Response to Review #2 (Thomas Armitage)

Date: 20 March 2019

We thank the Editor, reviewers and those who provided short comments on the manuscript for their inputs. The feedback has helped to improve both the clarity and content of the manuscript. We have provided responses to both the short comments and full reviews. We indicated by section, specified by paragraph, where revisions were made within the manuscript text. Since we include figures in response to both the short comments and the full reviews, there is a letter code to indicate if the figure is related to a short comment (e.g. SC1.1) or to a reviewer comment (e.g. RC3.1). The figures in the manuscript itself maintain normal numbering convention (e.g. Figure 1, Figure 2, etc.).

The following is a list of the major changes to the manuscript:

- We have revised the manuscript text for clarity and brevity. In particular we have shortened the Introduction and rearranged the text regarding the treatment of snow depth in each satellite data product.
- Based on input received during the review, we have revised the text of Section 2.1 and the information provided in Table 1 to clarify specific aspects of the processing chain and waveform retrackers used in each satellite thickness product.
- We have extended Table 1 so as to include the details of two additional satellite-derived sea ice thickness data sets, although these data are not included in the further analysis. This decision is a compromise between providing the pertinent details of publicly-available data products, while not overwhelming the reader with too much information in the figures and tables.
- We have replaced the CS2SMOS data set used in the original submission with an updated version of the data set, and revised all figures and tables containing CS2SMOS data.
- We have updated figures and tables wherever possible with new data that has become available since the original submission. In particular, we now include the BGEP ULS ice draft observations for the 2016-2017 season.
- We expanded our results to include winter growth rates, adding a new table (Table 5).
- The reviewers highlighted concerns regarding the original methods used to calculate the correlation between data products, and that using a near-neighbour interpolation with a search radius of 50 km could potentially artificially improve the correlation results. To address these concerns, we have revised the approach to calculate the correlation statistics between the satellite data products, as well as between the satellite and airborne observations. In the revised manuscript the thickness observations are placed onto a common grid (0.4° latitude by 4° longitude longitude) before common grid cells are compared and correlation statistics calculated. This follows the approach originally taken in Laxon et al. (2013) as well as in subsequent studies and allows the reader to place our results in the context of the published literature. We note that this did not change the results of the correlation analyses in a substantial way.

Within the manuscript text, all edits (additions/deletions) are indicated in red font. The manuscript version indicating track changes is posted as a separate author’s comment in order to keep the response document concise.

Review of "Assessment of Contemporary Satellite Sea Ice Thickness Products of Arctic Sea ice" by Sallila et al.

The authors present an intercomparison of several different sea ice thickness data products, principally derived from CryoSat-2, and compare these datasets against BGEP upward-looking sonar draft data and Operation IceBridge airborne thickness data. Unsurprisingly, they find good agreement between the three CS2 products which use empirical retrackers (CPOM/AWI/JPL), while the GSFC data appears to overestimate the mean thickness early in the season and shows quite different spatial variability, and the CS2SMOS data potentially underestimates the thickness throughout the season. The analysis is straightforward, presented clearly enough and, while there are no new science results, I believe this paper could be useful to the community. I'm happy to recommend publication once my points below have been addressed.

We thank the Reviewer for this recommendation for publication.
Here we review a range of publicly available, satellite-derived sea ice thickness data products, and show, for the first time, a side-by-side comparison of the thickness estimates. We disagree that there are "no new science results": we discuss the spatial and temporal variations in sea ice thickness and growth rates across the Arctic Ocean between 2010 and 2018, thereby providing scientific analyses of 8 years of CryoSat-2 observations, which is a longer time-series of satellite-derived thickness observations than has been previously published. In addition, the evaluation of the satellite products using in situ data (airborne and ULS) spans a time period that exceeds any previously-published result (e.g. Laxon et al., 2013; Kwok and Cunningham, 2015, Wang et al., 2016, or Tilling et al., 2018).

I have a few suggestions that I think will improve the paper:

RC 2.1. Introduction: I found the introduction cumbersome. I would combine the first two paragraphs into one, and reduce the length by ~half as some of the material you discuss here is not really relevant for the current manuscript. Most of the fifth paragraph (Page 3, lines 12-27) is unnecessary in the intro, particularly as you essentially repeat it all in section 2, so remove – the first sentence can go at the end of the previous paragraph, and combine this with the final intro paragraph.

We have modified the introduction (Section 1) according to this suggestion.

RC 2.2. Why don't you use the EM bird data? I believe it is a fairly extensive dataset, is commonly used by groups to assess CS2 thickness data, and is the only technique that directly measures thickness rather than a proxy (draft/freeboard).

We agree that the AEM data are very useful for assessment. However, one of our requirements for both the satellite sea ice thickness data products, as well as the validation data sets, is that they are publicly available to the user community. Unfortunately, we have not found a publicly available version, and the source information for the AEM data used in the previous studies you mention was not provided. It would have been possible to acquire the relevant data from one of the data providers examined here (AWI), as they kindly offered it, but such data do not currently fit the requirements of open access and may have raised concerns about the independence of the study, since it comes from one of the data providers. While additional evaluation against EM bird data might potentially provide further insights, we conclude that its absence does not impact the results shown in our study, nor is the presented evaluation of the satellite data compromised in any way.

