Summary: This paper investigates whether the considerable scattering of satellite radar altimetry at Ku-Band, namely CryoSat-2, that can be expected to occur at the air-snow interface can be exploited to estimate the elevation of the air-snow interface relative to the ocean surface and hence get an estimate of the total (sea ice + snow) freeboard. To do so a two-layer physical model is used together with least square fitting to obtain a fitted waveform to CryoSat-2 Level 1B data from which the elevations are obtained for lat winter / spring October months of 2011-2017. CryoSat-2 elevations are compared with observations from Operation Ice Bridge airborne topographic mapper for two quasi-coincident OIB-CS-2 overflights, one in 2011, one in 2012. Total freeboard is computed and averaged over the entire period 2011-2017 for the entire Antarctic and discussed and compared with ICESat total freeboard maps. Also, the potential to combine air-snow and ice-snow interface elevations for snow depth on sea ice retrieval is tested. While first evaluation results are promising and suggest that total freeboard derived from CS-2 could potentially be used complementary to ICESat and ICESat-2 data more work is required to better understand the observed differences between CS-2 total freeboard and independent data.

I find this an interesting and important contribution to the existing literature and suggest publication of the research results - provided that the authors take into account the various, partly substantial, suggestions for a major revision of their manuscript. I list my major concerns below in the general comments.
We sincerely thank the reviewer for his/her thoughtful and detailed comments on the manuscript. Particularly, we appreciate the suggestions on ways to strengthen the CryoSat-2/Operation IceBridge comparison as well as the encouragement to include more and (more applicable) references to recent studies on Antarctic sea ice freeboard retrievals. Our responses (in blue) to the reviewer comments (in black) can be found below.

General comments: GC1: The introduction lacks to present the state-of-the-art of seaice and/or total freeboard retrieval in the Antarctic. Several studies exist that are using radar or laser altimeter data for this kind of retrieval. In addition, the introduction lacks to present the state-of-the-art of freeboard-to-thickness conversion and inherent problems and uncertainties. While mitigation of the former lack is required to understand why it might make sense to try to derive total freeboard also from CS-2 data, mitigation of the latter lack is required to understand your attempt to retrieve snow depth on sea ice as well.
The introduction was indeed lacking in the original version of the manuscript. It has been substantially revised to focus on Antarctic retrievals of sea ice freeboard, referencing methods from laser altimetry (Zwally et al., 2008; Kurtz and Markus, 2012; Kern et al., 2016; Li et al., 2018) as well as radar altimetry (Giles et al., 2008; Schwegmann et al., 2016; Paul et al., 2018). Additionally, information on present day freeboard to thickness conversions is included as motivation for retrieving snow depth.

GC2: I don’t find the interpretation of Figures 3 and 4 particularly convincing as a motivation why there is substantial(ly more) information about the air-snow interface in the echograms of the snow radar and the Ku-Band altimeter. See my respective specific comments.
These figures are meant to show that Ku-band returns of the air-snow interface are similar to air-snow interface returns from the snow radar. Since the snow radar is utilized often for snow depth retrievals, similar returns from the Ku-band would provide evidence of Ku-band scattering from

the air-snow interface and motivation to retrieve the snow surface elevation from CS-2. Responses to the specific comments can be found below.

GC3: Even though Figure 6 and the interpretation is intended to stay qualitative (my guess), I strongly suggest to discuss these results in more detail. Putting more emphasis on increasing the credibility of the elevation estimates at this stage is very important in my eyes. You have the unique opportunity to have quasi-coincident air-borne and space-borne measurements. That's luxury and I have to admit that I am a bit disappointed that you do not exploit this luxury situation further. If I'd be allowed to recommend something, then I would i) quantify the temporal and spatial differences in the tracks and try to investigate whether a correction towards a better spatiotemporal match is worth an effort, ii) collocate the tracks with ice-type information (it might be sufficient to figure out where first-year ice and perennial ice was present), iii) collocate the tracks with meteorological information, e.g. from ERA-Interim or MERRA-2 or perhaps even from one of the higher-resolving weather forecast models to figure out whether air temperatures have been close to 0degC and/or whether and what kind of precipitation potentially occurred (Ideally you have a look at the meteorological conditions of not just the day of the coincident measurements but also of a 1-2 weeks period before to catch potential melt events and hence snow metamorphism near the air-snow interface.). See also my respective comments for Figure 6 and its interpretation.
You are correct that this comparison was meant to stay qualitative and be used simply as motivation for progressing to freeboard retrieval using this method. However, we realize it does provide a great opportunity to evaluate the retrieval and will address these points more in the revised manuscript.
We have collocated the tracks with meteorological information (both on the day of and 2 weeks leading up to the flight) and ice types on the day of the flight. Responses to the specific questions, including information on meteorological conditions and ice types, can be found below.

GC4: The interpretation of Figures 7 through 9 would also very much benefit from a more critical discussion which should also involve more work done by other researchers. I find a lack of attempts to explain the differences observed, e.g. in Figure 9. See my specific comments to these figures.
The results were indeed lacking as far as explaining the differences observed, which could be improved by relating the results to work done by other researchers. We will revise the conclusions to include a more robust explanation. Specifics can be found in the comments below.

Specific comments: Abstract: I suggest to add the standard deviation or Root-Mean-Squared difference in addition to the mean difference values given. If computed, also modal values of the difference would allow to give the obtained values more credibility.
Standard deviation values were added to the abstract. Modal values were computed and included with the figures in text.

