We would like to thank referee Sara Fleury for her review and appreciate her valuable comments and suggestions. Here, we would like to go through the made comments point by point and highlight our changes accordingly.

Specific Comments:

- The second step of the classification is qualified as "supervised" but for me this means supervised by an operator or guided by some external data. It does not seem to be the case here, so could you precise what you mean by "supervised" and "supervised training"?

In contrast to an unsupervised classification or clustering what we used in the first step, a supervised classification in general involves some kind of predefined or labeled training data. This does not necessarily mean external data and could also be selected manually by an expert or like in our case through an unsupervised clustering as a pre-processing step. In case of the here-used Random-Forest Classifier as a procedure qualified for the term of “machine learning”, the labeled training data from the k-means clustering is used to train the classifier so it is able in a second step to classify previously unknown data into the prescribed classes of ‘ambiguous’, ’lead’, and ‘sea ice’.

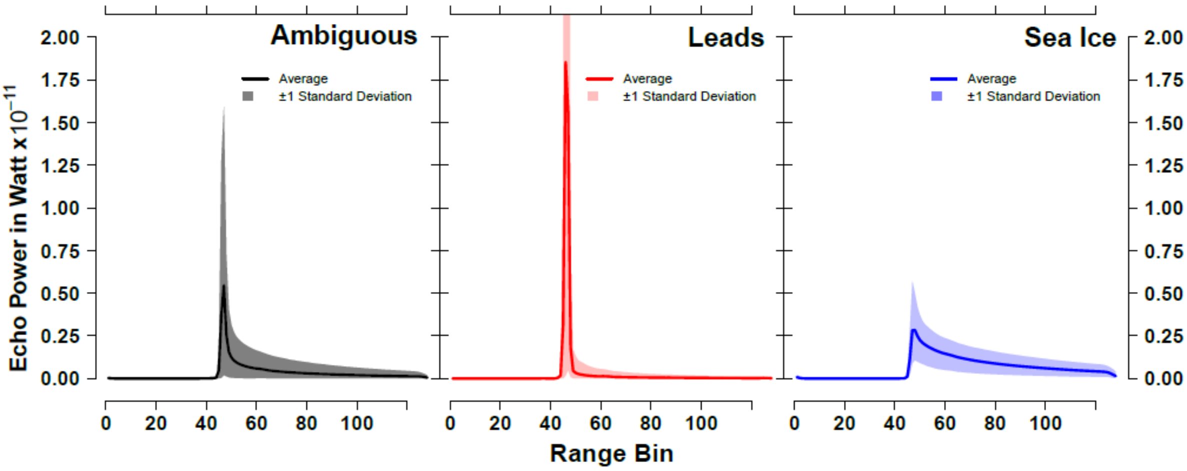
- The initial classification being done on selected surfaces, above 70N and avoiding marginal zones, could not this explains the later-on difficulties for these zones?

That is very likely. However, we decided to do it the way we did because of the much larger noise that stems from the higher degree of surface-type mixing, impact of ocean swell, presence of larger areas of thin and new ice in these areas. Because of these very challenging conditions, a proper fit that also ensures very good quality in the central Arctic would be rather difficult.

- There are no quantitative results of the progresses regarding the classification.

The reviewer is correct, however, we rate the presented results as of a higher importance compared stating specifics about cluster centers. We present both comparisons based on orbit as well as gridded data that can again be compared to the monthly gridded waveform parameters and the resulting freeboard maps.

The following Figure presents the average waveforms for all three classes. These are randomly picked from the Arctic for Envisat. However, compared to other Figures in the manuscript, we rate this information as less relevant for the overall methodology explanation.



- It looks like that you interpolate the heights of floes and leads - and thus the freeboard all along the track, independently from the surface classification or the distance to the nearest lead. Could you confirm this (defendable) strategy?

We are not sure what exactly the reviewer is referring to. We do not interpolate heights of leads and sea ice independent of the surface-type classification. What we assume the reviewer is referring to is that we reject interpolated sea surface height if the distance to the next lead tie point is greater than 200km.

