Author Responses (ARs) to each reviewer comment are in Bold.

The paper "Estimating snow depth over Arctic sea ice from calibrated dual-frequency radar freeboards" by Lawrence et al. deals with estimating snow depth by combining satellite-based measurements of snow and ice freeboard. The method requires prior calibration with independent freeboard measurements. Here, CryoSat-2 and AltiKa satellite freeboard measurements are calibrated with airborne Operation IceBridge (OIB) measurements.

The latter raises one of my main concerns: The method, as presented here, relies on having reliable independent freeboard data, which at the moment is only provided by OIB data. However, there is disagreement within the science community on how to interpret the OIB radar measurements, i.e. different retrieval algorithms differ in the way air-snow and snow-ice interfaces are detected and localized. A recent paper by Kwok et al. (2017) showed that this caused OIB snow depths as retrieved from different groups to differ on average by up to 7 cm (for first-year ice) and 12 cm (for multi-year ice) for the 2013-2015 data (see Fig. 7 & 8 in Kwok et al., 2017), which is used in the paper presented here. The variability of snow depths is also quite different (so it is not just a constant bias between the different products). Though this problem is briefly mentioned in the paper presented here, this is only done rather late (on p. 14, l. 14), which does not represent how severe this issue is for the proposed method of retrieving snow depth. I think that this should be mentioned and discussed far earlier and with more emphasis because it has major implications on the usability and accuracy of the proposed retrieval method! Ideally, the authors would perform their comparison not only for OIB quicklook data, but also for (at least) one of the other OIB-based freeboard retrieval data sets to estimate how much this can influence the results.

Author Response (AR): You have raised an important point which was not addressed adequately in the first submission of the paper. In the latest draft, we discuss the results of the Kwok et al. (2017) inter-comparison paper early on in section 2.4 (page 7 lines 15-23) when first introducing Operation IceBridge (OIB), and acknowledge that the variability of OIB Snow Radar data from different groups presents a limitation to our methodology. Following the results of the Kwok et al. (2017) study we now employ instead NASA JPL snow depths in our methodology. This data set shows best agreement with ERA-interim and the Warren climatology for the years 2013-2015 (Fig. 9, Kwok et al. (2017)). Regardless of the discrepancy between different OIB Snow Radar datasets, our methodology offers a means to extrapolate the OIB snow depth dataset (whichever is chosen or deemed "best") to the wider Arctic, both spatially and temporally. Our DuST product would benefit from the development of a next-generation Snow Radar data set, combining the best qualities of each existing processing algorithm, as is advocated by Kwok et al. (2017). This is discussed in section 3.2 (page 13 line 14 to page 14 line 6). The use of an optimised OIB Snow Radar dataset could improve our snow depth estimates but would not alter the methodology, which we feel therefore deserves publication.

A further concern is that the study of Guerreiro et al. (2016) also uses CryoSat-2 and AltiKa freeboard measurements to retrieve snow depth. Instead of calibrating these Ku and Ka-band measurements with independent data (as done here), they theoretically analyze the penetration depths of both radar altimeters in snow and use snow density estimates to modify the Ku-band radar signal's velocity through the snow. In their study, they compare their retrieved snow depths with OIB snow depths for the same years as in the study presented here (2013-2015). They seem to have somewhat lower RMSDs (4.1...5.4 cm) as compared to the results presented here (4.9...6.7 cm), although their results are independent of OIB measurements, while the results here are not. Why are these results not compared here? Is there any advantage of using the

method presented here as compared to the one used in Guerreiro et al. (2016)? This comparison and discussion is missing here!

AR: We now include a discussion of the Guerreiro et al. (2016) approach and outline why our methodology is different and has its own advantages (page 7 lines 1-13). We have aimed in this paper to outline a methodology that can be applied in future when AltiKa is no longer operational, and demonstrate the methodology applied to Envisat and ICESat in order to show that it could also be applied to CS-2 and ICESat-2 when ICESat-2 is launched. The method of Guerreiro et al. (2016) relies on the ability to reprocess CS-2 data to produce pseudo LRM CS-2 data in order to achieve a footprint similar to AltiKa. By doing so, the authors can then assume that the remaining elevation difference found between AltiKa and pLRM CS-2 is the result of snow penetration difference alone and thus snow depth can be found as the difference between the two. This methodology could not be applied to, for example, CS-2 and ICESat-2 because neither dataset can be processed such as to make the effective footprints the same.