RC 2.3. You do not discuss the impact of the different treatments of snow on your product intercomparison. There is a paragraph on the general difficulty of snow in sea ice altimetry in the conclusions (Page 19, lines 14-34) – I don't think this belongs here, I would move it to the discussion – but I would also like to see some discussion on how the different snow treatments impact your results. Importantly, you say that all the CS2 data use the 'modified' Warren climatology (Table 1), however I believe that the CPOM processing uses the basin-mean value of the Warren snow depth over the central Arctic region for MYI and half this value for FYI (at least this is my understanding from Tilling et al. (2018)), whereas I believe AWI/JPL use the spatially varying 'modified' Warren snow depth. I don't know if you realized this but it is not clear in the manuscript? [It might be worth checking with Tilling et al.] This could explain some of the differences in Figure 2 (AWI/JPL overestimate thick ice, underestimate thin ice relative to CPOM) as well as why the AWI and JPL maps look quite similar (they treat snow the same, but different to CPOM), and some of the differences in the in-situ comparison.

Thank you for bringing the Tilling et al. (2018) publication to our attention. While it was not available at the time of writing, it is now published, and we are happy to add this reference to the manuscript. Based on personal communication with the data provider we confirmed that the snow in the CPOM product is developed according to your description above.
We added details about snow treatment in the product descriptions in Section 2.1 and corrected Table 1 accordingly. We now describe how snow is treated per product in Section 2. We have also shortened and modified the paragraph about snow (page 19 lines 14-34 of the original paper) and moved it to the discussion section.

RC 2.4. The APP-x data is very poor, and I don't really see what value it brings to the paper, other than to say that it is very poor (you could be more unequivocal about this). It shows almost no interannual or spatial variability, and you find that the mean thickness grows by 1m between Feb and Apr (page 13 line 10) which is physically unbelievable. After showing the APP-x data in Figures 2 and 3, maybe Figure 4, I would consider dropping it from the rest of the manuscript and say in your conclusions that it is unrealistic and shouldn't be used over a CS2 product. I was very surprised that you suggest (page 19, line 8) users should use the APP-x data simply because it is available daily in NRT and Arctic-wide! Judging from your analysis, given that APP-x does not capture interannual or spatial variability, users would be better off simply using the monthly CS2 climatology because at least it captures the spatial variability and does not grow unrealistically thick ice at the end of winter.

Our goal is not to make decisions or assumptions about how potential end-users will use these data products. Rather our goal is to provide an independent assessment of the currently available satellite sea ice thickness data products so as to allow end users to make reliable and informed decisions, based on their "use case". The APP-x product is an operational data product, meaning that it is derived twice daily, provided in a routine data format, and is available year-round in a continuous data stream with consistent latency from a national data centre. Moreover, the APP-x ice thickness data is part of a climate data record maintained by the National Oceanic and Atmospheric Administration (NOAA). Because this is the only operational ice thickness data product currently available it is of wide interest to the community who rely on such observational inputs delivered in an operational setting. We have provided a balanced summary of both the advantages and disadvantages of the APP-x product in the second paragraph of the conclusions. However, we also acknowledge that none of the data products currently available meet the general needs of the end user, and, in the final paragraph of the manuscript text, we provide a set of recommendations for improvements.

Minor comments:

It's not clear what you mean by "contemporary" in the title – occurring at the same time, or occurring in the present? Either way I don't think it's necessary; consider revising: "An assessment of satellite-derived Arctic sea ice thickness data"

By contemporary we mean data that belong to the present. The word "contemporary" is a widely used adjective in scientific publishing and we use it here to indicate "current" data sets. This is to alert the reader to the fact that we do not consider historical satellite thickness datasets that are otherwise good or interesting, such as from ICESat or Envisat.

Page 1, line 19-20: Should read "Among the data compared, the blended..."

Corrected.

Page 2, line 26: Re-write: "The most widely-used thickness datasets are derived from the radar altimeter..."

Sentence revised.

Page 2, line 29: "..retrack*ing*..."

Corrected.

Page 2, line 32: "*...basin-scale* gradients..."

Added "basin-scale".

Page 3, line 4: "Given the variety of sea ice thickness data..."

Corrected.

Page 4, line 8: You should reference Tilling et al. (2018) as well as Laxon, as the data you are using was produced by Tilling et al and is somewhat different from the original Laxon data.

The publication by Tilling et al. (2018) was not available at the time of writing. However, as it is now published, we are happy to add this reference to the manuscript in Section 2.1.1 and, where applicable, we have updated the Tilling et al., 2016 references to Tilling et al., 2018.

Page 5, line 9, Table 1: The CPOM processing uses separate retrackers for leads (Gaussian+Exponential model fit) and floes (threshold), and they apply a correction to account for this.

We added details about the separate approaches taken for ice floe and lead retracking in the CPOM product description in Section 2.1.1 and we modified the description in Table 1.