- The impressive correlation obtains between Envisat and CryoSat-2 freeboards should be illustrated in order to make the fitting more demonstrative (or at least providing some other statistical characteristics).

Already several figures are dedicated to illustrate the in general very good agreement between the freeboard estimates from Envisat and CryoSat-2 (the histogram visualizations of all monthly gridded freeboard estimates in Figures 10&11 of mutually covered grid cells, the overall resulting cumulative frequencies of freeboard differences in Figure 12, and the resulting gridded freeboard estimates in Figures 13&14). Therefore, I think there are no further Figures needed and would otherwise just lengthen the manuscript without adding a lot of additional information.

- Some comments on the relative importance of the 3 considered parameters (pp, sig0, lew) for the classification and the range correction for Envisat floes would be appreciated.

By means of classification, all parameters are important, however, especially sig0 and pp show high importance scores in the random forest classification. For the retracker threshold estimation, sig0 and lew showed the most promising results through their capability to capture the seasonal cycle the best and focusing better on the characteristics of MYI (higher surface roughness e.g.).

- Some statements need to be argued (see the Technical Corrections part).

We would like to thank the reviewer for her suggestions and would like to refer to our detailed feedback on those comments in the Technical Comments section.

- Some references should be added for: the product that discriminate FYI and MYI, for OSISAF and for DTU15 (even if well known, it is nice to reference them).

We agree. Please see our changes in the Technical Comments part.

- Because of the distinction between sea-ice and leads all along the study, the expression "sea-ice backscatter" is ambiguous as most of the cases it refers to the "surface backscatter" (ie, a mix of sea-ice and leads). This expression can also be simply replaced by "backscatter" as it is a parameter that characterizes the waveform, like the pp or the lew.

We agree with the reviewer that this term is rather misleading as also the backscatter from lead-type waveforms are currently referred to as sea-ice backscatter. We changed sea-ice backscatter to surface backscatter in the complete manuscript.

- I recommend using the same color-bar theme for the map plots when the purpose is to compare some parameters (of course not necessarily with the same extreme values which depend on the units).

Due to the suggestion of another reviewer we added the pulse peakiness to the Figure Set as well as the resulting Envisat Freeboard. We now use the same color map for all three waveform parameters.

<u>Technical Corrections:</u>

- §1, p.2, l.5: what do you mean by "quasi-nadir" (off-nadir data do not measure the right range) and by "run-time measurements" (data are processed off-line).

The reviewer is right and we removed these misleading terms and changed the sentence to read:

"In a first step, the echo power waveforms are classified as returns from either sea-ice floes or returns from the sea surface of leads between sea-ice floes."

- §1, p.2, l.6: I dont agree with the sentence: "so accurate that one can see the difference in elevation of the snow surface or the sea-ice surface relative to the sea surface on the leads". All the along-track plots of the ranges show terribly noisy measurements, which justify all the studies to classify the surfaces and filter the ranges. At least you should illustrate or quantify this affirmation.

We change the sentence to read:

"These measurements are then converted into distance measurements that let one calculate the elevation difference of the snow surface or the sea-ice surface relative to the sea surface in the leads."

- §2.1.2, p.3, l.27-28: references for OSISAF and DTU15

We added the following references in the text:

"[…] sea-ice concentration data obtained from the Ocean and Sea Ice Satellite Application Facility (OSISAF; ftp://osisaf.met.no/reprocessed/ice/conc/v1p2) as well as the mean sea-surface height product provided by the Danish Technical University (DTU; Anderson et al., 2016; ftp://ftp.spacecenter.dk/pub/DTU15/) in its 2015 version"

- §2.1.2, p.3, l.28: you mean "data filtering"? I suppose you don't analyse the waveforms.

The reviewer is correct; we changed the sentence to read:

"Sea-ice concentration data is used mainly to discard waveforms based on a minimum required sea-ice concentration threshold of 5%, […]"

- §2.1.2, p.4, l.2-4: for me the discrepancies between W99 and the snow depth on FYI is mainly coming from the more and more late development of the new sea-ice in the season due to the global warning that strongly impact the Arctic. This delay limits the possible accumulation of snow on sea-ice. But this worth to be checked.

