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 an estimation of the sensitivity of the methods used to the input parameters with the goal 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. The 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 normally expect to be used, 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 kept the more recent Operation Ice Bridge (OIB) data."

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 not be bias free 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 fix snow densities (see below).
4. RA-2 sea ice freeboard is used to compute sea ice thickness with Eq. (1) using the standard set of densities (see below) and is also compared with OIB data.

The standard set of densities is: ρi = 900 kg m-3 (mean of the densities of first-year and multiyear ice) and ρw = 1030 kg m-3; snow density is taken from W99 and thus varies over space and time (see Fig. 1 e, f). In order to account for the effect of multiyear and first-year ice (in 3, see above) we use specific 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 fix 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 added: "The number of data points is substantially smaller in this region than in the Arctic Ocean region discussed in Table 5, e.g., there are only 11 data points for 2008, which could partly explain the negative correlations. Another aspect could be that the W99 snow depth is indeed more valid for the area overflown during the OIB campaign which is closer to the Arctic Ocean while the areas sampled during the CryoVEx campaigns are further south. As we discussed earlier, we are quite confident that the W99 snow depth is still valid in the areas considered here (see Fig. 4)."

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: "We note that the 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 and instead provided snow freeboard like the ALS sensor. Therefore CryoVEx ASIRAS data could not be used as an additional source of sea ice freeboard data and, in combination with the ALS instrument, of snow depth. Based on atmospheric re-analysis data internal or even surface melt are identified as a possible reason for the 2011 CryoVEx data but not for the 2008 CryoVEx data. This suggests that even under apparent 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 we will be short here and refer to the respective publications.

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. If for some reason airborne altimeter data need to be used it is essential to use consistent input parameters for the freeboard-to-thickness conversion. Otherwise a potential improvement in performance from utilizing a new set of input parameters cannot be quantified. In other words: 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."