We thank all reviewers for their extremely useful and constructive comments. We hope that we have been able to satisfy most of the concerns of the reviewers and are happy to provide the revised manuscript now. We have slightly updated the response to the reviewers because the revision of the manuscript changed some passages again after we sent out the response to the reviewers separately.

Thank you very much!

Stefan Kern, on behalf of the co-author team

Response to comments of reviewer #1 to

"About uncertainties in sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise"

by

Kern, S., K. Khvorostovsky, H. Skourup, E. Rinne, Z. S. Parsakhoo, V. Djepa, P. Wadhams, and S. Sandven

In the following we give the comments of the reviewer in *italic font* followed by our response in regular font. We thank the reviewer for the helpful comments.

General Comments:
This manuscript provides an assessment of the quality a RA and RA-2 derived freeboard and thickness data set. The altimeter data set is compared to ice draft from submarines and moorings, freeboard and snow depth from OIB, and snow depth from Warren et al., 1999. There are large differences in these data sets: each data set has limitations and not all of them are understood and there are a lot of issues in the conversion from freeboard to ice thickness. These limitations in available observations is acceptable as long as they are addressed.

There are two key findings: 1) the Warren climatology is not adequate for basin scale usage; and, 2) that ice density could cause significant biases in the estimation of thickness (although this needs to be demonstrated in more detail).

One topic that is missing is a discussion of the limitations/noise of the RA and RA-2 retrievals. With their large footprints, their estimates may be biased due to preferential sampling of areas with lower density of open water (i.e., thicker ice). In fact, that was one of the reasons why the CryoSat-2 design reduced the footprint of the radar by a Doppler processing technique. In any case, it would be important to understand the instrument issues before attribution of the biases to geophysical parameters. Perhaps this is not under the purview of the round robin exercise but certainty this is something that is required in this examination of the uncertainties of thicknesses from radar altimeters.

The reviewer is pointing out an important point. Indeed the title of the manuscript is promising more than we are providing. This was also marked by reviewers #2 and #3. We agree that more work needs to be done to assess the uncertainties in the RA and RA-2 retrievals. This is seemingly a difficult thing to do and most likely is not going to be solved during the first phase of the ESA-CCI sea ice ECV project. Based on this and the recommendations of reviewer #2 we changed the title of the paper to: "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". We further have changed the content of the manuscript such that it is more clear that we are not yet at a stage where we can say too much about uncertainties from RA and RA-2 data.

The text is somewhat difficult to wade through as there are a lot of details and could be improved. Detailed comments follow.

We have made an effort to both shorten the text and to make it more easy to read.

Detailed Comments:
page-line 1521-9: This is not in the reference list at the end of the manuscript.

A citation of this report has been added to the reference list.

1521-21: Just a clarification - is the starting point for the study here the averaged freeboard in these 2 deg by 0.5 deg grids? Please indicate so.

No. The starting point for the study are the non-averaged freeboards. For example the freeboards to compare with ULS of airborne data are calculated from the non-gridded freeboards. Main use for the gridded freeboards is visualisation in Figure 1 as well as comparison with other satellite data such as the AMSR snow depths.
We have re-written this part now: "… several measurements is required. RA-2 freeboard estimates obtained along single orbits are averaged according to the co-location areas defined in section 2.2, or into a 2 degree longitude x 0.5 degree latitude grid (approximately 60 km grid cell size). Averaging is always done over one calendar month. Depending … ."

1521-22: What is the expected uncertainty in altimetric height after the averaging process?

This is much dependent on the number of measurements. The biggest source of uncertainty in a single measurement is speckle, which for RA2 is of the order of 10 cm. The uncertainty due to speckle of an elevation measurement of 100 averaged elevation measurements would then be 10 cm / sqrt(100) = 1 cm. We note, however, that this is just one contribution to the uncertainty and that ice-type and surface roughness influence the altimetric waveform and the radar backscatter an can lead to substantial biases in obtained altimetric height.

1522-1: Are the freeboards and snow depths both 'monthly' averages within these grid cells?

Yes. Grid cells or other areas we have averaged over (such as for example all of the freeboards and snow depths within a certain radius from transect centre then comparing them to ULS data). This is described in more detail in section 2.2.

1522-12: The Baffin Bay area is outside the domain of validity of the Warren data set even though the polynomials could be evaluated in these regions. This is an incorrect use of the data set and thus the statement should be deleted.

We agree with the reviewer. We deleted the statement starting in 1522-10: "but reveals … Baffin Bay."

1523-2: Why is a 12 deg by 30 deg (lat/lon) box required? Please add a statement to justify size of box.

We are aware of the fact that this box is quite large and could possibly be reduced to half the size. The rationale behind it was to maximize the number of individual measurements, i.e. reduce speckle noise and to minimize the effect of drift.

We added: "For BGEP a 12 degree by 30 degree latitude-longitude box is used. This box may be oversized. The rationale behind using such a large co-location area was to maximize the number of valid RA-2 freeboard estimates and to minimize the effect of sea ice motion changing substantially the ice type composition in that area."

1524-5: Minimum should read lowest.

Corrected. We also added "of the magnitude" in front of 0.05 m as we are not talking about an exact value.

1527-6 & Fig. 4: Again, this is not a fair comparison. The W99 climatology is valid only inside the Arctic, it does not extend to the Fram Strait. If you mean that W99 should not be used outside the Arctic, then OK? But, the data set does not claim and is not expected to be useful in regions outside the Arctic.

We do not fully agree with the reviewer here. The Fram Strait area is pretty close to the region for which the W99 Climatology claims to provide reasonable snow depth data. In contrast to areas such as the Bering Sea, Canadian Archipelago and Baffin Bay sea ice is exported out of the Arctic Ocean directly into the Fram Strait area; the time to reach the area shown in Figure 4 is maybe one month. We can assume that during winter the snow depth on sea ice in that area is predominantly determined by the snow depth on the sea ice upstream. This is confirmed by rather similar values of W99 snow depth upstream, in the Arctic Ocean, and in the Fram Strait area. Furthermore, during winter it will be rather unlikely that snow melt would reduce the snow depth from about 40 cm to about 20 cm during the passage from the Arctic Ocean to the Fram Strait area shown. Therefore, because the link between the snow depth on sea ice in the Fram Strait area and the Arctic Ocean is much stronger than the link between the snow depth in the Canadian Archipelago or the Baffin Bay and the Arctic Ocean, we keep the figures and statements related to snow depth in the Fram Strait area. In order to comply with the reviewers' concerns we added the information stated above further down in the discussion of Figure 5. There we now write: "…one in the Canadian Archipelago. For the latter region we only compare OIB and AMSR-E snow depth data in the following because W99 snow depth and density data rely solely on extrapolation in this region. The same applies to the Fram Strait area (see Figure 4). However, the sea ice cover in the Fram Strait area is quite dynamic and originates from the Arctic Ocean while the sea ice cover in the Canadian Archipelago is much more static. Hence it can be assumed that at least during winter sea ice and snow properties in the Fram Strait area are similar to those upstream in the Arctic Ocean, which is actually confirmed by the W99 data (not shown), while those in the Canadian Archipelago are determined by local processes and the sea ice which entered the region during the previous summer season."