I found it confusing that the authors first declare that radar altimetry penetrates through to the snow-ice-interface, while laser altimetry does not (p. 2). AltiKa is presented as a radar altimeter (thus suggesting that it penetrates through the snow), but it is later compared with OIB's ATM laser freeboard (section 3.5). From what I understood, Guerreiro et al. (2016) conclude that the radar signal from AltiKa does not penetrate the snow, while Armitage and Ridout (2015) concluded that the AltiKa signal is scattered from roughly the midpoint of the snow layer. This discrepancy is not even mentioned here. What do your results suggest? Please comment/discuss/specify.

AR: Please re-refer to the introduction since it has been restructured in line with your subsequent criticism. Following an overview of the Guerreiro et al. (2016) and Armitage and Ridout (2015) studies, we have added a paragraph (page 4, lines 3 - 22) to clarify our position on AltiKa and CS2 snow penetration.

Another issue is that I think the structure of the paper could be improved:

a) In an "Introduction" I would mainly expect to read about the importance of the presented study, how it fits into the context of already existing studies and what is the new contribution of the presented study. Instead, we here get a general introduction on the importance of snow (ok) and we are presented the equations used to convert ice/ snow freeboard to snow depth (more appropriate for the "Data and Methods" section?). This is followed by a chapter that lists existing snow depth products, where I would prefer to read more about the differences to the presented study and the implications these have instead of a list of methods.

AR: On your advice, we have changed the structure of the paper. The introduction no longer contains any equations but is rather an overview of: the importance of snow, the current methods to retrieve it, their limitations, and our proposed methodology and a justification for its necessity.

b) The "Results" section contains a lot of what I would consider discussion (or speculation as some of the statements on p. 11 are not based on citations), while the "Discussion" section on p. 14, l. 20 starts with showing more results...

AR: We have now combined Results and discussion into one section.

Otherwise, the manuscript is, in general, well written and I was able to follow the method.

Specific comments:

p. 1, l. 3: "*...can be applied to any coincident freeboard measurements*" -> to any coincident snow and ice freeboard measurements? (would be clearer)

AR: The methodology can be applied to any coincident satellite freeboards they do not have to measure the snow and ice freeboards (indeed we suggest that AltiKa and CS-2 do not measure snow and ice freeboard but rather that we do not know where in the snowpack their signal originates).

p. 1, l. 19: "*...snow depth estimates could be usefully assimilated...*" -> "usefully" is a vague (and strange) expression here...

AR: This has been removed. See page 1 lines 18-20.

p. 1, l. 23-24: "*The implications ... is*" -> The implications ... are

AR: removed

p. 2, l. 4: Eq. (1) -> Is this formula from Beaven et al., 1995?

AR: This section of the introduction including this formula has been removed

p. 2, l. 27-28: " *The granular nature of snow acts to scatter and dissipate microwave energy radiating from the Earth's surface, reducing the surface brightness temperature.*" -> This statement is only true for part of the frequency spectrum of microwaves! Not true for low microwave frequencies.

AR: This statement has been removed. See page 2 line 19 onwards.

p. 2, l. 30: "*for a given frequency*" -> Too vague, I'd prefer to see the frequency (range) that you mean here.

AR: This section has been condensed. See page 2 line 19 onwards.

p. 3, l. 30-31: "*AltiKa was designated with a maximum penetration depth of 0, i.e. no penetration, and CS-2 a maximum penetration of 1, i.e. full snow penetration...*" -> What does this mean? Is it possible to retrieve snow depth using this method? Could you compare these with your method?

AR: This statement has been clarified (see page 3, lines 14-19). We now discuss their methodology and its limitations further on page 6 (lines 34 onwards).

p. 4, l. 8-14: You write about the issues raised by different satellite footprint sizes, please also give the CS-2 footprint size here to make the comparison easier.

AR: This has been included (page 3, line 30)

p. 4, l. 28: "*retrieves surface elevations up to 81.5*◦" -> a) Please add "latitude" (to avoid confusion with "geometrical elevations", which can also be given in degrees). b) I think this should be mentioned earlier in the manuscript because it constitutes a major limitation for polar applications of AltiKa.

AR: "latitude" has been included (page 4, line 27). This is now also mentioned earlier on page 3, (line 11).

p. 6: References for statements in l. 10-15 ?