Page 1 L30: I guess, since you are focussing on Antarctic sea ice it would not hurt do use citations referring also to the albedo observed over Antarctic sea ice: Brandt RE, Warren SG, Worby AP and Grenfell TC (2005) Surface albedo of the Antarctic sea ice zone. J. Climate, 18(17), 3606–3622 (doi: 10.1175/JCLI3489.1) and Zatko and Warren, Annals of Glaciology 56(69) 2015 doi: 10.3189/2015AoG69A574

Agreed – We have removed the citation for Perovich et al., 2002 (which focused on Arctic sea ice) and replaced it with citations for Brandt et al., 2005 and Zatko and Warren, 2015.

Page 2 Line 2: I suggest to cite Comiso et al., J. Climate, DOI: 10.1175/JCLI-D-16- 0408.1 instead of Beitler 2014; the former is a peer-reviewed paper.
The reference was changed.

Line 6: Please add "Antarctic" or "Southern Ocean" to make clear that these shipbased observation based sea-ice thickness data set is valid there but not general in the polar regions.
Added "in the Southern Ocean".

Line 5-13: - Is there a reason why you refer to multiyear ice only for the Arctic? - Is there a reason why you refer to sea-ice thickness in the Arctic only while for the Antarctic you refer to sea-ice thickness in volume? Is the sea-ice thickness retrieval in the Antarctic more accurate so that it makes sense to also derive the volume?
Multi-year ice was referenced because the Arctic study (Kwok et al., 2009) estimated trends over multi-year ice alone, while in the Antarctic study (Kurtz and Markus 2012) did not discriminate. This paragraph was revised and no longer includes these references alone.

Line 14-20: - I am not sure I like the mentioning of Kwok et al. (2009) and Kurtz and Markus (2012) as the role models for sea-ice thickness measurements from active satellite sensors. Since you are basically referring to the principle, wouldn't it be sufficient to simply write what you wrote without these two references? I I guess my dislike comes from the fact that there have been earlier papers that describe how laser altimetry (which is the main focus in this paragraph) can be used to get an estimate of the total freeboard of snow-covered sea ice: Kwok et al. (2004) or Kwok et al. (2006) for the Arctic and Zwally et al. (2008) for the Antarctic.
The introduction has been revised. Zwally et al. 2008 (among other works) has been included for describing how laser altimetry can be used to retrieve total freeboard.

L21-30: - I suggest to re-organize the sentences starting in Line 24 to avoid that sea-ice freeboard is used before being explained. How about you write along these lines: "... 2010-2012. The difference between laser ... [continue until Line 29] ... above the sea surface, known as the "sea-ice freeboard", and is used to calculate sea-ice thickness applying appropriate assumptions (see previous paragraph)."
Thanks for the suggestion – the introduction has been substantially revised and this sections has been changed accordingly.

L31-39: - I suggest to expand rightaway in Line 31: "by the depth and variable vertical structure of the snow on top ..." - In Line 32, I suggest to add that it is not simply more precipitation but "... more and more frequent precipitation ..." - Line 34: "sea ice down near the" –> perhaps better: "sea-ice surface down near or even below the"
- I suggest to break in Line 38 for a new paragraph, starting with "While ... ".
These suggestions have been made in the text, page 2.

- Is there perhaps also the chance that you quantify how large or weak the scattering at the air-snow interface is compared to that at the ice-snow interface? This could make your motivation

stronger about why it might be reasonable to look for the snow surface scattering contribution even in Ku-Band.
This explanation is done in section 3 ('Observed Ku-band scattering of radar from Antarctic sea ice'). I've added "(discussed in section 3)" to page 3 lines 5-6.

I strongly suggest to seek for evidence in the literature about the possible strength of the snow surface backscatter at Ku-Band (at nadir) to underline that it is physically reasonable to use CS-2 SIRAL returns for snow freeboard retrieval. I am stressing this because there exists literature in which one aims for snow-depth on sea ice retrieval by confidently assuming that Ku-Band penetrates to the ice-snow interface and combining it with a Ka-Band radar such as from SARAL AltiKa (Guerreiro et al., 2016). Your attempt is clearly conter-acting their assumptions.
We agree that more literature surrounding the strength of the snow surface backscatter at Ku-band needs to be referenced in this section of the manuscript. In particular, we have added Giles et al. (2008), which deals with CS-2 sea ice elevation retrievals in the Antarctic, and Willat et al. (2010), which shows that the strongest Ku-band return can come either the snow surface, within the snow layer, or the snow-ice interface.
We acknowledge that the dominant backscatter from Ku-band altimetry often occurs within the snow layer on Antarctic sea ice (Schwegmann et al. 2016 and others) and also that it is often approximated to penetrate to the snow-ice interface over Arctic sea ice (Geurreiro et al. 2016, Kurtz et al. 2014 and others). Here, we are introducing the fact that although the dominant scattering occurs below the snow surface, there exists some scattering from the air-snow interface (as shown in Willat et al. 2010) that we can exploit for snow freeboard retrieval. In that regard, we do not believe we are directly counteracting the assumptions of Guerreiro et al., 2016 and other published works, but instead expanding on the utility of Ku-band returns.

Page 3: Line 10: "builds off" ?
Changed to "builds on".

Line 22: I don't understand the mentioning of the "originally 128". What is this for?
Removed.