We write further:
"… depth; data from the Canadian Archipelago are excluded. Figure 5 b) suggests that W99 snow depths are twice as large as AMSR-E ones over FYI in the Arctic Ocean; the difference… "

1527-6 to 20 (Fig. 5): Once more, W99 should not be used in the Canadian Archipelago Please delete all comparisons with W99 in the Canadian Archipelago.
Fig. 5: Isn't there more extensive coverage of the Arctic by OIB than just the data shown here?

We apologize for not having written clearly what we did.
The maps shown in Figure 5 show only OIB data of April 2010; we omitted to show data from March 2010 and also from March 2009 in this figure for better visibility.
We decided to keep the OIB data from the Canadian Archipelago in the map of Figure 5 c), and added a statement in the text.

Yes, there are more OIB data available. However, the strategy in the ESA-CCI sea ice ECV project was to use only a part of the OIB data in the Round Robin Exercise (RRE) to keep some data for the evaluation of the product at a later stage. Similar comments by the other two reviewers point into the same direction. We deliberately only used a sub-set of independent observations in the RRE.

1528-23: Okay, so the authors recognize that the W99 estimates are extrapolated into these the Canadian Archipelago. So, what's the rationale of showing these results? Is there something to be learned by showing comparisons that are not expected to be valid.

We have deleted all aspects of the comparison between W99 snow depth and OIB or AMSR-E from the manuscript.

1529-1: W99 does show the expected interannual variability, that should be quoted.

We are not sure whether we understood the reviewer correctly because we would tend to see the opposite: While OIB gives 0.36 m and 0.23 m, W99 gives 0.35 m and 0.34 m. We made a statement in the text that W99 does NOT capture the interannual variability. We added the sentence: "It could be that the W99 snow depth climatology does not capture the inter-annual variability in snow depth over MYI in the Arctic Ocean."

1529-7: Was there a reason why ASIRAS did retrieve the ice-snow interface? Is there a reference one could provide?

We don't understand the comment. ASIRAS, as a Ku-Band radar altimeter is supposed to sense the snow-ice interface (similar to Cryosat-2). This was apparently not the case as we found a very good agreement with ALS (laser scanner) data which senses the snow surface. We did not write that ASIRAS retrieved the ice-snow interface.
We added: "at the used frequency in Ku-Band according to laboratory experiments (Beaven et al., 1995)" and further, to account for the fact that this might not be state of the art anymore: "There is growing evidence that this assumption does not hold for more cases the previously thought (e.g. Ricker et al., 2014)."

1529-15: So, perhaps the rationale for using the W99 data should be clarified at the outset otherwise it would be very confusing to try do understand why one would attempt such comparisons.

We have responded this comment further up and have also included some clarification in the text further up. We keep Fram Strait W99 snow depth data and we have deleted the part of the sentence starting with "… and that W99 … 1999)."

Fig. 6: Are these comparisons at all OIB tracks?

These comparisons include OIB data from 2009 and 2010 in the Arctic Ocean. We have added: "OIB snow freeboard observations are compared with RA-2 snow freeboards computed from RA-2 sea ice freeboard and OIB or W99 snow depth in the Arctic Ocean (Table 3, Figure 6)." at the beginning of the paragraph on 1529-17

1529-26: I think this is expected as W99 pertains to multiyear ice.

Sure but still worth mentioning.

Table 4: What is BS 1994, BS 1996, BSS 2007, etc.? Please provide legend.

We added: "See Table 1 for data set acronyms." in the caption of Table 4.

1530-23: Please show that the area is mostly multiyear ice during these years or pro-vide a reference. The statement, as is, is too qualitative.

Actually, there is evidence that the region of interest was covered by a mixture of FYI and MYI in most winters except 2007/08 (e.g. Kwok, et al., JGR 114, 2009; Swan and Long, TGRS 50, 2013). We took the information whether we have MYI in that region or not from the AMSR-E snow depth data product. We added this information in the text after "in winter 2007/2008": "(taken from AMSR-E snow depth data set, Cavalieri et al., 2004)"

1531-5: So, perhaps it is attributable to snow or ice densities. But, isn't it also possible that the RA estimates are biased towards retrieval of higher freeboards?

From our experience with the RA-2 freeboard data we would say no. We rather have problems at the other end, i.e. that we have too much low freeboards because the ocean surface is not referenced properly under the absence of leads.

Table 5, Table 6 and Fig, 9: pretty discouraging for RA-2.
Fig. 9: Which OIB year is this?

This was for 2010. However, this is obsolete now as we have changed Figure 9, and Tables 5 & 6. Reviewer #2 pointed out the unrealistic negative sea ice thickness value from RA-2 data in Figure 9. Triggered by this comment we recognized (VERY LATE, we know) that also a varying number of OIB sea ice thickness values is negative. Because the RRDP does not contain any negative OIB sea ice thickness values we have to assume that something fundamentally went wrong when performing this analysis. As the reviewer will note the agreement has not really improved with the new version but at least we are more consistent now with physics. Note that we have omitted sea ice thickness values computed for the Fram Strait region for the CryoVEx campaigns. This is motivated by the fact that because we don't have an independent ice thickness estimate as in case of the OIB flight but need to compute sea ice thickness from either ALS or ASIRAS data with snow information from, e.g. the Warren Climatology.

1532-paragraph starting at 19: Again, it is possible the RA freeboards, because of the large footprints of the radar, are themselves biased (thicker) in areas with thinner or FY ice? Thus, the signal you are seeing could actually be due to freeboard issues rather than density issues.

We are demonstrating how density issues will cause significant uncertainties in thickness regardless of freeboard uncertainty. So yes, disagreement of RA-derived thickness could be due to freeboard uncertainties too, but that is not the point we are making here.

1533-3 to 15: The authors are too eager to attribute the issues to snow densities. I am not entirely convinced.

We are not sure whether the reviewer is referring to the right paragraph here. We take the note that we should equally weight the contributions of the different factors and to not emphasize only one parameter. We will change the manuscript accordingly.

1534-20: almost all ULS data are acquired under MYI ice? What percentage?

Based on the AMSR-E snow depth data set, of the BGEP data 98% are of MY ice and of the Tireless data 96% are of MY ice. We note that the actual MYI fraction is possibly smaller because we assigned the 100 km area co-located around a 50 km submarine transect already to MYI if one AMSR-E snow depth grid cell is flagged as MYI.
We did not check the fraction for the two U.S. submarine cruises because AMSR-E snow depth data are not available for the 90ties. But we assume that the fraction of MY ice is as high as for the other two data sets.