AR: Added (page 6, line 16)

p. 6, l. 20: "*It*" -> it + "*this criteria*" -> this criterion

AR: Corrected (page 6, line 23)

p. 6, l. 21: Is "*snagging*" a word generally used for this? (just asking)

AR: Yes. Not sure when the term was coined but it appears as early as 1992 in "F. M. Fetterer, M. R. Drinkwater, K. C. Jezek, S. W. C. Laxon, R. G. Onstott, and L.M. H. Ulander, "Sea Ice Altimetry," in Microwave Remote Sensing of Sea Ice. Washington, DC, USA: AGU, 1992, ch. 7"

p. 6, l. 22: "*To overcome these problems,...* " -> refers to which problems? the off-nadir ranging of leads or also roughness?

AR: This has been changed to "To overcome the CS-2/AltiKa freeboard bias" (page 6, line 26)

p. 6, l.26: "*... we instead adopt an approach...*" -> Did you come up with this approach? Or did you take up an existing approach? (If yes, which one?/Reference?)

AR: We came up with this approach. We have changed this to "we here introduce an approach" accordingly (page 6 line 31)

p. 6, l. 30-31: "*the appeal of this methodology is its applicability to any freeboard data sets*"

Why would this (i.e. applying to any freeboard data sets) not be possible for the method described in Guerreiro et al. (2016), for example? Wouldn't both have to be re-evaluated for their performance with different freeboard measurements anyways?

AR: The method of Guerreiro et al. (2016) relies on the ability to reprocess CS-2 data to produce pseudo LRM CS-2 data in order to achieve a footprint similar to AltiKa. By doing so, the authors then make the assumption that the remaining elevation difference found between AltiKa and pLRM CS-2 is the result of snow penetration difference alone and thus snow depth can be found as the difference between the two. This methodology could not be applied to, for example, CS-2 and ICESat-2 because neither dataset can be processed such as to make the effective footprints the same. This is explained in section 2.3, page 7 line 5 onwards.

p. 6, l. 31: "*By calibrating satellite freeboards with an independent data set, biases are systematically corrected for*" -> I think this statement is too "optimistic"/general. Whether or not biases are systematically corrected for depends to a large extent on the quality, accuracy, and temporal + spatial resolution of the independent data. Not to mention that the bias is not the only thing to worry about...

AR: This sentence has been removed

p. 7, l. 7: "*snow depth, retrieved with the Kansas Snow Radar to within 5 cm accuracy* " -> Here (and also already in the introduction) it should be mentioned that different snow depth retrieval algorithms give very different snow depths! (Kwok et al., 2017)

AR: This is now discussed in this section (see page 7 lines 15-23)

p. 12, l. 5: Asterisk too high?

AR: Asterisk has been removed. (Page 14 line 9)

p. 14, l. 22: "*Spring*" -> spring

AR: corrected

p. 16, l. 3-5: Did you use 2016 OIB data for calibration when comparing with the 2013, 2014 and 2015 OIB data? If not, why not?

AR: yes we did. This has been clarified (page 17, line 11)

p. 16, l. 14-15: remove parentheses around "*Kwok et al., 2017*"

AR: corrected

p. 17, l. 1: "*onto a onto a*"

AR: corrected

p. 17, 9-10: "*Snow depth agrees with expected spatial distribution and magnitude"* -> Compared to what? How do you know? Or do you mean just with regard to the statement that follows (on thicker snow over multi-year and thinner snow over first-year ice)?

AR: With regard to the statement that follows. This has been clarified (page 18, line 28 onwards)

Fig. 18, l. 8: " *...this evaluation does demonstrate the ability to up-scale OIB snow depths to the wider-Arctic* " -> Do you mean ability as in "we do not get nonphysical snow depth values" or how is the ability for this demonstrated here without comparing the snow depths to independent data?

AR: we mean that the scatter plots suggest the ability to use the calibration functions to predict OIB snow depths for an unsampled year and region. This has been clarified (page 19, line 13)

Fig. 1 & 2: For the sake of completeness, it would be good to mention what the dashed grey line is (the zero line I guess).

AR: included

Fig. 3: Why are the snow depths smoothed? Is there a physical reason for this? The spatial variability contains information too (about real variability or about the "consistency" of the method, for example), why not show this?

AR: Snow depth maps are now shown unsmoothed.

Fig. 6: It is very hard to see the OIB measurements on top of the snow depth map. Maybe zooming into the campaign area would be useful? L. 3 of caption: "*each grid cells*" -> each grid cell"

AR: Maps are no longer included and scatter plots for all years have been combined into a single plot.

Fig 6 & 7: In none of the scatter plots there is snow depth values <0cm or >60cm, why would you show the data for a range of -20 to 80cm? This raises the question whether this was made to make the regression look "better"... and also creates unused white space that could be used for information instead.

AR: Plots now scaled from 0-60 cm