Response to Review #1 (Jack Landy)

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) 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.

Assessment of Contemporary Satellite Sea Ice Thickness Products for Arctic Sea Ice
Sallila et al., 2018

This study provides the first genuine inter-comparison of pan-Arctic sea ice thickness data products obtained from three different satellite sensors. It is systematic in approach, equitably comparing the datasets to identify seasonal and interannual patterns of ice thickness variability, as well as assessing biases versus independent reference data. Results from the assessment will prove to be a valuable contribution to the field of Arctic climate science and potentially to operational stakeholders.

We thank the Reviewer for their encouraging comments on the usefulness of our study for the science community and stakeholders.

However, I have some concerns about the robustness of methodology used to process and compare datasets. These concerns focus on the procedure used to resample data, as described in the comments below. I have also made a few suggestions to improve the analysis and impact of the author's findings. I'd recommend this manuscript is published in The Cryosphere, following these revisions.

Please do get in contact if you have questions regarding these comments. Kind regards, Jack Landy

General comments:

RC 1.1. The method used to re-grid the satellite products may be introducing errors into the data inter-comparison and artificially improving the correlations. By gridding to 5 km with a search radius of 50 km you are introducing significant spatial autocorrelation between adjacent grid cells (through your method cells can theoretically only be treated as independent samples at a length scale 25 km). It is not statistically robust to be calculating the pixel-wise correlation between spatially-dependent samples, and this may be artificially improving the coefficient and significance. Perhaps you could try recalculating the correlation coefficient for a down-sampled 25-km version of your current grid? Take a look at this paper https://www.cambridge.org/core/journals/annals-of-glaciology/article/impact-of-spatial-aliasing-on-seaice-thickness-measurements/AB5DB924171B219D25E64A829E418840 to consider your gridding approach. (This is also worth considering for your OIB comparisons in Fig 9).

Agreed. Originally data were placed on a grid with 5 km resolution using a nearest neighbour interpolation function with a search radius of 50 km. As you and other reviewers have commented, this could lead to an autocorrelation between adjacent grid cells, and an overestimate of the correlation between independent data products. As suggested, we have revised our approach and now place the data sets onto a common 0.4° latitude by 4° longitude grid before calculating the correlations between products. We have revised the description of the gridding procedure in Section 3.1, updated figures 4, 9 and 10, and Tables 2 and 3. We have updated Sections 3, 4, and 5 to reflect the changes in the Figures and Tables. Our revised method follows the approach originally taken in Laxon et al. (2013), and then used in subsequent studies (e.g. Tilling et al., 2018), thereby allowing the reader to place our new results in the context of the previously published statistics.

RC 1.2. Linked to the above comment, you suggest at Page 14, Line 19 that the satellite products do not capture the thinnest or thickest ice because of the grid-averaging process. It surely makes sense then, for a fair comparison between satellite obs and the ULS, to average the ULS measurements over some window, rather than binning individual obs in your histograms. I understand this is difficult because it means averaging over time and comparing it to grid averages over space. However, could you identify a time-averaging window based on recorded ice drift speeds (either a seasonally- representative drift speed or obtained from satellite ice motion data)?

We believe the Reviewer may have misunderstood a detail in the original text, regarding the satellite data. The monthly data sets as provided are actually gridded at 25 km resolution (without any additional gridding by us). Regarding the assessment of the satellite data products through comparison with the ULS analysis, we follow the methodology of Laxon et al. (2013) and Kwok and Cunningham (2015). This is to enable the reader to place our results in the context of the existing literature. We have revised the text in Section 4.1 for clarity.

RC 1.3. Your correlations between all the CS2 products are understandably very high (Page 12, Line 12) and there's not too much new analysis here on the differences between the CS2-only datasets. I would be very interested to know how closely each product picks up small-scale spatial variability in ice thickness. For instance, you could calculate the anomaly maps of each year compared to the climatological average, then assess the pixel-wise inter-product correlations between positive and negative anomalies. Do some of the CS2 products identify more or less of the smaller-scale thickness features than others? Are they realistic? This analysis would improve e.g. the first paragraph of the discussion.

Our results show for the first time the correlation between a variety of CS2-only products, which differ due to waveform retracking algorithms. We have also significantly extended the evaluation of the CS2 products, through comparisons with independent data over much longer time period (2010-2017) than has previously been published. We are unsure exactly what the Reviewer means by "small-scale spatial variability in ice thickness", since the majority of the satellite data products are provided on grids of resolution 25 km. Sea ice thickness varies on much smaller scales than this, but is not possible to assess it in this analysis of gridded data. However the results in Table 2 provide details of the spatial variability between products at the regional scale. APP-x shows the least variability from region to region, with mean ice thickness in spring across the product record falling within a range of 0.19 m for the southern regions of the central Arctic basin (regions 3-6). JPL shows the largest regional-scale variations, with a range of 0.97 m in average spring thickness across regions 3-6. Table 3 also provides some interesting results on differences between CS2-only products, in terms of seasonal variations relative to the 2011-2015 baseline. There are also differences in the growth rates (Table 5) and wintertime trends (Fig. 5) across the CS2-only products, where the GSFC product has the lowest daily growth rate. Based on these results, we have included additional discussion in Sections 4.2 and 5 about the regional differences.

RC 1.4. There is not a lot of analysis or discussion of regional variations in the products: whether any obvious differences are regionally-dependent etc., except from Table 2.

Can you add some deeper discussion on regional differences in the products, and offer some interpretation as to why these differences may exist? This would be a really valuable addition to the work.