1536-3 to 24: The argument is difficult to follow: there is noise in all these measurements. There is the underlying assumption that the average freeboard within each grid cell provide the true mean, deviations from that due to sample size and biases in RA freeboard could be issues. It would be useful to convince the reader of that before discussion of ice densities.

This paragraph received comments from the other reviewers as well. We have shortened it substantially and removed the dubious parts.

1536-15: Should you obtained the same ice density that OIB used? I am confused.

See comment to 1536-3 to 24

1538-5: This is over seasonal ice?

No. We moved the part “This results is … same direction” up so that it comes behind “…within 0.12 m in 2010.”

1539-5: the length scale issues are important as there could be significant gradients in freeboard within your large grid cells.

We edited the manuscript and rewrote parts of the discussion and conclusion. Yes, we agree, within the 100 km grid cells or the 100 km diameter discs used for the RA-2 data there can be large gradients in sea ice freeboard. On the other hand, as we collocate RA-2 data at the center of each 50 km transect segment (either OIB or submarine or CryoVEx) these gradients contribute to both values. If, e.g., the OIB track is parallel to isolines of similar sea ice freeboard in the RA-2 data, then the OIB 50 km average should give the freeboard of that isoline. Of course in the RA-2 disc there needs to be a Gaussian distribution around that isoline to make the mean RA-2 freeboard similar to the mean OIB freeboard. If the OIB track crosses these isolines then both sensors sense the same mean freeboard and also have a more similar freeboard distribution within the averaging area (which is not investigated here though).

1540-1: The issues of radar biases that is ice type dependent need to be understood as well.

Noted. We agree with the reviewer. We did not investigate this in the current paper and refer to those recent papers where light has been shed on this issue.

Response to comments of reviewer #2 to

"About uncertainties in sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise"

by

Kern, S., K. Khvorostovsky, H. Skourup, E. Rinne, Z. S. Parsakhoo, V. Djepa, P. Wadhams, and S. Sandven

In the following we give the comments of the reviewer in *italic font* followed by our response in regular font. We thank the reviewer for the helpful comments.

General comments

The authors present an analysis of the effects of snow depth, snow density and ice density on the retrieval of sea ice thickness from pulse limited radar altimetry. They conclude that the Warren climatology is no longer representative of snow conditions, in line with several other studies. They also make recommendations regarding the sensitivity of ice thickness to ice type dependent sea ice density.

One of the main objectives of the study is to 'characterize the uncertainties in the sea ice thickness product based on the uncertainty of the input parameters'. However, as has already been mentioned by another reviewer, the authors do not present an analysis of the error on the freeboard retrieval itself. This is not an insignificant source of error or possibly bias and is dependent upon several factors including the radar speckle noise, the local sea level interpolation (the abundance of lead measurements), radar penetration into the sea ice pack, filtering of contaminated waveforms, retracker noise/biases as well preferential sampling of larger (thicker) floes. See Ricker et al, 2014, 'Sensitivity of CryoSat-2 Arctic sea-ice volume trends on radar-waveform interpretation',The Cryosphere, 8 1831-1871 for a more detailed discussion of freeboard retrieval uncertainty. The abundance of lead measurements, and sea level interpolation error, in particular is considered to be a significant source of error even for CryoSat-2, and will be more severe for pulse limited systems. Kurtz et al, have also looked in some detail at the freeboard retrieval (again for CryoSat-2) in their paper "An improved CryoSat-2 sea ice freeboard and thickness retrieval algorithm through use of waveform fitting", The Cryophere, 2014, 8, 721-768. There is also the issue of off-ranging to leads biasing the sea surface elevation low, which was examined for CryoSat-2 by Armitage et al, 2014, 'Using the Interferometric Capabilities of the ESA CryoSat-2 Mission to Improve the Accuracy of Sea Ice Freeboard Retrievals' IEEE Transactions on Geoscience and Remote Sensing. Similarly, this effect can be expected to be more severe for pulse-limited systems. Unless this paper is extended to address the impact of the freeboard retrieval uncertainty I feel that the scope of the study should be limited to 'examining the impacts of snow depth, snow density and ice density on ice thickness uncertainties', or something similar. In effect, what the authors have presented thus far is an assessment of the sensitivity of ice thickness retrieval to snow depth and the density of ice and snow that is somewhat independent of the altimeter measurements *themselves.*

We take this note and thank the reviewer for this suggestion. Based on this we indeed changed the title of the manuscript to: “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”.
It is beyond the scope of our current capabilities to come up with an uncertainty estimation of RA-2 and RA data which would satisfy the readership. This needs to wait until a later stage of the project.

The manuscript is in general well written and the figures are clear however the discussion section, in particular, is quite long-winded and hard to follow. I feel that this section could be trimmed down to 2/3 or even half of it’s current length without a significant loss of content.

Noted. The revised version of the manuscript will be more concise in the discussion section.

Detailed comments
P1519 L16: Given the claim that this is the first time a combined time series of ERS1/2 and Envisat data has been constructed, has an inter-satellite comparison been performed to check for inter-satellite biases? These could arise from e.g. differences in the on-board tracking systems. An inter-satellite comparison could be indicative of the accuracy of the freeboard retrieval algorithm and is fairly standard practice when looking at multi-mission time series.

Yes, we agree that this should be done. However, at the time of writing the manuscript we did not have the re-processed REAPER data of the ERS1/2 RA data in our hands so that this part of the work needed to be postponed to a later stage of the project.

P1521 L9: What does this stand for?

We have included “(SICCI project)” in line 2 of the introduction now. ATBD stands for Algorithm Theoretical Basis Document. We have added a reference in our reference list for clarity.

P1521 L22: The number of freeboard measurements per grid cell affects the thickness uncertainty as well as the use of external datasets like snow depth & density. This should be addressed in this paper.

The biggest source of uncertainty in a single freeboard measurement is speckle, which for RA2 is of the order of 10 cm. The uncertainty due to speckle of an elevation measurement of 100 averaged elevation measurements would then be 10 cm / sqrt(100) = 1 cm. We note, however, that this is just one contribution to the uncertainty and that ice-type and surface roughness influence the altimetric waveform and the radar backscatter and can lead to substantial biases in obtained altimetric height. Also different floe sizes are likely to introduce biases. But this is not what we are after in the current study. We are after finding a proper set of input parameters for the freeboard-to-thickness conversion and assume – in a way – that the freeboard obtained from the RA-2 data is correct. Actually, the comparison between OIB total freeboard and total freeboard computed from RA-2 sea ice freeboard using OIB snow depths point into the direction that this assumption is perhaps not the worst one. It was the nature and the rationale of the Round Robin Exercise to do this. The rationale of the RRE was not to play around with different re-trackers to obtain different realizations of RA sea ice freeboard but rather to check out how reliable are the input data sets and assumptions of constant densities etc. The validation of sea ice thickness will happen in a later part of the project.