Line 34-37: The motivation for choosing data from October is clear. You could have stated why you did not also use data from November. Most of the ICESat spring measurement periods last well into November. Here you state years 2003 to 2009 for ICEsat as years with measurements but actually using you are only data from 2003 to 2007. I can understand that the main motivation for this is to use the data produced by one of you. But from NSIDC and potentially also from University of Hamburg you could possibly have obtained ICESat freeboard data for 2003 through 2009, i.e. from an equally long period as you have CS-2 data from. You stated yourself explicitly, that "Seven years of data allows for a longer-term average to be computed"
We felt that because the comparison between CryoSat-2 and ICESat is (and will always be) indirect, that one month of data would be enough to compare to the spring ICESat campaigns. We agree that both months would be better, but feel one was adequate to assess the distribution. In the later years of the ICESat campaigns, the lower laser energy led to questionable freeboard retrievals, and therefore was not processed using this method. Again, we felt that with the indirect nature of this comparison, the time span chosen was long enough. We have, however, more explicitly stated the differences when discussing the results in the revised manuscript.

Page 4: Line 1: - It would be helpful for a better understanding of Figure 3 to mention the frequency range of the FMCW snow radar. - "First, ..." –> Where is the "Second"?
Added both suggestions to page 4, lines 6 and 9 respectively.

Line 8-12: Is this a gridded product? If yes, which grid resolution does it have and what is done to fill the gaps between the ICEsat overpasses?
Added to page 4, line 14.

Line 13-16: You are using sea-ice concentration data obtained with the NASA-Team algorithm. While when choosing a 50% sea-ice concentration threshold is might not really matter which product to use there is published evidence that the NASA-Team algorithm often severily underestimates sea-ice concentrations in the Antarctic compared to the truth - particularly in late winter / spring. You could avoid students and early-career scientists being trapped by your choice by choosing a more appropriate sea-ice concentration product right from the beginning, i.e. based on the ComisoBootstrap algorithm or the Eumetsat OSI-450 algorithm.
Thank you for pointing this out. In the revised manuscript, we have used the Comiso Bootstrap algorithm at the 50% threshold. This has been included in the datasets section (P4L13-16) and in the updated figures.

Line 18: "... altimetry tend ..." –> I suggest to insert: "for ice freeboard retrieval"
Added, thanks for the suggestion.

Line 20: You could cite Willat et al. (2011) here.
Added.

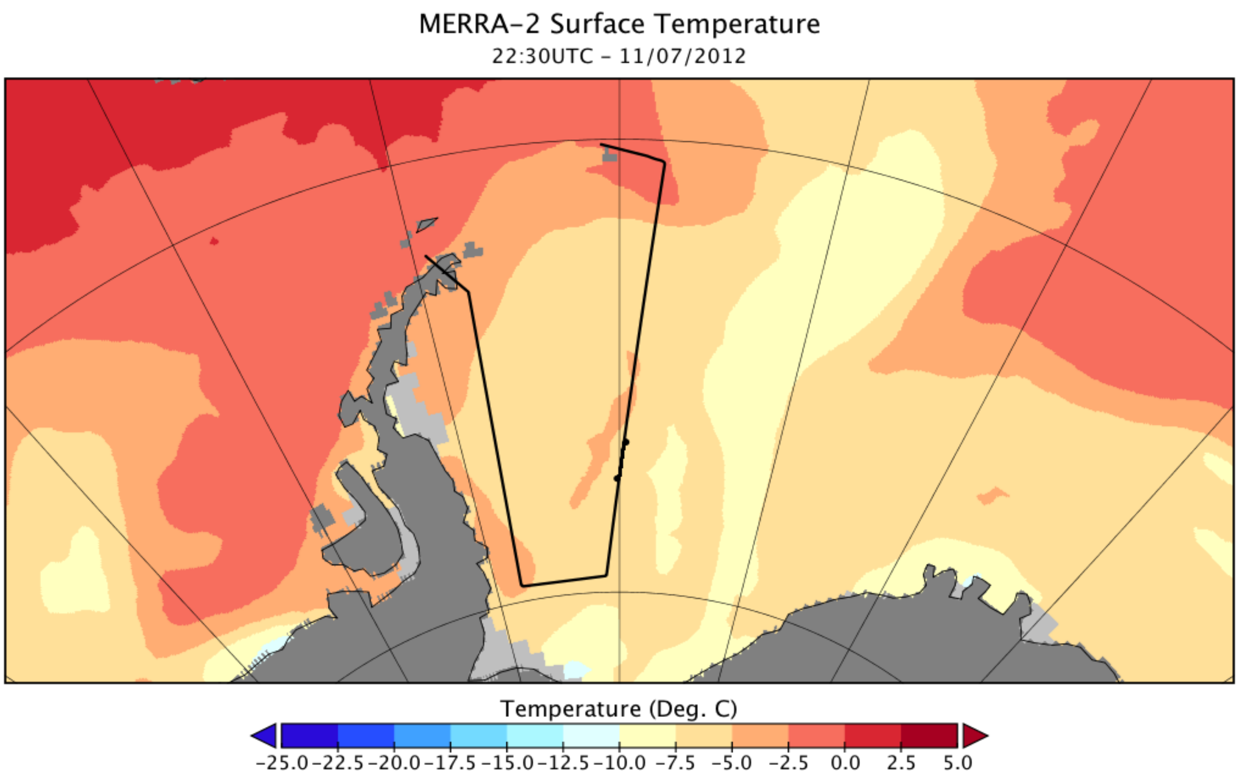
Line 35/36: "This result is expected, as it means that the scattering power from the air-snow interface is closer in magnitude to that of the snow-ice interface in snow radar returns" I do understand your conclusion from the smaller difference in power (about 13 dB for snow radar and 14 dB for Ku-Band altimeter) but I don't understand why this is expected. Lets assume for simplicity that the peak power is 20dB at the icesnow interface for both instruments. Then the power at the air-snow interface would be about 7 dB for the snow radar and 6 dB for the Ku-Band altimeter and with that the power at the air-snow interface would be SMALLER at that frequency from which you assume that the backscattering at the air-snow interface is LARGER. How does this fit together?
What is expected is that the power at the air-snow interface is larger for the snow radar than the ku-band altimeter. This is expected because the snow radar is often used to resolve the two interfaces and calculate snow depth. We do not assume that the air-snow interface power from the Ku-band is larger, but state that because it is similar to the snow radar, that we can expect to have scattering from the air-snow interface in the ku-band.