We agree with the reviewer. This doubt further encourages us to use only 50% of the Warren values over FYI.

- §2.1.2, p.4, l.6-17: could you provide with the name of the used product and if possible a reference?

As mentioned in the manuscript, this product results from the reprocessing of co-author Stefan Kern based on the mentioned data products in order to have a consistent product for the complete combined lifespan of ERS, Envisat and CryoSat-2. So far, there is no specific reference for these data other than a section in the ESA-CCI sea ice ECV phase 2 (SICCI2) ATBD for sea-ice concentration: SICCI-P2-ATBD(SIC), Version 1.0, Sep. 2017.

- §2.2, p.4, l.29-32: could you precise the percentage of removed data ?

This filtering step remains from the processing done in during SICCI-1. While the number of rejected data values is potentially small, flag names suggest that it is better to have them removed nonetheless. We do not capture the exact number of waveforms that are removed in this step.

- §2.3.1,p.5, l.4-6: could you precise what makes you tell that the sea-surface height products are not reliable? Which products?

In accordance with a comment from reviewer 1 we changed that paragraph to read:

"The surface-type classification is a crucial part in the processing chain, because the detection of leads is essential for determining the instantaneous sea-surface height anomaly with respect to the mean sea-surface height at the ice-floe location. The resulting sea-surface height at the ice-floe location in turn is used as the reference from which the sea-ice freeboard is calculated."

The general principle of estimating freeboard from radar altimetry is already summarized in the introduction and does not contribute any benefit at this point in the manuscript.

- §2.3.1,p.5, l.10: I would say more precisely that off-nadir measurements provide wrong ranges.

We changed that.

- §2.3.2,p.6, l.9: with "three classifiers", you mean "three (classifier) parameters"?

The reviewer is correct. We made changes at several occasions throughout the manuscript to clarify the term "classifiers" as "classifier parameters".

- §2.3.2,p.6, l.10: what is the limit for the southern ocean?

This is explained in line 15 on page 6:

"For the Antarctic the same restrictions apply, but waveforms are geographically limited to an area *south of 65°S to exclude the majority of the marginal-ice zone to reduce the impact of ocean swell"*

- §2.3.3,p.8, l.2 and 7: could we state that 1 « 3? What are the possible impacts?

While this suggestion by Breiman was definitely made for larger amounts of input parameters, setting m=1 is the best possible way. While the Random Forest Classifier is capable and powerful enough to deal with very large amounts of input parameters, there is no doubt about its quality using only fewer, but very suitable parameters for the data at hand.

- §2.4,p.9, l.13: please precise the smoothing function that is used.

We clarified this bullet point:

"- Smoothing of the oversampled waveforms with a running-mean window-filter size of 11 (Envisat, CryoSat-2 SAR) or 21 (CryoSat-2 SIN) range bins respectively;"

- §2.4,p.9, l.25: in what way a 50% threshold for leads and floes is "consistent". Why is it more consistent for CryoSat-2 than for other altimeters?

Maybe the word "consistent" was misleading at this point. What it meant is that we follow previous work conducted at AWI and use the same retracker threshold setup (being consistent with that work) that so far showed good results. However, we removed the word at this point in the text.

- §2.4,p.9, l.30: "However" can be removed as the same conclusion is drawn in Guerreiro et al 2017

We changed that.

- §2.4,p.10,F.2: use a unique color-bar

Please refer to our response in the Specific Comments section.

- §2.4,p.10,l.9-10: is there any reason to prefer sig0 than PP ? It could be nice to have also a plot with PP. Visually, the matching with lew is impressive.

As the reviewer mentions, the visual correlation between sig0, pp and lew is quite good, and especially high between sig0 and pp for what we saw (please refer to our updated Figures 2-5). To keep things as simple as possible, which one might argue about in the case of fitting a 3rd order polynomial plane, we decided to stick to as few parameters as possible and chose in that case sig0 over pp for its more direct relationship to surface roughness.

- §2.4,p.10,l.5: the sentence here could let imagine that only one monthly value is used in Guerreiro et al. 2017 to establish the correlation. Perhaps you could remove the 2 words "monthly" or precise that all the monthly cells are used.