P1522 L3-13: As has been pointed out by another reviewer, the Warren climatology cannot be considered to be valid in Baffin Bay, the Canadian Archipelago or Hudson Bay (as well as the Fram Strait on P1526 L7, figures 4&5) as the polynomial fits are not constrained by measurements in these regions. See figure 1&3 from Warren et al, 1999, for an indication of where there are observations in their climatology. It is not justified to present comparisons in regions where the data are not valid.

We have answered the concerns of reviewer #1 in a similar way as we will do here. The Fram Strait area is pretty close to the region for which the W99 Climatology claims to provide reasonable snow depth data. In contrast to areas such as the Bering Sea, Canadian Archipelago and Baffin Bay sea ice is exported out of the Arctic Ocean directly into the Fram Strait area; the time to reach the area shown in Figure 4 is maybe one month. We can assume that during winter the snow depth on sea ice in that area is predominantly determined by the snow depth on the sea ice upstream. This is confirmed by rather similar values of W99 snow depth upstream, in the Arctic Ocean, and in the Fram Strait area. Furthermore, during winter it will be rather unlikely that snow melt would reduce the snow depth from about 40 cm to about 20 cm during the passage from the Arctic Ocean to the Fram Strait area shown. Therefore, because the link between the snow depth on sea ice in the Fram Strait area and the Arctic Ocean is much stronger than the link between the snow depth in the Canadian Archipelago or the Baffin Bay and the Arctic Ocean, we keep the figures and statements related to snow depth in the Fram Strait area. In order to comply with the reviewers' concerns we added the information stated above further down in the discussion of Figure 5. There we now write: "…one in the Canadian Archipelago. For the latter region we only compare OIB and AMSR-E snow depth data in the following because W99 snow depth and density data rely solely on extrapolation in this region. The same applies to the Fram Strait area (see Figure 4). However, the sea ice cover in the Fram Strait area is quite dynamic and originates from the Arctic Ocean while the sea ice cover in the Canadian Archipelago is much more static. Hence it can be assumed that at least during winter sea ice and snow properties in the Fram Strait area are similar to those upstream in the Arctic Ocean, which is actually confirmed by the W99 data (not shown), while those in the Canadian Archipelago are determined by local processes and the sea ice which entered the region during the previous summer season."

We write further:
"… depth; data from the Canadian Archipelago are excluded. Figure 5 b) suggests that W99 snow depths are twice as large as AMSR-E ones over FYI in the Arctic Ocean; the difference… "

P1523 L10: Provide references for this statement i.e. lab work by Beaven et al 1995. It should also be mentioned that this idea has come under question, in particular by Ricker et al, 2014 (above), and justification for continuing use of this assumption should be made.

We added the suggested reference and also added the sentence: "There is growing evidence that this assumption does not hold for more cases the previously thought (e.g. Ricker et al., 2014)."

P1523 L15,26: Do you use different averaging distances for the different datasets? If so, why? Or is this a typo?

We do use them. The reason for a shorter averaging distance for FS flights is that they were much shorter than OIB flights. 100 km would had been optimal to use for FS too, but that would had resulted in 5 and 10 data points for years 2008 and 2010 respectively. For OIB we have thousands of kilometres of track available and we can use a longer averaging span.

P1523-24: How come airborne Electromagnetic sounding (i.e. EM Bird) data has not been used? This could be a useful comparison as it measures the ice thickness directly.

We agree. This is driven by the nature of the SICCI project and the layout of the Round Robin Exercise. It was decided to keep data such as EM-Bird measurements or ICESat sea ice thickness data for the validation of the product. We would like to mention though, that EM-Bird data provide the total (sea ice + snow thickness) and are thus not a direct measurement of the ice thickness. Again additional information would be needed such as snow depth and / or the location of the ice-snow interface as can be obtained, e.g., from the OIB data.

P1525 L15: Please see my above comments about addressing the freeboard uncertainty. As the paper stands you have not addressed all of the input parameters.

We agree. We refer now to the fact, however, that we like to follow the suggestion of the reviewer to let the paper indeed focus on the impact of snow depth, density and ice density on ice thickness retrieval using radar altimetry and not try to provide some half-ready information from our preliminary investigations about ice freeboard uncertainties based on RA and/or RA-2 data.

P1526 L9,13: Where are these values taken from?

The sea ice density value of 900 kg/m³ is simply the arithmetic mean of the densities of FYI and MYI given further down. A reference for the water density used is now given.
Snow densities 240 and 340 kg/m³ have been chosen according to the mean seasonal cycle of the snow density according to Warren et al., 1999. We tried to make this more clear in the revised manuscript. It reads now:
"For the standard computations as given above under C) and D) the following values are used: ρi = 900 kg m-3 (mean of the densities of first-year and multiyear ice) and ρw = 1030 kg m-3. For multiyear and first-year ice we use sea ice densities published elsewhere (e.g., Timco and Frederking, 1996; Alexandrov et al., 2010): 882 kg m-3 and 917 kg m-3, respectively. Snow density varies over space and time in the W99 data set (see Fig. 1 e, f). For C) we vary snow density values accordingly between 240 kg m-3 and 340 kg m-3."

P1526 L15: Both radar and laser altimetry techniques are sensitive snow depth, snow density and ice density, but in different ways. This is not clear from your statement.

We agree. We found that the paragraph "Note that the sea ice … according to W99" is not required here and deleted it from the manuscript.

P1527-1528: See above comments on the Warren climatology validity.

See our reply to this topic above.

P1529: See comments on Warren climatology validity. Also, considering that you have shown the Warren snow depth is about double that of OIB and AMSR-E snow depths, it is rather trivial to say that the altimeter 'snow freeboard' is higher than the campaigns freeboard when you apply the Warren climatology. The rationale for converting altimeter freeboard to a 'snow freeboard' should be clarified anyway, or else left out and focus on the sea ice thickness.

The rationale behind this was that we wished to inter-compare RA sea ice freeboard values as much as possible with independent data. As there are no other sources for sea ice freeboard (except ASIRAS which unfortunately failed) we needed to find a way to compare the data at a different level. OIB provides snow freeboard. This is a direct measurement. OIB provides snow depth. This is also a direct measurement but more uncertain than the snow freeboard measurements. Hence our strategy was not to fiddle around with OIB data and claim that the sea ice freeboard one can retrieve from two uncertain products can be compared well with the RA sea ice freeboard but rather look how close snow freeboard computed from RA sea ice freeboard – which we assume to be correct in this case – comes to the observed snow freeboard.
We would like to again stress that the manuscript does not aim for the validation of sea ice thickness obtained from RA data. We are not there yet. The manuscript aims at providing useful information about the limitations and caveats of utilizing RA data. Validation / inter-comparison of RA sea ice thickness will come at a later stage. At the time of writing RA-2 sea ice thickness was not yet retrieved.

P1531 L5: The seasonal range of ice draft derived from altimetry will also be influenced by the seasonal range of snow depths and densities that are applied, as well as the FY/MY ice densities. Would it be possible to examine the effect of using different snow depth/density values as well as ice density values? If not, this should at least be mentioned.

