Answers to Referee 1

The paper is well put together and offers detailed assessment of the validity of Ice measurements at Baseline D. However, I would suggest to the authors to include more information on the specific changes/evolutions that have been implemented between the Baseline C and D processing chains. I have provided a set of minor comments and recommendations but have no significant concerns with the author's methods or results. My review is focused on the sea ice validation, since that is my area of expertise, although I have made a few minor comments elsewhere.

General comments:

1. It is important for tracking the history of each baseline to describe here what issues led to poor quality L2 data in baseline C (e.g. Section 3.3.2) and then what specific modifications were made to the retracking algorithms or processing chains that have led to vast improvements at baseline d.

Reply: We would like to thank the reviewer for the comment. The intention of the authors was to avoid making the text too technical, therefore reporting only the major evolutions applied to the new Ice Baseline-D. The complete list of improvements and evolutions implemented in both L1B and L2 Baseline-D processing chains is detailed in paragraphs 2.1 and 2.2, but we acknowledge the suggestion of the reviewer and we will add a summary table with the major differences between the two baselines.

Minor comments/edits: Line 40-41. Reword to explain why the 12 km is relevant.

Reply: The sentence will be rephrased as following: "Over sea ice, Baseline-D improves the quality of the retrieved heights inside and at the boundaries of the Synthetic Aperture Radar Interferometric (SARIn or SIN) acquisition mask, removing the negative freeboard pattern which is beneficial not only for freeboard retrieval, but for any application that exploits the phase information from SARIn Level 1B (L1B) products."

L48- 49. Are the exact same set of auxiliary measurements used for this ice draft analysis at baselines C and D?

Reply: The exact same set of auxiliary measurements is used to compare the Baseline C and D. In order to keep a smooth reading of the abstract we add the following sentence in section 3.3.2.

L403: The same set of auxiliary measurements is used to compare the Baseline C and D. (before the sentence: Relative to OIB....)

Fig 1. Please include product acronyms in the captions.

Reply: These will be added in the revised version of the manuscript.

Section 1. It would be useful here to include some introduction to the observations produced in the L2 data product. What specific measurements are provided by the ice processor at L2 for land ice, sea ice and lakes?

Reply: Thanks to referee 1 for this comment. The main outputs of the L2 Ice processing chain are the radar freeboard estimates, the difference in height between ice floes and adjacent waters well as ice sheet elevations, tracking changes in ice thickness. The text will be amended accordingly.

L 140-141. How can the SARIn mode be used to reduce uncertainty?

Reply: According to (Di Bella, 2018) the phase information available in the SARIn acquisition mode can be used to estimate the across-track location of leads, correct for the range overestimation and ultimately get a more precise value of the along-track SSH. The higher precision of the SSH enables, in turn, to reduce the uncertainty of the sea ice freeboard retrievals. The reference to (Di Bella, 2018) will be added at the end of the statement.

L 143. Need to explain what is meant by 'bad phase difference calibration'.

Reply: The phase difference calibration in Baseline-C did not consider CAL4 at the beginning of the SARIn acquisition. The statement will be rephrased and it will be added a reference to Section 3.3.3 where the issue is described together with the impact of its the in Baseline-D.

L150-153. What are these parameters for and how Cn they be used by the community?

Reply: The parameters are the stack peakiness and the position of the centre of the Gaussian that fits the range integrated power of the single look echoes within a stack as function of the look angle. Stack peakiness can be used to improve the sea ice discrimination. The position of the centre of the Gaussian that fits the range integrated power of the single look echoes within a stack as function of the look angle gives additional information on the shape of the Range Integrated Power, similar to the other stack characterisation parameters already present in the product.

L159. OK to refer to another study, but you need to at least include a definition here of this correction.

Reply: The statement will be rephrased to clarify that the mispointing angle accuracy was improved by considering a proper correction for the aberration of light when the data from Star Trackers are processed on-ground. In fact, the Star Trackers compute the satellite orientation in an inertial reference frame starting from comparison of the stars in their field of view with an on-board catalogue, therefore the aberration of light needs to be compensated for on ground to give accurate information about the satellite attitude.

L170. What is specific about the SARIn mode retracking? Specific in comparison to SAR mode?