We assume the reviewer is referring to page 11, l5 here instead of page 10 as there is no reference to Guerreiro et al. on page 10. We added "monthly-gridded" to clarify that all cells of a month are used for their estimations.

- §2.4,p.12,l.18: Could you also provide R2 which is more frequently used and for which we have more references. A plot showing the distribution and the fitting curve would be very welcome. The correlation is just one characterization, among many others, of the fitting and it is not very intuitive.

Please refer to our answer to your general comment. The here-used adjusted R^2 is always lower or equal to the normal R^2 by definition. We therefore assume there is no valuable additional information from it.

- §2.4,p.12,l.19: could you display the central Arctic region on one of your maps?

The central Arctic region is pretty much what one would expect. However, we agree our text in the manuscript suggests a rather vaguer delimitation. We changed the text in the parenthesis to read:

"[…] (i.e., we excluded the Canadian Arctic Archipelago and the Hudson Bay, but also extensive fast-ice areas like the Laptev Sea) […]"

- §2.4,p.12,l.29: could you show on a map the regions where the sig0 and the lew are less correlated? In particular for the lew it is not so obvious.

Similarly to the reviewers comment above, we clarified this in the text by adding the following text:

"[…] as well as patterns in surface backscatter and leading-edge width are less correlated in some areas (e.g., the MIZ but also in the central Weddell Sea; […])"

- §3.1,p.14,l.12-13 and p.15: Could you provide with some quantitative values to illustrate the progress regarding SICCI-1?

Schwegmann et al. (2016) focus on the Antarctic and a quantitative comparison is rather speculative due to the limitation we have through diversified snow stratigraphy and surface flooding on Antarctic sea ice. However, visual comparisons between both studies suggest a substantial improvement. For the Arctic, Kern et al. (2015) provide an exemplary visualization of the Envisat SICCI-1 freeboard for the Arctic in March 2010. While we do not cover this period in time in this manuscript, visual comparisons between both studies again suggest a substantial improvement

Kern, S., Khvorostovsky, K., Skourup, H., Rinne, E., Parsakhoo, Z. S., Djepa, V., Wadhams, P., and Sandven, S.: The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise, The Cryosphere, 9, 37-52, https://doi.org/10.5194/tc-9-37-2015, 2015.

- §3.2,p.16,l.23-25: it is not clear whereas all the numbers are related to the current study or some of them concern SICCI-1. For instance the "three cm" line 23 seem in contradiction with the "2.2cm" line 25. Could you provide some quantitative comparison with SICCI-1?

The reference to SICCI-1 is based primarily on the paper of Schwegmann et al. (2016) for the Antarctic and some internal analysis that lead to the improvements made in this study. These general impressions and limitations are summarized in the Introduction. There is so far no citable publication concerning the SICCI-1 results for the Arctic. Concerning the 2.2cm statement, we clarified the last sentence to read:

"The overall maximum monthly average freeboard differences is 2.2cm"

- §3.2,p.19,l.10: typo "Shown are the same months".

We changed that.

- §3.2,p.19,l.21: I don't understand the sentence: "In Antarctic, while the differences are lowered, the overall differences remain larger".

We thank the reviewer for pointing this out to us. We clarified the sentence to read:

"In the Antarctic, while the freeboard differences between both sensors are lowered through applying the here-presented methodology, the overall resulting differences remain larger than the ones estimated for the Arctic."

- §4,p.22,l.20-21: and how far are you confident in the AMSRx solution in Antarctic?

It has been shown in a number of publications that the snow depth based on passive microwave data can be substantially biased due to various physical properties of the sea ice and the snow itself, making the retrieved snow depth noisy and unrealiable at times. Using a climatology suppresses this noise. As the focus of this manuscript is on the possibility to match Envisat freeboard retrievals to those of CS-2 ones based on Envisat waveform characteristics (see last paragraph on page 1) we find it justified if not even mandatory to use a consistent snow depth on sea ice data set. We are aware of the fact that using a climatology is not ideal when it comes to the derivation and geophysical interpretation of a sea-ice thickness time series.

.