We hoped that Figure 8 will give enough information into this direction as we here "play" around with both different ice densities (MYI and FYI) and snow densities.

Figure 9: Whilst these results are not particularly encouraging for RA-2 thickness retrievals, I am still not convinced about your freeboard retrievals and data filtering given that you are reporting such large negative thicknesses. Considering that these are monthly, gridded and smoothed data you would not expect to see RA-2 ice thickness of -2.5m. Negative ice freeboards are understandable from a data processing point of view (and also can occur in reality) but on the spatiotemporal scales that you are presenting the data one would not expect to see these large negative thickness values. This requires some explanation.

We agree. We thank the reviewer for pointing this out to us. In fact, when reviewing Figure 9 we became less worried about the one negative RA-2 sea ice thickness value but more worried about the varying number of also negative OIB sea ice thickness values. We really should have captured this MUCH earlier and we apologize that we have been that sloppy in our assessment of the figures. However, as the RRDP does not contain negative sea ice thickness values from OIB flights we have to assume that something with the analysis went completely wrong when performing this analysis for the first version of the manuscript. Meanwhile we repeated the analysis and are coming up with a new version of Figure 9 and changed numbers in Tables 5 and 6. Note that we have omitted sea ice thickness values computed for the Fram Strait region for the CryoVEx campaigns. This is motivated by the fact that because we don't have an independent ice thickness estimate as in case of the OIB flight but need to compute sea ice thickness from either ALS or ASIRAS data with snow information from, e.g. the Warren Climatology.

P1532 L3-9: In line with the points made by another reviewer, and by myself, I think that any comparison between RA-2 and CryoVEx data in the Fram Strait that uses the Warren climatology in any way should be left out. This includes the "W99", "AMSRE+ W99" and "KF11" results.

As we have pointed out further above: we are pretty sure that the W99 snow depth data are still valid in the Fram Strait area because of a) its proximity to the upstream area of sea ice from which its sea ice cover originates, b) W99 snow depth data are quite similar between the two regions (north of Fram Strait and Fram Strait area), and c) the ice export takes place on the scale of weeks to 1-2 months which, during winter, does not justify large snow depth changes between the upstream area and the Fram Strait area. We therefore have kept the comparison.

P1535 L25-P1536 L2: I think this is an important point. It should perhaps be put into the previous section where you talk about the sea ice thickness results, to add some balance to the very negative results that you have presented! I think it is almost certain that OIB is much more sensitive to the small-scale range of ice thicknesses – it will pick up pressure ridges where RA-2 will likely not, and will be sensitive to smaller, thinner floes where RA-2 will not.

Agreed. The revised manuscript will contain a corresponding discussion into this direction.

P1336 L10-24: I do not understand what you are trying to demonstrate here. You seem to be calculating the sea ice density from the ice thickness equation, using values of ice thickness that already assume a particular ice density? What is the benefit of doing this? The derived ice densities are not meaningful since they depend on the density assumed to derive the OIB ice thickness.

This was pointed out by reviewer #1 as well. We have considerably shortened this paragraph.

P1538 L18-22: Again, I would emphasize that this over/under estimation of ice thickness does not necessarily mean that either OIB or RA-2 are incorrect. Rather it is probably indicative of the spatiotemporal differences between the two retrieval methods – i.e. OIB being more sensitive to e.g. ridging and smaller, thinner floes, RA-2 measuring over one month.

We agree with the reviewer and tried to formulate this more clearly in the revised version of the manuscript.

Response to comments of reviewer #3 to

"About uncertainties in sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise"

by

Kern, S., K. Khvorostovsky, H. Skourup, E. Rinne, Z. S. Parsakhoo, V. Djepa, P. Wadhams, and S. Sandven

In the following we give the comments of the reviewer in *italic font* followed by our response in regular font. We thank the reviewer for the helpful comments.

The authors present uncertainty analysis of satellite altimetry data of a combined time series of ERS-1/2 and ENVISat freeboard data in the Arctic. They use airborne, moored and submarine validation data to validate measurements and/or parameterizations of freeboard, snow-depth and densities of sea ice and snow. Their main findings are summarized in a list of 4 recommendations, which focus on the correct choices for snow depth and sea ice density.
The paper is well motivated and I fully agree with the authors that a validation of pulse limited radar altimetry is a pending issue, as well as a consistent conversion of (either ice of snow) freeboard into thickness over the range of existing and future altimetry missions. I am however not fully convinced that the methodology and chosen datasets do the job.

My main concerns can be summarized in three categories:
1) Pulse-limited radar altimetry
The authors focus too little on the potential biases freeboard from radar altimeters may have, especially the low resolution pulse-limited systems. The influence of physical snow properties and surface roughness on the radar range retrieval is not yet sorted out even for higher resolution data (CryoSat-2). Part of this problem is visible in the CryoVEx validation data the authors use here, where the difference of ASIRAS and ALS freeboard appears negligible. It seems therefore premature to use radar freeboard from ERS-1/2 and ENVISat without a bias analysis and contribute differences of satellite and validation data to assumptions of snow depth and ice density. Recently submitted studies of Kurtz et al. and Ricker et al. (in TCD) have shown how large the impact of radar waveform interpretation on freeboard retrievals can be.

We agree with the reviewer. His/her comments to this topic are pretty much in line with the comments of the other two reviewers. Here we oversell the work which has been done by the consortium. Basically we have only looked at the impact of snow depth, snow density and ice density on sea ice thickness based on sea ice freeboard obtained from radar altimetry but we have not quantified at all any uncertainties in the sea ice freeboard we have obtained. In terms of the sea ice freeboard the manuscript describes an inter-comparison with

independent data rather than discussing uncertainties. Therefore we followed the suggestion of reviewer #2 and changed the title to:
"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".

2) Sea-ice thickness validation data
The authors use airborne altimetry datasets as validation data for sea-ice thickness. In the manuscript is this partly done in a confusing way (please see my detailed comments below). But since the authors acknowledge that the conversion of freeboard to thickness is yet containing significant uncertainties, even high resolution altimetry data cannot be used as a reference for sea-ice thickness, but only freeboard. If the satellite and airborne freeboard is converted into thickness in a consistent way, there is no gain in comparing more than freeboard (except visualizing the impact of freeboard differences in units of thickness). An independent validation of sea-ice thickness requires datasets which either directly measure sea-ice thickness (EM induction) or are much less affected by the uncertainties of densities and snow depth (e.g. ULS draft data used in this study).