Page 5, lines 24-29: I don't think it is true to say that the AWI processing "does not differ significantly" from the CPOM processing. They are 'similar', but a number of important differences I can think of from the top of my head: 1) different retracking approaches (I believe AWI apply a 40% or 50% threshold retracker to all waveforms), 2) different sea level interpolation, 3) different waveform discrimination criteria (e.g., right and left sided peakiness), the MSS which you mention, different treatment of snow (see my comments above).

We have modified the lines in question, in paragraph 2 in Section 2.1.2. We added more detail about the AWI processing chain in Section 2.1.2, and we modified Table 1 for the CPOM product following the description in Tilling et al. (2018), page 1211, Section 4.2.2.

Page 6, line 9: I think Kwok and Cunningham (2015) is missing from the bibliography? Please check all references in the text actually appear in the bibliography.

Thank you for pointing this out. We added Kwok and Cunningham (2015) to the references and we checked that all of the other references were complete.

Section 3.1: Calculating a pixel-by-pixel correlation after oversampling the data to 5km doesn't make sense, as it may artificially increase the correlation stats. I would suggest sampling all the data onto an identical 25 km grid for the analysis. Also, why use a 50km search radius for the interpolation when the data are posted on 25 km grids? Surely this will act to smooth the data, also improving the correlation?

We thank the Reviewer for their comment which is consistent with RC1.1 and RC3.2. The gridding approach has been modified. We have clarified this in the methods, Section 3.1. We have also revised all text where the results of the gridded data are discussed. Please see further details in our response to RC1.1.

Page 9, line 13: You should provide some justification for using the CPOM data as your baseline, it seems rather arbitrary.

We choose the CPOM data product as the baseline against which to compare the other products partly because it was the first CryoSat-2 sea ice thickness data made publicly available and is thus the most widely used product. Also, according to the number of citations, the method is the best known with roughly four times the citations for the associated product publications compared to the next most cited product, which is the AWI product. We added an explanation of our choice in Section 3.1.

Page 11, lines 5-9: The phrasing here jarred a bit, as it suggests that the "cooler summer =

thicker ice in 2013/14" result is a "finding" of yours which was "also noted" by Tilling et al. (2015). Obviously, you haven't shown any linkages between temperature and thickness, and I know you didn't mean to imply this, so please rephrase this section to say that "Tilling et al found that... The thickness changes in Figure 2 are consistent with their result."

We modified the text in the first paragraph of Section 4.1 as suggested.

Page 11, line 25: The AWI CS2 data appears to show extensive thin ice (<0.5m) in fall but this is the noise floor of CS2, so not necessarily believable – is there something in the AWI algorithm that causes this? I would like some more discussion/explanation of this.

In response to Reviewer 3, we modified the range and increments of the colour scale used in Figure 2, to more clearly show the thickness gradients. We have also modified the text in the first paragraph of Section 4.2 to more appropriately describe the data shown in Figure 2.

Page 14, lines 1-10, Table 3: You should calculate the anomalies relative to the same baseline period to make the numbers comparable i.e., 2011-2015 considering this is the common baseline.

Thank you for this suggestion. We have modified the analysis so that now the baseline period is 2011-2015, and all anomalies are calculated relative to the mean over that period. We updated the results in Table 3 and the associated interpretations in the text, which occur mainly in Section 4.3, paragraph 4. Also, the term "climatological mean" has been changed to "baseline mean".

Page 14, line 16: "CS2-only products" – judging from figure 7, all of the satellite products miss the thickest/thinnest ice including CS2SMOS?

We agree, and we have modified the text in Section 4.4, paragraph 2.

Page 14, lines 20-21: by what measure does the AWI product most closely align with the ULS draft? Likewise, by what measure does the GSFC data least agree? I would say the CPOM/AWI/JPL show the same agreement from figure 7?

Our statements are based on the agreement between modal draft and modal thickness (assuming a ratio of 0.9), the width of the distributions (full width half maximum/standard deviation), the shape of the thickness distributions, and how well the distributions capture both the distribution of thin ice <1.5 m, and the decay in the distribution for thicker ice >1.5 m. The results shown in Figure 7 have been updated through the addition of ULS draft and satellite thickness data for the 2016-2017 season and the use of the revised CS2SMOS data set. Consequently, we have revised our interpretation of the draft to thickness comparisons and results. We have modified the text in Section 4.4, paragraph 2 to more adequately describe the comparisons.

Section 4.5, paragraph 1, Figure 10: I would think the negative freeboards in the GSFC data must have something to do with the different retracking, as this is the major difference between the data products? I would speculate that the GSFC floe retracking might be sensitive to off-nadir scattering, as they use a functional fit to the entire waveform and power in the trailing edge could skew the fit to later delay times.

Thank you for this very interesting insight. We agree that the abundance of negative freeboards in the GSFC product is anomalous compared with the other data products and may be a result of the empirical retracker. We also agree that it would be insightful for the community to be able to understand the sensitivity of the CryoSat-2 freeboard estimates to the specific retrackers. This analysis is unfortunately outside the scope of our current study which is targeted towards end users of the existing satellite thickness data products, rather than an assessment of retracker methodology and sensitivities