Reply: The height value is still that from the SARIn mode specific retracking (phase has been used to relocate the height measurement across track), but new fields have been added to contain the sea ice processing height result (not relocated, and different retrackers for specular and diffuse waveforms), and freeboard and sea level anomalies are now computed in SARIn mode (previously SAR mode only).

L172-173. Define retracking before this discussion. You also need to include details of this retracker and how it is implemented.

Reply: the sentence will be changed to:

"In addition, a new threshold-of-first-maximum retracker is used…"

And after that sentence, the following text will be added:

"Retracking is the process whereby the initial range estimate in the L1B data is corrected for the deviation in the first echo return within the waveform from the reference position."

L176. 'Records' is quite ambiguous. Returns?

Reply: thanks for the comment, replaced with "waveform".

L214-215. This was an issue with baseline c data, or just an issue with the selected TDS for baseline d?

Reply: this was an issue with the Baseline-C processing chain in general and not specific to the particular TDS used in this study. The issue has now been fixed in Baseline-D.

L238-239. Clarify whether the angular correction is implemented by the data provider for baseline d L1B products? Can you explain in a little detail here the source of the angular error and its spatiotemporal dependence?

Reply: In Baseline-D a new Star Tracker Processor was developed to create files containing the most appropriate Star Tracker data. In addition, new fields were added to the L1B products to include the antenna bench angles (roll, pitch and yaw) and the sign conventions of these fields were updated.

L247. What are these retrackers? What are their differences? It would be extremely useful generally for the altimetry ice community if the authors could provide a table here with details of all the retrackers implemented for each surface type and sensing mode.

Reply: The details about the implemented Baseline-D retrackers are given in the Product Handbook document available at:

https://earth.esa.int/documents/10174/125272/CryoSat-Baseline-D-Product-Handbook

This reference will be added in the revised version of the manuscript.

L249. Citations?

Reply: the following citations will be added to the revised version of the manuscript.

Simonsen, S. and Sørensen, L.: Implications of changing scattering properties on Greenland ice sheet volume change from Cryosat-2 altimetry, Remote Sens. Environ., 190, 207–216, https://doi.org/10.1016/j.rse.2016.12.012, 2017.

Schröder, L., Horwath, M., Dietrich, R., Helm, V., van den Broeke, M. R., and Ligtenberg, S. R. M.: Four decades of Antarctic surface elevation changes from multi-mission satellite altimetry, The Cryosphere, 13, 427–449, https://doi.org/10.5194/tc-13-427-2019, 2019.

L250. Up- dated surface mask derived from what? By whom? Fig 3. Include an inset map of the location.

Reply: The Level 2 products contain a flag word, provided at 1 Hz resolution, to classify the surface type at nadir. This classification is derived using a four-state surface identification grid, computed from a static Digital Terrain Model 2000 (DTM2000) file provided by an auxiliary file to the processing chain. We will add an inset map in figure 3.

L327-328. Explain why.

Reply: The sentence:
"… projects. Based on recommendations from the ESA project, CryoVal-LI, the 2016 CryoVEx airborne campaign (Skourup et al. 2018) revised the traditional satellite under-flights to fly parallel lines with spacing of 1 or 2 km next to the CryoSat nadir ground tracks."

Will be revised to

Traditional airborne validation campaigns for satellite radar altimetry have targeted satellite under-flights as close to the satellite nadir as possible. This approach is favourable when surveying a flat surface, however, a sloping surface will induce an off-nadir pointing of the radar returns, and the number of coinciding observations will be limited. The ESA project CryoVal-LI quantified this off-nadir pointing based on CryoSat SARIn L2 data and based on the project recommendations, the 2016 CryoVEx airborne campaign (Skourup et al. 2018) revised the traditional satellite under-flights to fly parallel lines with a spacing of 1 or 2 km next to the CryoSat nadir ground tracks. Figure 7 shows the Austfonna flight path, which is optimised to ensure as many coinciding observations between CryoSat and airborne surveys, within the possible range of the aircraft.

L349-350. Add explanation on the latest ESA baseline d retracking algorithm and processing chain. Does it follow one of the other group's processing chains? Are the retracking solutions from other group's algorithms available in the baseline d L2 ice processor data product?

Reply: The details about the implemented Baseline-D retrackers are given in the Product Handbook document available at:

https://earth.esa.int/documents/10174/125272/CryoSat-Baseline-D-Product-Handbook

This reference will be added in the revised version of the manuscript.