The rationale of the Round Robin Exercise was not to validate a sea ice thickness product. This would have been the case if there would be at least a handful of different algorithms to compute sea ice thickness from radar altimeter data – like it is the case for sea ice concentration retrieval algorithms. So instead of trying to validate results of an approach from which we don't know whether it is the correct one (taking all the assumptions and input parameters into account) it was decided to rather carry out the comparison one level below. This means that we look into the consistency of the obtained sea ice freeboard in comparison to independent data (like total freeboard or sea ice draft). Furthermore this means that we look how reasonable the input parameters (and assumptions) used are. This explains why we do not consider EM-Bird data or ICESat data. To our humble opinion EM-induction sounding is as good or bad as OIB data when it comes to ice thickness estimates. EM-Bird data provide the total (sea ice + snow) thickness by combining the EM technique with a laser scanner but do not provide information about the location of the snow-ice interface. OIB data measure total freeboard and provide an estimate of the snow-ice interface. In both cases assumptions need to be made to obtain the actual sea ice thickness.

3) Choice of datasets
It is my guess that the choice of data package in this paper was chosen at the beginning of the ESA CCI project. To my knowledge, the authors use only subsets of the airborne campaigns and especially the lack of comparisons against ICESat freeboard map is a missed opportunity in terms of consistency between missions. Even a comparison of CryoSat-2 and ENVISat would be possible in in early 2011. Other data sets (e.g. sea-ice thickness from EM-induction) are very briefly mentioned in the introduction but never used again.

We refer to the comment we made above and add the following. Indeed the goal of the Round Robin Exercise (RRE) was not to validate an ice thickness product. This is going to happen at a later stage of the project. For this reason we kept a) EM-Bird data, b) more recent OIB data, c) more submarine and moored draft data, and d) other satellite data such as ICESat for the validation part of the project when we are after the validation of the thickness product. The goal of the RRE was to find an appropriate set of input parameters for the freeboard-to-thickness conversion on the basis of the freeboard and the input parameters. We have added the following sentences for clarification in the introduction:
"The main goal of the RRE is not to validate a sea ice thickness product but rather to carry out a consistency check of the sea ice freeboard data obtained from satellite RA. Another important part of the RRE is the investigation of the quality of the data used and the estimation of the sensitivity of the methods used to the input parameters. The goal is to find

an optimal set of assumptions and input data for the freeboard-to-thickness conversion – assuming that the RA sea ice freeboard is correct. Validation of sea ice thickness obtained from these RA freeboard data will be carried out at a later stage of the SICCI project. This is the reason why a number of data sets one would expect to be used in this study, such as e.g. sea ice thickness derived from ICESat data or total (sea ice + snow) thickness derived from electromagnetic (EM) induction sounding are missing in the present study. For the same reason we did not use more recent Operation Ice Bridge (OIB) data."

We added another paragraph at the end of the data section: "Many other validation data are available, e.g. for ULS from submarines and the BGEP array, more recent OIB flight data, observations of total (sea ice + snow) thickness from electromagnetic (EM) sounding (Haas et al., 2008; 2010), and from ICESat (Kwok et al., 2009). The reason for not using this data for the present study lies in the nature of the project. We deliberately kept a large number of validation data for a later stage of the SICCI project to validate the prototype sea ice thickness product."

Summarizing, the authors do not provide a convincing case for their recommendations. Mainly because the validation of the ERS-1/2 and ENVISat data lacks radar altimetry specific biases and the choice of validation data sets is limited. The sensitivity study for radar freeboard to thickness conversion and the comparison of snow-depth products does not produce novel insights than earlier publications from Giles, Kurtz, Kwok and others (all cited in the study).I also feel that often the comparison of satellite and validation data is not explained well enough.
But I definitely see the need and importance to extend the time series of Arctic freeboard data with the early pulse-limited data and I would strongly recommend that the authors focus their study on an estimation of freeboard bias and uncertainty of the pulse-limited radar systems and how they relate to other missions that complement (ICESat-1) or extend the time series (CryoSat-2). The consistent conversion to ice thickness would be only the next step after the sensor specific biases between different missions are approximated.

The novel insight of the manuscript is that we make the reader aware of

a) the limitations of using a RA freeboard product with a grid resolution of the order of 50 km to 100 km for sea ice thickness retrieval
b) the difficulty to use the suite of available evaluation data to decide which parameter combination is best for freeboard-to-thickness conversion (no, we cannot yet state that using an ice-type dependent sea ice density makes THE difference because the validation data do not yet allow us to quantify what the improvement is)
c) that as long as densities are used in an inconsistent way no appropriate uncertainty estimation can be carried out

Minor points:
The title includes the term "Round Robin Exercise". Have the different approaches been taken out independently by the different co-authors? A short explanation would be valuable.

Yes. We have added information explaining this at the end of the introduction: "We note that the results presented reflect the work of the SICCI project consortium and have been carried out in the respective institutions."

The Discussion section is long and mixed with "Results". Also, the Summary & Recommendation section should be only a Summary, with the Recommendations following anyway.

We tried to condense the manuscript and be more concise.

Detailed comments:
P1520 L03 ff Please define "precise". One of the main points of this study is that freeboard to thickness conversion has not been always done in a consistent way and depends on the choices of densities and snow depth.

We agree but are not sure whether we understand the reviewer correctly here. We take his/her comment as a hint that also in the validation data sets we listed assumptions are made and that these data sets might not all be free of a bias and/or carry a substantial uncertainty; we added: "We note that for all methods mentioned in the previous three paragraphs assumptions need to be made about, e.g., penetration depth of radar waves into the snow, ice and snow density, vertical sea ice structure, location of the dynamic sea surface height, and snow depth distribution. The only direct sea ice thickness measurement is a drill hole. Therefore it is important to keep in mind that the data of the above-mentioned sources might have a bias and do have a finite uncertainty."

P1521 L09 Sea Ice CCI Algorithm Theoretical Basis Document? Link or citation?

We added the respective reference in the reference section.

P1521 L10 Do the authors take the slower wave propagation speed of the radar waves in the snow layer into account?

No. This is one of the sources of uncertainty, but a minor one compared to, for example, radar penetration or speckle. It should be noted that the mean effect of slower propagation speed is removed from the freeboard signal because bias between floe and lead elevations due to different re-trackers is removed from the freeboard signal.

P1521 L20 Is the grid optimized for RA/RA-2 data?

This grid results into a number of measurements we think to be at the limit of usability per one grid cell.

P1522 L03 ff Why is the area in Figure 1 limited to the Beaufort Sea and Canadian archipelago? It might be outside the region of available validation data, but interesting to see whether RA/RA-2 based freeboard shows basin-scale gradients.

We agree that it might have been interesting to see basin-scale gradients and we only show approximately half of the region possible. But the reviewer guessed correctly that we deliberately chose this region because of the availability of OIB and ULS data.

P1522 L09 ff It is very questionable that W99 is valid in this region. I think it is stated later in the text, but it would be good to mention it already here

Ok, we added: "In these regions the W99 snow depth is based on extrapolation."

P1523 L01 ff The RA grid cell size is latitude dependent, the AMSRE and OIB data is a constant radius of 100 km. Can this introduce a bias?