Page 5: Line 2: At the end of this paragraph interpreting Figures 3 and 4 I have a few questions.
i) How accurate are the two instruments with respect to the dB values shown?
The two instruments aren't radiometrically calibrated, so it is difficult to know the accuracy with respect to the dB values shown (Some radar parameters from each can be found in Panzer et al. 2013 (snow radar) and Gomez-Garcia et al. 2012 (Ku-band, conference paper)).

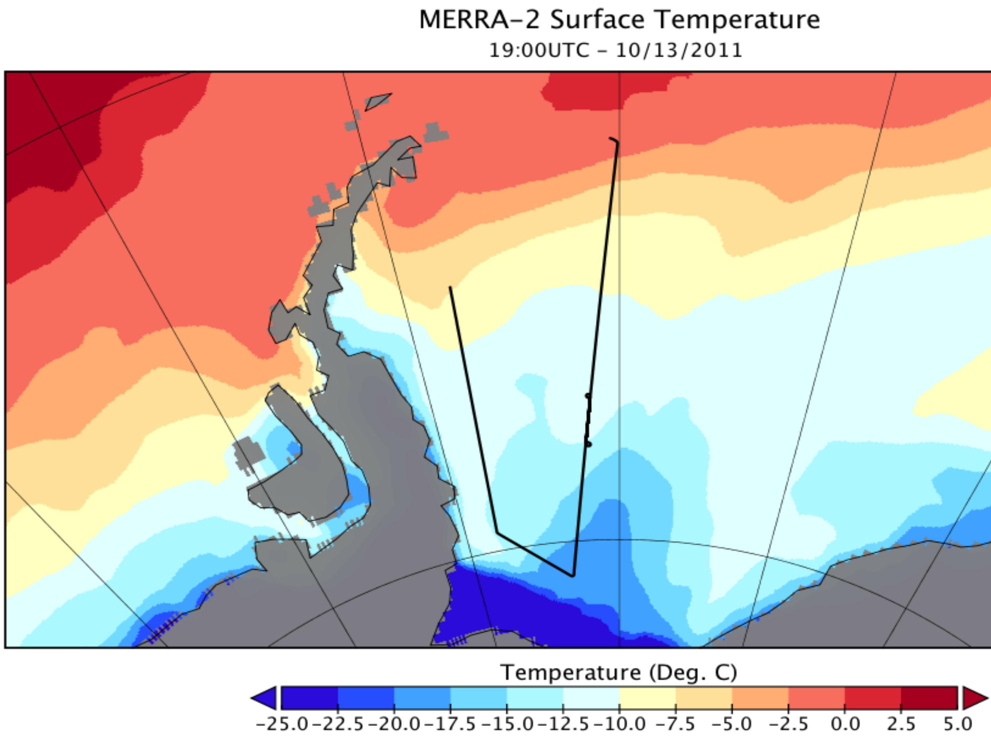
Along the flight lines presented here, the snow radar had a mean noise level of -29.1 dB (std:1.39 db) and a mean air-snow interface signal level of -16.8dB (std: 1.47). The Ku-band had a mean noise level of -30.3 (std:1.32) and a mean air-snow interface signal level of -17.9 (std: 2.09), showing that the return from the air-snow interface is well above the noise level in both cases. We have included these values in the revised manuscript as further comparison.

ii) How relevant is the similarity in the histograms shown in Figure 4 with respect to the shared mode at about 11dB while at the same time the histogram shows secondary modes at 16dB (snow radar) and 20 dB (Ku-Band).
The point of this figure was to show that the histograms are indeed similar to strengthen our case for tracking the air-snow interface from Ku-band. The shared mode is encouraging in that regard, while the secondary peaks could mean that the snow radar is more sensitive to resolving both interfaces when the power difference is larger between the two (as is expected, since the snow radar is used for deriving snow depth).

In addition I have a few comments: iii) What were the meteorological conditions during that flight? Can we expect homogeneous snow properties in terms of snow wetness etc.
iv) How would Figures 3 and 4 look for the October 2011 campaign? Would they result in the same result?
We would expect relatively homogenous conditions during the 2012 flight, judging from the surface temperatures (from MERRA-2, below) and the ice type (taken from EUMETSAT Ocean and Sea Ice Satellite Application Facility) which gives the same “first year ice” classification everywhere along the line.