In Baseline-D Ice L2 products only the retrackers described in the above document have been used.

L375-376. Clarify.

Reply: It is unclear what the reviewer refers to, since Lines 375-376 are in the middle of two sentences. If the Reviewer refers to "The Lead areas....minimum" (Lines 376-377), then the Clarification is already provided in the subsequent sentence (377-380).

If the reviewer refers to "...but the lead returns also influence the measurements nearby", this can be reformulated as "...is easily identifiable; the measurements close to the peak are characterised by a decay SP, which is still higher than the value found in the absence of a lead, since the latter can be the dominant return in the waveform up to about 1.5 km away from the sub-satellite point (Armitage et al., 2014) "

L385.The hyperlink doesn't seem to work.

Reply: The sentence will be rephrased in the revised version of the manuscript to include updated reference in the following way:

"Previous analyses carried out by the CryoSea-Nice ESA project ~~(https://projects.alongtrack.com/csn/)~~ highlighted important over-estimations in the freeboard values of the ESA CryoSat Baseline-C products relative to in-situ data (see the recommendation Rec.9 in [CSEM Report 2017])
Following these conclusions, modifications have been made to develop the new ESA CryoSat Baseline-D freeboard product. We present here the first assessments of this updated version."

[CSEM Raport 2017] Summary and Recommendations Report of the CryoSat-2 Expert Meeting, CSEM, 2017, ESRIN, https://earth.esa.int/documents/10174/1822995/CryoSat-CSEM-Summary-and-Recommendations-Report.pdf

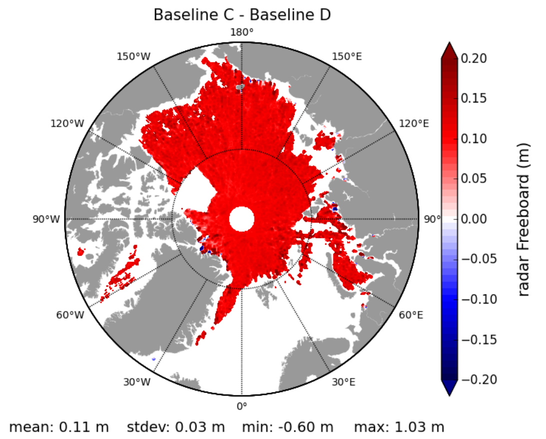
L394. Is this correct? I expect this rms measure is a convolution of the noise with valid signal at the sub grid-cell level. A better estimate for the noise distribution would be obtained from along-track rms of height observations over smooth level ice. Fig 8. Very difficult to see the difference map. Can you enlarge the points and ensure the color scale is cantered so that white = zero. Almost impossible to visualize the positive anomalies here.

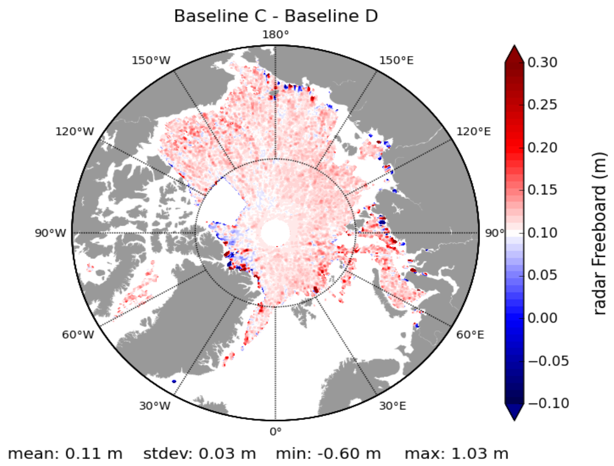
Reply: We do agree with the comment that "real" RMS should be calculated along track. This is actually the procedure we use to estimate freeboard uncertainties in the products. Here we wanted to insist on the fact that the Baseline-D improvement is more a bias correction than a decreasing of noise in the product. We agree that this sentence is confusing, not necessary and not entirely true. Then we have reformulated the sentence as it follows:

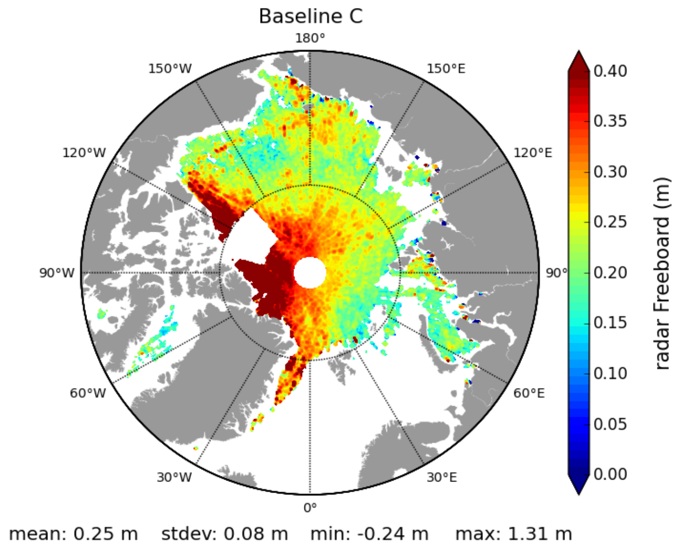
L393: In addition, the Root Mean Square (RMS) in each 20 x 20 km2 pixel, referring for a small scale freeboard variability, is similar for the 2 Baselines (about 15 cm).

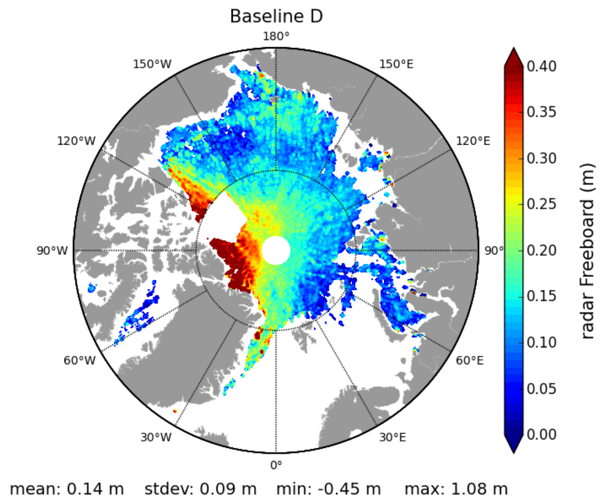
For a better visibility, all figures have been replotted (maps for figure 8 are given at the end of the document).

However, a colour scale centred on zero does not provide much information (see the figure below).









L404. You need to explain in detail the processing changes that have led to such extreme improvements here. Fig 9a and b. Please include the best-fit line so the reader can see the deviation from 1:1. How were the OIB freeboard observations processed? Are they an official NSIDC product? How are the CS2 observations converted to draft from freeboard? Most importantly what assumptions were made about the snow load?

Reply: To process the OIB freeboard observations, we use the ATM laser and the snow radar NSIDC official data. The ice freeboard is calculated from the difference between the laser total freeboard and the snow depth of the OIB snow radar (equation (1)) and the radar freeboard is then calculated taking into account the decrease of velocity of the radar wave into the snow following equation (2):

$$FB_{ice} = FB_{laser} - snowdepth \; (1)$$

$$FB_{radar} = FB_{ice} - snowdepth \times (1 + 0{,}51 \times \rho_s)^{(-1.5)} (2)$$

with$\rho_s = 0{,}3$

The CryoSat sea ice draft is calculated from the difference between the gridded SIT product and the gridded ice freeboard product. This last one corresponds to the radar freeboard corrected for the decrease in radar velocity in the snow pack with the same formulae presented above. For this calculation the snow depth of the Warren99 climatology is used.

In order to add this information in the manuscript we have modified the paragraph from L400 as it follows:

L400: Figure 9 presents scatter comparisons with the Beaufort Gyre Exploration project (BGEP, https;//www.whoi.edu/beaufortgyre) and NSIDC Operation Ice Bridge official product (OIB, https://daacdata.apps.nsidc.org/pub/DATASETS/ICEBRIDGE/Evaluation_Products/IceBridge_Sea_Ice_Freeboard_SnowDepth_and_Thickness_QuickLook) in situ measurements. To compute OIB sea ice freeboard, we calculate the difference between the ATM mean total freeboard and the snow depth estimated from the snow radar. The freeboard radar is then deduced taking into account the decrease in radar velocity in the snow pack as follows:

$$FB_{radar} = FB_{ice} - snowdepth \times (1 + 0{,}51 \times \rho_s)^{(-1.5)} (2)$$

with $\rho_s = 0{,}3$

To compare with BGEP data, we compute a CryoSat ice draft from the difference between the gridded sea ice yhickness (that integrates the snow load) and ice freeboard data. Note that the ice freeboard is calculated from the radar freeboard taking into account the decrease in radar velocity in the snow pack using the formula specified in Eq 2.