For comparison with OIB data, RA data is averaged over a constant radius around transect centre, so it is not latitude dependent. For some parts of the comparison we do use a latitude dependent grid because number of RA measurement per area is also latitude dependent. However we do not see how this could introduce a bias.

P1523 L10 There are more and more studies that raises the question how radar data has to be interpreted to yield ice freeboard (see Willatt 2011, Kurtz in TCD or Ricker in TCD).

That is true. We come back to this issue later on when we discuss the failure of the ASIRAS to locate the ice-snow interface.

P1523 L13 The statement that ASIRAS measures ice freeboard is contradicting to statements later in the text

We removed that statement from this line.

P1523 L20 How do the authors derive the ALS error of 10 to 15 cm. This range seems to be a rather high and only justified in regions which very few leads (which the Fram Strait data used here is typically not).

We agree. Unfortunately we do not reach the responsible person from DTU so that we can only make a statement based on own experience that possibly the accuracy on the scale we look at this data (i.e. 50 km) is an order of magnitude better, i.e., 0.01 m. We added: "As measurements are averaged along 50 km transects located in an area of frequent lead occurrence the accuracy relevant for this study is of the order of 0.01 m for the ALS data."

P1524 L18 Figure 2 caption: Change mooring to moorings

Corrected.

P1525 L07 ff An additional assumption must also be made that pulse-limited radar altimetry yields a radar freeboard that is not biased by surface roughness. And with the different backscatter signatures of open water, level and deformed ice, this is not very likely. Higher resolution data may be less affected or differently biased, like oblique laser scanner data over open water. Therefore one important objective should also be the investigation of potential biases of space-borne radar altimeter data.

We agree with the reviewer that this should be an objective, too. However, I guess we all agree now that the main focus of this paper is (unfortunately) not the evaluation of the RA gridded freeboard but that we needed to keep this for a later stage for some reason. What we could do and did was an investigation of the input parameter and assumptions. As we have changed the title accordingly we think that the readers' expectation is now in a different field.

P1526 L8 Does "standard" mean no distinction between MYI and FYI?

Yes. Given the fact that we have maybe written this a bit sloppy this part reads now:

3. RA and RA-2 sea ice freeboard is used to compute sea ice draft using different input data and compared to ULS sea ice draft data. This is done using our "standard set of densities" (see below). For BGEP mooring ULS data we compute in addition sea ice draft separately for MYI and FYI densities and two different fixed snow densities.
4. RA-2 sea ice freeboard is used to compute sea ice thickness combining the standard set of densities with various snow depth realizations and compared with OIB sea ice thickness.

The standard set of densities is: ρi = 900 kg m-3 (average density of MYI and FYI) and ρw = 1030 kg m-3 (Wadhams et al.,1992); snow density is taken from W99 and varies over space and time (see Fig. 1 e, f). In order to account for the effect of different densities for MYI and FYI (in 3, see above) we use sea ice densities published elsewhere (e.g., Timco and Frederking, 1996; Alexandrov et al., 2010): 882 kg m-3 and 917 kg m-3, respectively. The two fixed snow density values used in 3 (see above) are 240 kg m-3 and 340 kg m-3 and correspond to the mean wintertime minimum and maximum snow density (Warren et al., 1999).

P1526 L15 This statement is somewhat vague. What does control the dependence and which sensor is more/less affected by which factor? In the beginning of the next section it is stated that snow depth is crucial for all altimeter data

This has been noted by reviewer #2 as well. Since we think that at this point in the manuscript this information is not required we deleted the part starting with "Note that the sea ice …" and ending with "… according to W99."

P1528 L21 Correct, but one could even state it more clearly that W99 is invalid in the CA completely unconstrained by observations. But what are the prospect of getting ice thickness retrievals inside the CA with RA/RA-2 anyway (closed ice cover, land contamination)?

We agree. We kept the CA data in Figure 5 c) but do not include them in our analysis anymore. Corresponding parts in the manuscript have been removed as well as corresponding entries in Table 2.

P1529 L05 ff Add: Based on laboratory experiments (Beaven et al., 1995)

We added: "at the used frequency in Ku-Band according to laboratory experiments (Beaven et al., 1995)" and further, to account for the fact that this might not be state of the art anymore: "There is growing evidence that this assumption does not hold for more cases the previously thought (e.g. Ricker et al., 2014)."

P1529 L13 I still don't see how the uncertainty of the ALS can be that high. How many leads were in the data as tie points for the sea surface correction? Could the difference not only be part of the sampling bias?

See our comment to P1523 L20

P1529 L17 Is there a missing curve in Figure 6? I see the red OIB freeboard and the OIB and W99 snow depth but not the blue RA-2 data

No. We have changed the caption now to read: "Histograms of OIB (red lines) and RA-2 (blue bars) freeboard. RA-2 freeboard is derived using OIB snow depth (light blue bars) and W99 snow depth (dark blue bars). Both MYI and FYI are included. Note the different y-axis scaling."

P1530 L01 ff I miss an explanation how the data was collocated. I am sure there was not always a good temporal and spatial match between the individual orbits and the submarine data.

As the goal is to look at the consistency of the gridded 100 km freeboard product which enters the sea ice thickness retrieval we did not carry out the comparison on an orbit by orbit basis. We added information about the co-location of RA-2 data where these were missing in section 2.2 so far.

P1530 L22 The authors state that the mooring is mostly in multi-year sea ice but in Table 4 the average draft (_1.6m) is more typical for first-year sea ice. Is there any explanation why the ice was untypically thin for multi-year ice?

We agree. In fact the choice of a very large area for comparison with the BGEP data plus the limitation to cases where the AMSR-E snow depth data set indicates no MYI at all was misleading us here. The area comprises in reality a mixture of FYI and MYI, with MYI being the larger contributor, though, which explains why the draft is relatively small.

P1531 L06 ff What is the reasoning of using airborne altimetry datasets as reference for sea-ice thickness when the main objective of the study is to determine how to get sea-ice thickness out of altimetry datasets? Sea-ice thickness from OIB is a product of observations (freeboard, snow depth) but not an observation itself.

We agree. The only proper way to do such a comparison would be lots of in-situ drillings. None of the air-borne instruments flying around can measure directly the sea ice thickness. It is always a combination of different measurements which makes the estimation of an ice thickness value possible. In case of OIB data this is snow depth from snow radar and total freeboard from ALS; in case of CryoVex this is sea ice freeboard from ASIRAS and total freeboard from ALS, in case of EM-Bird this is total (ice+snow) thickness from the EM induction device and total freeboard from ALS. So in a way, it doesn't matter which data set we use here. We would have used CryoVex data if they would have revealed proper results. As we stated further up, the "real" inter-comparison of sea ice thickness with other sea ice thickness data is going to happen at a later stage of the project. Then we will include more recent OIB data, EM data, and ICESat into the analysis as well as the left out ULS data.
The main rationale of showing these figures is to see whether there is a difference in the agreement between OIB campaign sea ice thickness and sea ice thickness derived from RA data with DIFFERENT treatments of snow depth.
We added: "The rationale of this section is to show the impact of utilizing different snow depth data sets for freeboard-to-thickness conversion based on the gridded, monthly average RA-2 freeboard data."