In 2011, the ice type was again classified as first year ice everywhere along the line, however, there was a stronger surface temperature gradient along the flight line (shown below). This gradient may have resulted in inhomogeneous snow properties that could have altered the radar returns. In this regard, figures 3 and 4 in the 2011 campaign may have looked slightly different than the 2012 campaign and could be responsible for features observed in figure 6.



The meteorological conditions and ice types during these flights will be added and discussed in the revised manuscript.

v) What is the length (in kilometers) of the transect (or echogram) shown in Figure 3?
The echogram here is taken from a single datafile, which in this case covers just over 3km (3.016km). This has been added to the figure description.

vi) The Ku-Band altimeter histogram in Figure 4 has a substantially longer right tail with high dB values. It is almost certain that these values are responsible for the 1dB difference observed between the snow radar and the Ku-Band altimeter data. Are these particularly large differences the result of a particularly low power at the air-snow interface compared to the peak power or is this the result of peak powers being generally elevated at Ku-Band compared to the snow radar?
Agreed – the tail is likely responsible for the difference between the two observed. The differences between the histograms are probably a combination of both, since the snow radar generally has higher power at the air-snow interface (as it's used for snow depth retrievals) and also since peak Ku-band returns tend to be larger than peak snow radar. However, this figure was just meant to highlight the similarity in the histograms, which we feel it adequately does.

vii) What explains the larger time difference between the locations denoted by red and black points in Figure 3 for snow radar echogram compared to Ku-Band?
The snow radar tends to resolve the interfaces better, as can be seen in the more separate "interfaces" tracked bt the peak-picker. The Ku-band, at a higher frequency, tends to be slightly less sensitive to the air-snow interface and also attenuated more by the snow volume, and thus has less-defined interfaces and smaller time differences.

viii) What is the source for the staggered echos above the air-snow interface at Ku-Band? Such echos are not at all present for the snow radar data.
These staggered echoes are returns from radar sidelobes, which are primarily a function of non-linearities in the transmit pulse. Over strong ice returns (seen in the figure) sidelobes can be present in both radar echograms, however in this case, the Hanning time-domain window filter applied to the data removed sidelobes from the snow radar but not the Ku-Band. Our peak-picker is designed to ignore staggered echoes such as these.

Line 21: What is "n"?
The variable "n" is part of the summation ($\sum_{n=0}^{N_b}$) and stands for each of the synthetic beams (N_b) in the sum.

Page 6: Line 2: I guess "thicker snow depth" and "scattering effects from the snow surface" are not as much linked with each other as scattering effects from the snow volume. What makes a thicker snow cover to have more surface scattering than a thinner snow cover?
This line was intended to mean that the entire snow layer (encompassing both the surface and volume) cannot be neglected. It has been rephrased to read "…scattering from the snow layer cannot be neglected…" to reduce confusion.

Line 27: Why "Though"?
Removed "though" and added "however" to line 28 for clarity.

Page 7: Line 16: What is the motivation to only use a different initial guess for alpha in case resnorm is too high after the first fitting attempt?
The motivation to only use alpha comes from the fact that it is the least well-constrained initial guess variable, and that re-fitting with a new alpha tended to reduce the fitting error. While other initial guesses could have been modified and re-fit, we felt that it was not necessary due to the facts that a) they are all empirically derived and b) the added computation cost did not outweigh the benefit.

Page 8: Line 8: Are these PP and SSD thresholds also taken from Laxon et al. (2013)?
Yes, the PP and SSD thresholds are taken from Laxon et al., (2013) though there is scaling factor of 100 difference present in the PP. As the PP was not explicitly defined in Laxon et al., 2013 the values are calculated following from Armitage and Davidson (2014), and SSD comes from the CryoSat-2 product. These references have been added.

Line 12: "seasonal average freeboard datasets" –> perhaps better "datasets of the seasonal average total freeboard"?
Changed.

Line 15: A very good place to cite the paper by Kwok and Maksym, 2014 in J. Geophys. Res., doi:10.1002/2014JC009943
Added.

Lines 23-25: Please explain how the OIB data area used. Did you take data from both flights? Did you average over all valid points? What is meant by "respective surfaces"? Two final question at the end of this section: What happens in the special case of a snow free ice floe? What happens in case of a wet snow surface, where penetration of the Ku-Band into the snow cover is almost zero?
Added some clarification to this paragraph. Data were taken from the Ku-band radar from both flights, and the values found were the average of all valid points. The "respective surfaces" refer to the surface backscatter coefficients of the air-snow and snow-ice interfaces.
As far as the cases you mentioned: From CryoSat-2, it depends on the return waveform parameters (PP and SSD). In both of these cases, the return may be more specular than a typical snow-covered floe and be non-classified or classified as a lead. The chosen thresholds tend to limit misclassifications, but as with any retrieval process, misclassifications can still occur. We have added some more verification figures to show how the algorithm performs over different regions and surface types.

Line 32: "found" –> perhaps better "computed" or "derived"?
Agreed, it has been changed to "derived".

Page 9: Lines 7-13 & Figure 6: - Please provide a measure of the total distance along the tracks shown in Figure 6 a) and b). This would make referencing to certain feature more easy in addition to simply providing an easier interpretation of the spatial scale we are looking at. - I suggest to add a vertical line at zero difference ATM minus CS-2 in images c) and d). - What is the average difference in successive measurements in images a) and b)? - Suggest to use the same y-axis scaling for a) and b) for a better visual comparability of the elevation variations.
The suggested changes will be made to figure 6.
The distance between successive measurements is 0.38km (corresponding to the along-track footprint of CryoSat-2).

