. Thus, a small lead located near nadir would have a stronger return than a physically larger lead which is located off-nadir.

P 734 L18-22: how does this differ between pulse-limited and SAR waveforms?

The physics of the behavior of the returns is exactly the same for traditional pulse-limited radar returns (such as from Envisat) and the SAR processing used by CryoSat-2. There will be significant differences in the waveform shape due to the summation and weighting of the off-nadir look angles which is used in the CryoSat-2 processor.

Section 3.2.1, 3.2.2, and following: The use of terms Sea Ice Leads and Sea Ice Floes sounds kind of wrong. A floe is something rather confined and limited although they can be vast. I suggest to consistently use the terms Leads or Ice Cover or Sea Ice Cover or Ice Surface?

Agreed, this is confusing particularly with the use of the term “sea ice leads”. The sections have been renamed to “Lead returns” and 'Sea ice surface returns”.

P735 L 15: unclear: all amplitudes seem to be equal in Figure 3

This was a typo, it has been changed to the “black color waveform in Figure 3”.

P737 Fitting routine: What do you do with waveforms which possess multiple peaks, or where the highest peak is not the first peak, respectively?
P740 L8-22: More generally? What is the rejection rate of waveforms that cannot be fitted? Later on you mention that there may only be very few points in each grid cell. What is the average number, or how does it vary?

The fitting routine can provide a fit regardless of whether a waveform possesses multiple peaks and if the highest peak is not the first peak. However, when a waveform does not meet the stated requirements for the first peak then the waveform is not fit and no elevation is retrieved. In total, about 60% of the CryoSat-2 waveforms are used for elevation retrievals during the March campaigns, this rate is largely determined by the use of the pulse peakiness and stack standard deviation requirements. For floes, about 90% of the waveforms which meet the pulse peakiness and stack standard deviation requirements are fit. This has now been added to the text.

The average number of points in each grid cell is mainly dependent on the orbit, points near the upper limit latitude of 88 degrees have many more observations than lower latitude points.

P739 L15 ff: Is this based on the fact that sea ice returns are wider than lead returns?

Exactly, different values need to be used since the lead returns are typically much thinner and only extend over a short number of range bins.

P741 L5 ff: I would argue that Willat et al show that penetration is actually quite variable and that complete snow penetration cannot be generally assumed. This is a key point for all following discussions in the manuscript, and therefore I would suggest to not go into too much detail. More research needed...lets go CryoVEx...

We have changed this statement to now quote specific numbers from the Willatt et al., (2011) study which showed that 80% of the radar returns from the 2008 CryoVEx field experiment were closer to the snow-ice interface. While for the 2006 CryoVEx field experiment only 25% of the returns were closer to the snow-ice interface. This was done to put a less generalized statement. More research into this area is certainly needed.

P741 L9ff: This discussion is very short and I would be afraid that these steps introduce high uncertainties. Is your assumption that they affect waveform and threshold retrievals equally? I think this might be a fair assumption but should be stated?

Yes, it is a fair statement to say that applying the sea surface height corrections will

affect both methods equally. This has been added to the text in the following section where it is more appropriate.

P743 L20 ff: The discussion of different thresholds for leads and ice, and the usage of certain geoid models creates significant uncertainties. I think it is important to point these out, but Section 5.2.1. becomes quite confusing. Can these issues be explained a little better? What geoid model was used in the present study?

The use of the EGM08 geoid was inadvertently left out of the description of the data in the original manuscript. We have now added this in to the paper and have rewritten the section to clarify the discussion.

P744 L1-2: Important sentence, state earlier on!

This statement has now been added to the abstract.

P744 L20-L23: This is also very important and should be stated at the outset of the paragraph. Does the paragraph first assume that no snow is present at all?

The sentence has been moved up in the text. In the comparison between the waveform and threshold retrackers, the impact of snow was indeed not considered. Again, this does not affect the comparison since the impact is the same regardless of the retracker used.

P745 L5: does A valid measurement mean ONE valid measurements? As mentioned above, it would be good to know more about the rejection rates of invalid versus valid measurements.

We use the term “valid gridded measurement” to be consistent with the fact that we are comparing gridded data. Statistics on the rejection rates have now been included.

P746 L8: In what context is SSH mentioned here? How is SSH determined for CryoSat (see above)?

SSH is determined by gridding the sea surface height elevations from a given month. This has now been more clearly explained in the text in Section 5.1.

P746 L19 ff? Correlation: What has been correlated with what, and on what scales? Is this now a point-to-point comparison along the same tracks? Or grid point by grid point?

The correlation is done using the gridded data (grid point by grid point). It is now more explicitly stated in the section that all comparisons are done with the gridded data.

P748 L5: How does 0.57 relate to the values stated on P747 L5?

The values stated on P747 L5 are from the Monte Carlo simulation of the expected correlation given the estimated uncertainties in the data. The correlation of 0.57 on P748 L5 is the correlation which was found between IceBridge and CryoSat-2 for March 2013, while the estimated correlation from the Monte Carlo simulation is 0.60 +-0.04 for the same time period. Thus, for this period the uncertainty estimates appear to be

reasonable and the correlation is as would be expected from theoretical arguments.

P749 L8-11: But there will still be uncertainty from dynamic SSH variations?

Yes, dynamic SSH variations will still be present and errors in this could perhaps overwhelm any gain introduced from using a mean sea surface height. This statement has been amended to account for this fact.

Technical comments:

P723 L 5: impact ON climate
Fixed

P723 L23: clarify limited regional data
Clarified this to refer to the data covered by the submarine cruise tracks.

P724 L24: FROM the use...
Fixed

P 728 L 14: add ERS
Done

P741 L1: What do you mean by frequency range?
Added the word 'electromagnetic frequency range' to refer to the 13.575 GHz +- bandwidth used byt he instrument.

P742 L7: BE added...
Fixed