P1532 L05 Is this the result of a completely wrong snow depth assumption or could it be that the number of data points for the comparison is insufficient to stick out of the noise level?

We have left out any CryoVEx airborne data from the sea ice thickness comparison in the revised manuscript.

P1533 L08 0.02 cm as a mean (?) difference sounds unbelievable good. What is the standard deviation?

The standard deviation is of the order of 0.1 m. I would not say that this is too surprising because a) we are talking about averages over 50 km along track (OIB) and 100 km diameter disc and b) we use the OIB snow depth – which seems to be quite reliable - to convert the sea ice freeboard to the total freeboard.

P1534 L03 ff Please use consistent naming for the realizations (Numbering not included in the legend or caption of Figure 8)

That is true. We do not show these realizations as figures. We only show the results of these in the Table. Figures 7 and 8 are based on A1. All thin lines given are variations of A1 where the indicated ice densities and snow densities are used.
We added at the end of that paragraph: "Of these realizations only A1 is shown Figures 7 and 8."

P1534 L13 "any snow depth" means a statistically chosen value?

This is a typo: It should read: "… multiyear ice without including any snow depth information."

P1534 L17 Figure 8 gives the impression to me that none of the realization is able to capture the trend of ULS sea-ice draft in any year. Often, the entire range of realization is necessary to explain the winter cycle. Also if A3 (one fixed ice density) and A1 + A4 (both ice type dependent densities) "agree equally well", does this not mean that the ice-type dependent ice density is overruled by the choice of snow?

See our comment to P1534 L03ff. We don't agree with the reviewer as the agreement between RA2 and ULS is quite good during winters 2005/06 and 2006/07. The difference in the seasonal draft range is about 0.3 m in both winters. Once again, Figure 8 does not show the different realizations mentioned above. The rationale here is to give a feeling how good or bad the standard retrieval (from Laxon et al., 2003) which does NOT include ice type dependent density variations matches with the ULS data and what is the range in the obtained draft when we use the typical lowest and highest sea ice and snow densities. The rationale is to show that the range in draft values obtained with these different density settings within each month is as large as 50% or more of the entire seasonal draft range observed by the ULS.

P1535 L27 Was not the RA-2 sampled on 2_ x 0.5_ grid and the validation data on a 100 km sphere?

We apologize for the misunderstanding. We did not specify correctly what we did. For Fig. 1 we indeed sampled the RA-2 data onto a 2 degree longitude by 0.5 degree latitude grid which is approximately 60 km grid resolution; this was purely for visualization. For comparison and the colocation with all the other data we used data from single orbits and averaged them over the respective areas – as is detailed in section 2.2.

P1536 L01 This is a bold assumption, given the mix of surface types in the large footprint. It is questionable the deformed ice and level contribute equally to the backscatter signal and this has to be proven.

We agree. We added: "This depends, however, on the degree by which different ice types and ice surface properties impact the radar backscatter and the waveform (Zygmuntowska et al., 2013, Ricker et al., 2014). More studies need to look into the different backscatter of sea ice of different type and roughness to quantify the impact of sea ice property variation on the radar altimeter signal and thus the retrieved sea ice freeboard."

Please note: Reviewer #2 pointed out the unrealistic negative sea ice thickness value from RA-2 data. Triggered by this comment we recognized (VERY LATE, we know) that also a varying number of OIB sea ice thickness values is negative. Because the RRDP does not contain any negative OIB sea ice thickness values we have to assume that something fundamentally went wrong when performing this analysis for the first version of the manuscript. Meanwhile we repeated the analysis and are coming up with a new version of Figure 9 and changed numbers in Tables 5 and 6. Note that we have omitted sea ice thickness values computed for the Fram Strait region for the CryoVEx campaigns. This is motivated by the fact that because we don't have an independent ice thickness estimate as in case of the OIB flight but need to compute sea ice thickness from either ALS or ASIRAS data with snow information from, e.g. the Warren Climatology.

P1536 L03 I do not understand. Does this study not use the OIB freeboard and snow depth data and can produce the thickness with consistent assumptions? Or is this the Round-Robin part of the exercise?

The rationale is not to change evaluation data according to our needs. The OIB data products are used in the form they are provided to the users and we did not manipulate them.

P1536 L013 ff Here I am lost. Why do the authors reverse-engineer the sea-ice densities with (obviously) different snow densities than the provider of the ice thickness data?

This part received comments from the other two reviewers as well and has been deleted.

P1536 L25 ff The reason is that Ku-Band radar data may be influenced by density contrast in the snow or volume scattering in general and that the final word of the "correct" interpretation of SAR altimetry waveforms is not yet spoken. The data of ASIRAS is a very good example why Ku-band radar data should be taken with a grain of salt. It must therefore be the first step to understand the bias and uncertainties of radar freeboard before the conversion into thickness.

We assume the reviewer wanted to mention this in the text. The paragraph reads now: "Our interpretation of the CryoVEx data remains inconclusive because the ASIRAS instrument, which is supposed to sense the ice-snow interface and thus provide an independent sea ice freeboard measurement, failed to do so. Instead it provided the total freeboard like the ALS sensor. By means of atmospheric re-analysis data we identify snow cover property changes as a possible reason for CryoVEx data from 2011 but not for 2008. This suggests that even under freezing conditions sensors like Envisat RA-2 or Cryosat-2 might not sense the sea ice surface. It is likely, that vertical snow density gradients and/or volume scattering in the snow in general influence the radar signal, resulting in a less distinct signal from the ice-snow interface or in similarly strong returns from the snow surface or interior as was shown for Antarctic sea ice by Willatt et al., (2010)."

P1538 L18 ff (ii) What are the uncertainty factors in the airborne campaign? How do these uncertainties relate to those in the satellite data?

We take this comment as the request for being more detailed. Given the request by all reviewers that we should be more concise and shorten the paper we refer to the publications we are referring to in the data section.

P1540 L11 ff I downright disagree: The validation of sea-ice thickness retrievals from altimetry needs independent and non-altimetry validation data. There might be consistency between different freeboard data sets but that does not mean the thickness of both datasets is correct.

We changed the last two bullets of the recommendations as follows: "3. For a sophisticated inter-comparison and validation of the final sea ice thickness product from satellite altimetry it is mandatory to use independent and preferably non-altimetric validation data. The amount of such contemporary sea ice draft, snow depth and sea ice thickness data is clearly sub-optimal and needs to be improved.

4. Potential improvement from utilizing new sets of input parameters, e.g. densities, cannot be quantified without consistent input parameters for freeboard-to-thickness conversion. We call for a consistent internationally agreed standard set of densities to be used for freeboard-to-thickness conversion to be applied to air- and spaceborne altimeter data."