- I'd say that the overall agreement, i.e. the large-scale agreement is better for 2011 than 2012. - For 2011 the mean is very close to zero, right. But the modal value is between 5 and 10 cm with CS-2 underestimating ATM elevation. - CS-2 elevations quite often exhibit strong variations in magnitude; in 2011 more during the first third of the track, in 2012 during the first two thirds of the track. These strong variations (or jumps) are as large as about 40 cm and except in one case do not have a counterpart in the ATM elevations. Please try to give an explanation to these.
We believe these jumps are a product of the threshold chosen to represent "good fits" in the retrieval process. Currently, we use a residual value ('resnorm' in text) of 0.3 as the cut off for acceptable waveform fits. Fit waveforms with larger residuals are not included in the retrieval. When they are included, we see more points with anomalous elevation (i.e. more jumps). Lowering the threshold below 0.3 does help to remove some of the jumps, but also removes non-anomalous points and results in worse overall agreement between CS-2 and ATM. Therefore, we still find 0.3 to be the most appropriate threshold, and understand that the jumps are caused by

“bad” fits that are still within our threshold of being “good”. An explanation has been added to the text.

- While in 2011 the large-scale agreement is quite good (you could even stress this impression by adding elevation profiles with largescale smoothing applied), in 2012 there appears to be a systematic under-estimation of the ATM elevation by CS-2. Please try to give an explanation to these as well.

This potentially comes from the smaller footprint size of ATM compared to CS-2. ATM is able to resolve small-scale peaks that would get washed out in the CS-2 return. While it is not as apparent in 2011, there still exists a slight underestimation of ATM by CS-2 in the last third of the profile. This potential explanation will be added to the text.

- You argue that differences between the two elevation data sets might be caused by "initial temporal and spatial discrepancies between the two data sets. Would you be able to quantify these differences? Would it make sense to do a correction of the track of one sensor with respect to the track of the other sensor?

The discrepancies mentioned here mainly refer to the sampling differences between the two instruments, which would be difficult to quantify. The CS-2 footprint is around 1.6km across track and 0.38 km along track, while the ATM icessn footprint is much smaller (~250m across and 30m along). This itself would lead to different ice being sampled between the two instruments, and very likely different mean elevations per shot.

From this initial validation effort, we felt that the agreement was good enough to justify a freeboard calculation, and that it wouldn’t make sense to correct one track with respect to another.

- You write that both datasets "appear to detect similar locations of troughs and ridges along the flight line". I don’t find this statement particularly convincing because there are also many cases where troughs in one dataset and ridges in the other dataset coincide.

Agreed. This statement has been removed from the text.

Lines 19-33: - Line 25: "fewer than five data points" –> What is the distance of successive data points? How many data points would fall into one 25 km grid box at maximum, i.e. diagonal crossing? Can you comment on the data density as well? How many CS-2 overpasses or days with CS-2 overpasses in one grid cell do you have in one month?

Each successive CS-2 point is .38km, which would result in a potential maximum of about 93 data points.

We will add more verification figures that show the density of usable points as well as the lead/floe classification fractions.

Lines 29/30: "Any points within each grid box ..." –> Could it be that this sentence should be placed before the previous sentence? I am asking because the previous sentence already describes the method used at grid level.

We are updating our freeboard calculation following comments from the other reviewer. This sentence has been removed.

– Line 32: "are smoothed" –> Why is this? Why do you do that? Is it because of the gaps between the overpasses? Please state so in the paper.
This is done mainly to reduce noise but also to fill gaps in the data. There are some gaps in the data between the overpasses and in grid cells that have been filtered out.
The following sentence has been added to the end of the paragraph: "Smoothing is applied to reduce noise in the CryoSat-2 data and also to fill in gaps in the data."

Page 10 Paragraph ending in Line 7: - While there is not too much work yet about freeboard distribution from radar altimetry in the Antarctic I still suggest that you consider comparing your results with the results published by Giles et al. (Geophys. Res. Lett., 2008), Schwegmann et al. (Annals of Glaciology, 2015), and Paul et al. (TC, 2018). - While the work of Nghiem et al. (2016) is really interesting and certainly not invalid over parts of the Antarctic MIZ I suggest that you also take into account (and at least mention if not discuss) the potential effect of ocean swell, lower CS-2 data density and hence a larger representativity error, and ice types being different in the MIZ than in the pack ice; a large fraction of the Antarctic MIZ is formed by the often several hundreds of kilometers of pancake ice or cake ice or first-year ice with small floe sizes (< 100 m) for which I doubt that CS-2 is going to provide reasonable elevation and hence freeboard estimates.
We have added many references that were lacking in the original manuscript, and our results have also been compared to the freeboard distributions shown in Paul et al. (2018).
Additionally, we will expand our explanation of the ice edge in the revised manuscript to include effects of ocean swells and surface waves.
As far as pancake ice is concerned, it is likely that this algorithm is classifying these surface returns as lead points or more likely unclassified points. We will add a figure on lead/floe fractions shows to showcase surface characterization in different regions, including the MIZ.