L422-42. This paragraph seems more appropriate for the introduction.

Reply: this paragraph will be reformulated and eventually moved to introduction.

L438-439. By what degree can it be reduced? I would also expect it to reduce systematic uncertainty associated with biases in the SSH retrieval.

Reply: The sentence at lines 437-439 will be changed as: ...accounting for the range overestimation to off-nadir leads (Armitage et al., 2014). Thus, the ONC can correct for biases in the SSH retrieval due to off-nadir ranging, estimated to be 1-4 cm by Armitage et al., 2014. Additionally, the more precise SSH obtained from SARIn measurements can reduce by ~29% the average random uncertainty of freeboard estimates (Di Bella et al., 2018).

L471-472. Most of these citations do not correspond with the AWI data product.

Reply: This is intentional. The citations provide references to the range of sea ice thickness data sets that rely on CryoSat L1 data.

L479-480. So is the ESA retracking algorithm used to derive freeboard, then the remaining processing uses the AWI chain? Or is the full processing from AWI but using different data baselines?

Reply: The processing from AWI includes retracking. We have clarified the sentence:

Processing steps consist of a L2 processor based on L1 waveforms for the estimation of sea ice freeboard and thickness at full along-track resolution and a L2 processor for mapping data on a space-time grid for a monthly period with a resolution of 25 km in the northern hemisphere.

L489. Is the phase used to produce an ONC in this processing chain?

Reply: No, the phase information is not used for the SIN mode.

L504-508. This passage requires explaining in more detail.

Reply: We have extended the passage:

This impact analysis however does not provide any insights into the specific algorithm changes that are causing the observed ΔSIT. We therefore speculate that the change in power scaling for SIN data between IPF1C and IPF1D is the reason for the larger impact on SIN data as the AWI surface type classification depends partly on total waveform backscatter. <u>Specifically, we observed that fewer IPF1D waveforms are classified as lead or sea ice (not shown) with a classification algorithm previously used for IPF1C. Therefore, the gridded thicknesses in both baselines in SIN mode areas are based on a different subset of input waveforms, which is far less the case in SAR mode areas.</u>

L536. Identical? Fig 12. It is very difficult to observe any differences between these classifications if indeed there are any. If there are, can you use extra panels to highlight the differences?

Reply: We entirely re-write section 3.3.5 adopting stack peakiness in lead classification, so the classification results are different.

L562-563. Explain.

Reply: We will add some additional text to explain, in the revised version of the manuscript.

"Here we assume that the observations follow a mixture of a Gaussian (70%) and Cauchy (30%) distributions. The mixture of distribution is more robust and ensures that the estimated standard deviations are not too influenced by erroneous observations (Nielsen et al, 2015)."

L564-565. Why is one meter considered to be good? Do you mean the lake mean height from a single track?

Reply: The one meter threshold was chosen as a reference for comparing the two baselines. The point is to quantify the difference in valid observations between the two baselines. As suggested we could also choose a threshold of 0.5 meters as the reference. The results of this threshold are illustrated in the figure below. We will add the following sentence: "The one meter threshold is arbitrary and simply selected to establish a common reference".

L575. Why was such a large offset present at baseline c?! Fig 13. Is this a stacked bar chart? If not, move the BD bars next to each BC bar.

Reply: following this suggestion we have modified the graph with Baseline-D bars next to the Baseline-C ones.

L590. The lower noise level is not really confirmed here, as I explained in the comment above this would require a different approach to ascertain.

Reply: this has been further explained in the comment above.

L599. Which statistic? Mean bias, rmse..?

Reply: The statistics are improving in terms of mean and standard deviation when comparing results obtained with the new Ice Baseline-D. This has been specified in the new version of the manuscript.