- I note that the distribution of total freeboard shown in Figure 7 is quite patchy and contains several artificial south-north oriented freeboard variations (possibly caused by sampling issues). I note that the freeboard in the southern Ross Sea is indeed lower than further north. However, given the fact that this is an area of extensive new-ice formation and export paired with low precipitation and hence thin snow cover, the freeboard values shown are certainly at the higher end of what is typical there - if not a proper overestimation. Sea-ice thicknesses in the southern Ross Sea are 20 to 50 cm ... total freeboards (without snow) therefore in the range of between 2 and 5 cm and not between 10 and 20 cm as indicated in the maps.
We are updating our freeboard calculations (to be performed on orbit, at the request of the other reviewer) and thus the distribution will differ slightly in the areas mentioned. However, the total freeboard (while given ice thickness) is heavily dependent on the snow depth, which is unknown. Also, these total freeboards in the Ross Sea do compare reasonable well with that found using ICESat. We will add some additional explanation on retrievals over new ice formation, such as that found in the Ross Sea, to the revised manuscript.

Lines 8-15 & Figure 8 - "smallest measured freeboard" –> perhaps better "smallest measured mean October freeboard"
Changed.

- 25.77, 27.6, 12.97 ... I suggest to give these figures with the same number of digits, i.e. 25.8, 27.6 and 13.0.
Changed.

- Showing the mean total freeboard together with the sea-ice area is certainly fine even though, as you stated correctly, it is not too clear why you find a good correlation between these two quantities. However, instead of the sea-ice area one could plot other variables as well. One would be the standard deviation of the mean total freeboard as a measure of the scatter of the mean values. A second one would be to show either the number of 25 km grid resolution grid cells with valid CS-2 data or even the number of individual valid freeboard (or elevation) measurements. Since you have many gaps in the original CS2 data it would potentially be a very interesting additional information. - I note that the maximum inter-annual difference of the mean October freeboard is 2 cm. Is this within or outside the retrieval uncertainty?
We definitely agree that we could plot different variables here, however, the point of the figure was not meant to be further validation of the algorithm, but instead to simply show that a relationship may exist and could be explored further. We have added additional figures / information throughout the manuscript that act to further validate the performance of the algorithm. More comments on the uncertainty of this retrieval method will be included in the revised manuscript.

Lines 17-25 and Figure 9: - I suggest to color the open water in a grey tone (different than Antarctica of course) to ease discrimination between areas with differences close to zero and open water. - I suggest to reduce the range of the differences shown to +/- 30 cm to show more details. The way the range is chosen currently only reflects the larger differences.
We have revised this figure to account for these issues, and will include it in the revised manuscript.

- I find it essential that you mention three things in your discussion of this Figure: i) the larger number of years for CS-2 (7 instead of 5), ii) the fact that the ICESat data cover different time periods with at least 2 of the five years have a substantial if not dominating overlap in time with November and hence conditions changed towards spring (As far as I recall you analysis you did make the effort the average CS-2 from exact those dates from which also ICESat measurements exist.), iii) the fact that we look at years 2011-2017 for CS-2 but 2003-2007 for ICESat, i.e. two different, not overlapping time periods. While this might not have an effect it needs to be stressed once more in the context of this discussion. Finally, iv) one could ask whether you used the same method for averaging the CS-2 data (and filling gaps, extra- or interpolating gaps) than was done in Kurtz and Markus, 2012? Because of these four issues I warmly recommend to delete the last sentence in Lines 24/25.
We tried to emphasize that these were not direct comparisons, but agree that more could be added to fully explain the differences between the two datasets. We used the same method for averaging and filling gaps as was used in Kurtz and Markus (2012), however with an updated freeboard calculation, the averaging will be slightly different in the revised manuscript, though the gap filling will remain consistent.
These points have been added to the description, and the last sentence has been removed.

- When talking about a difference of only 1.9 cm: What are the uncertainties in monthly mean freeboard from CS-2 and from ICESat? Is the difference about the uncertainty?
The uncertainty was not explicitly quantified in the original manuscript, however, the revised version will include uncertainties derived from the freeboard comparison with IceBridge.

- You do not make any attempt to explain the highlighted negative freeboard differences CS-2 minus ICESat in the Weddell and Amundsen Seas. Why?
This has been inadvertently left out of the original manuscript. It will be added to the revised version.

- How do your results compare to the work of independent researchers: Yi et al., 2011, Kern and Spreen, 2015, Kern et al., 2016, Li et al., 2018?
The freeboard distribution results will be compared to other works in the revised manuscript. Specifically, the freeboard distributions will be compared to data from Paul et al. (2018), and qualitatively compared to other works mentioned here.

Lines 27-36, Figure 10: - Figure 10 is indeed quite interesting because the highest "snow depths" are not observed in the Weddell Sea but on East Antarctic sea ice. Puzzling. This is even contradicting your own work (Kurtz and Markus, 2012), where the freeboard maps shown are assumed to represent the snow depth while assuming sea-ice freeboard to be zero. In that work maximum freeboard and hence "snow depth" was observed in the Weddell Sea. What is further interesting is that the "snow depth" is nowhere considerabler smaller than 10 cm - even not in the southern Ross Sea where there is little or no snow on the young sea ice. - I would have found it again very useful, if you would have related the results shown in Figure 10 also with other work, i.e. snow depth based on satellite microwave radiometry (Markus et al., various) or based on ICESat data (Kern and Ozsoy-Cicek, 2016).
We didn't compare these retrieved “snow depths” to other sources, since we believe the snow-ice interface is not being tracked well in this algorithm, and thus these results do not accurately represent the actual snow depth distribution. Further work will need to be done to improve the retracking of the snow-ice interface to retrieve snow depth, if indeed it is possible given the attenuation of the radar signal in flooded snow..

Page 11: Lines 1-4: That the peak-picking algorithm provides snow depth along that flight line which is within 10% of the values published by Kwok and Maksym (2014) is an encouraging result and should be highlighted more.
We agree that it is indeed encouraging, but have done little in the way of close verification of the interface detection for each data file along the flight path. We feel it is best to refrain from highlighting this result until more validation is done.

Lines 4-6: I am pretty sure that this is a frequency issue and not an issue of bandwidth or footprint size: The snow radar used on OIB is a 2-8 GHz radar, right?, while CS-2 operates in Ku-Band.
With the "correct" snow conditions it is very likely that CS-2 does not penetrate down to the ice-snow interface, explaining the considerably lower "snow depth" value estimate from the two elevations. Actually, the OIB snow depths are potentially even higher because of the difficulties to retrieve snow depth in areas of deformed sea ice and on multiyearlike ice reported elsewhere.

You're correct that CS-2 likely does not penetrate down to the snow-ice interface, and this is something we included in explanations in the previous paragraph. Assuming that CS-2 is tracking the correct snow-ice interface (as other works have done) we would still expect the snow radar-derived snow depth to be larger due to the footprint size and bandwidth discrepancies between the two instruments.

Comment: Again I doubt that the precision and accuracy of the data warrants to give mean values with 3 digits = millimeter precision here. I guess 0.29 m, 0.26 m and 0.15 m would do it.
Changed.

Line 8: I suggest to delete "slightly". It is considerably greater.
Removed "slightly".

Line 12: "validating" –> perhaps better "evaluating" or even only "understanding".
Changed to "evaluating".

Line 13: "to better understand the snow depth distribution on sea ice" –> perhaps better: "to use it together with the air-snow interface for snow depth on sea ice estimation."
Changed.

Lines 15-27: - Line 16: "air-snow interface of sea ice" –> perhaps better "air-snow interface of snow on sea ice".
Included "snow on".

- Line 20: "validate" –> "evaluate"
Changed.

- Line 24: "data from comes from" ...?
Removed the first instance of "from".

- Line 22-27: I strongly suggest you revise these conclusions based on the additional analysis, interpretation and discussion that is recommended in the general comments. In particular, figures for standard deviation and potentially also uncertainties should be given in addition to the mean values. One can have a mean value close to zero with one part of the data pairs having -50 cm difference and the other part having +50 cm difference ...
The conclusions will be revised based on the changes mentioned in these comments, and reflect the updated freeboard calculations, figures, and comparisons to other sources.

Lines 28-33: - Line 28: "retrieved ice freeboard" –> you did not really retrieve ice freeboard, did you? You computed the snow depth from the difference between air-snow interface elevation and snow-ice interface elevation.
Correct. This was changed to refer to the retrieved snow-ice interface elevation values.

- Line 29: "lower than typically expected" –> since you did not show any other results about snow depth on Antarctic sea ice - expect for the case of East Antarctic sea ice - it is difficult to follow this statement.

The “lower than typically expected” was broadly referring to other studies that measured snow depth on sea ice, though a citation was missing from this manuscript. The revised manuscript includes a reference to snow depth measurements made from passive instruments (Markus and Cavalieri, 1998).

– Line 31: I agree about the potentially wide-spread flooding of Antarctic sea ice but sea ice is mostly flooded when the ice-snow interface is submerged be low the sea level and in that case an ice-snow interface does not exist in that sense anymore. If it still exists, e.g. through to lateral flooding it is possibly close to zero. It might therefore be more correct to again refer to sea water and brine wicked up into the snow, creating a saline snow - non-saline snow interface which is possibly the interface seen by Ku-Band.
Agreed, we have revised this section to mention the saline snow/non saline snow interface that is likely detected in the Ku-band.

Question: For the location of the air-snow interface you have a-priori information from seasonal mean ICESat snow freeboard. For the ice-snow interface you don’t have any a priori information, do you? This could be one explanation for the sub-optimal performance with regard to detecting the ice-snow interface as well.
This could be one explanation for the sub-optimal performance, yes. Another could be the fact that although the threshold retracking used here has shown to be successful at retrieving the snow-ice interface in other works (Laxon et al., 2013; Kurtz et al., 2014) the addition of more free parameters could lessen the usefulness of the threshold chosen for snow-ice interface tracking.

- Line 34: I guess it would be fair to cite the already existing literature about using CS-2 data for Antarctic freeboard (sea-ice thickness) retrieval: Paul et al., 2018, and change "... observing Antarctic sea ice." into something like: "... observing Antarctic sea ice with satellite radar altimetry in addition to Paul et al. (2018)." Maybe there is even more work out using CS-2 data in the Antarctic? Please check!
We will revise the conclusions to include the Antarctic freeboard retrievals made by other authors, and focus on the novelty of retrieving snow freeboard (as opposed to ice freeboard) from CS2.

Page 12, Line 7: Again I think it would not be too bad to add the work of other authors here to avoid the impression that you are the first on this field: "... for improved retrievals of Antarctic sea ice thickness, complementary to sea-ice thickness retrievals based on the 15+ years long time series of combined Envisat - CryoSat-2 freeboard estimates (Paul et al., 2018)."
The work of other authors, including Paul et al. (2018), have been included in conclusions of this manuscript.

Page 23: Line 23: Giles et al. This paper was in Geophys. Res. Lett. and not The Cryosphere
Changed – thank you.