Based on the results shown in Table 2 we have added new discussion to the second paragraph of Section 4.2, per this suggestion.

Minor comments/edits:

Page 2. Lines 15-16. By what mechanism does Laxon 13 speculate that lower ice volume could be a factor in recent minima in ice extent?

The direct statement is found in paragraph 27 of Laxon et al. (2013), which is as follows: "Finally, we can speculate that the lower ice thickness and volume in February/March 2012, as compared with February/March 2011, may have been one factor behind the record minimum ice extent reached in September 2012". The statement points out that the anomalously low ice volume at the end of winter in April 2012 may have contributed to the sea ice minimum observed in September 2012. We have revised the text (Section1, second paragraph) to clarify that we are referring to the September 2012 minimum.

P2. L29-31. Can you briefly introduce here some of the main differences in approach?

Section 2.1 is dedicated to the description of the satellite data products. We therefore elected to introduce information concerning the retrackers there, and we do not feel that this information is critical for inclusion in the introduction section.

P3. L17-18. This is confusing: you're not including the ICESat data in your evaluation right, so why do you discuss evaluating ICESat here? Do you mean they have been used as a tool for evaluating satellite products in the past?

We have revised the introduction section based on comment RC2.1 and subsequently removed this sentence about ICESat.

P4. L6. What is the time period under investigation?

The period under investigation is October 2010 - April 2018. For some analysis, e.g. the time-series of ice thickness shown in Fig. 5, data are shown based on the availability for each data product. The baseline period for comparisons is the period with full product overlap (spring 2011-fall 2015). We added a brief clarification in Section 2.1, paragraph 1.

P4. L8. What do you mean by mean scattering horizon?

We have revised Section 2.1.1, paragraph 2, based on updated information provided in Tilling et al., 2018 and the comments of Reviewer 2.

P4. L30+. I understand that estimates for snow depth are pretty similar between the CS-only datasets, but you do not refer to them here and they are critical. Can you add some info on how snow depth and density are estimated for each product?

For each product description in Section 2.1 we have added text about snow depths and densities, wherever that information was available.

P6. L16-18. The description of the waveform-fitting method used for the GSFC product needs to be improved. The physical model of Kurtz 14 does not calculate expected returns empirically. Their model is semi-analytical, producing a simulation of the expected echo shape based on statistical parameterizations of the sea ice backscattering properties and height distribution of surface roughness.

We have changed the explanation to better describe the method used to derive the GSFC thickness product, in Section 2.1.4, paragraph 2.

P7. L26-27. What is the physical relationship between the surface EB and ice thickness?

The OTIM model, used in the APP-x product, derives sea ice thickness as a function of heat fluxes, surface albedo and radiation, which all contribute to the surface energy budget. Furthermore, most of the flux and radiation parameters in the equations are functions of surface skin and air temperatures, surface pressure, surface air relative humidity, ice temperature, wind speed, cloud amount and snow depth, which are input parameters in the model. We refer the Reviewer to Wang et al. (2010) where the equations and a thorough explanation are provided. We revised the description of the APP-x data product, in Section 2.1.6, paragraph 2.

P7. L30-31. Why does sea ice roughness affect the surface EB?

We reference Wang et al. (2010), where this was stated in the conclusions. We added a reference to Wang et al. (2010) to the line in question, Section 2.1.6, paragraph 3.

P8. L 5. How is snow depth estimated in the model?

Based on personal communication with the data provider, the initial snow depth estimates are based on those from Warren et al. (1999), but these estimates are then adjusted using field observations and the final snow depth values have been chosen experimentally. We have added information about the treatment of snow in the APP-x sea ice thickness data product to Section 2.1.6.

P9. L2. What measurement of uncertainty is that? Bias? RMSE?

This measurement comes from the error propagation typically used in sea ice thickness remote sensing and follows Giles et al. (2007). The thickness uncertainty is calculated using an assumption of the probable errors associated with the variables used in the thickness equation. We added an explanatory subordinate clause in Section 2.2.2 and added the reference to Giles et al. (2007).

Giles, K. A., Laxon, S. W., Wingham, D. J., Wallis, D. W., Krabill, W. B., Leuschen, C. J., McAdoo, D., Manizade, S. S., Raney, R. K.: Combined airborne laser and radar altimeter

measurements over the Fram Strait in May 2002. Remote Sens. Environ., 111(2–3), 182–194, doi:10.1016/j.rse.2007.02.037, 2007.

P9. L20. It is unconventional to subtract the focus product from the reference – typically it's the other way around. This makes the plots e.g. Fig 3 harder to interpret, because you have products with higher thickness than the CPOM data given as negative numbers.

We have changed the order of product differences in Equations 1 and 2, leading to changes in Figure 3 and the corresponding interpretations.

P10. L3. Why do you pick a 200 km radius? Is this an arbitrary threshold or based on estimated ice drift displacements over the season? Do you use the same threshold in winter and summer? Have you tried varying the threshold +/- 100 km to see if you get the same results?

We used a radius of 200 km, since it is consistent with the original evaluation presented in both Laxon et al. (2013) and Kwok and Cunningham (2015). This allows the reader to place our results in the context of the conclusions of prior studies.

P10. L9. So, one correlation coeff for the whole time period, using each monthly average as a single sample?

We have added clarification on paragraph 2, Section 3.2.

P10. L17. Above you refer to the issue of oversampling at the start or end of a seasonal period and then treating products as the same when comparing. The IceBridge spring data are collected in discrete campaigns, so the period of observations will often not be spread evenly over the March-April period. This seems like the same issue – so how might this be introducing uncertainty into your satellite-OIB product comparisons?

We have revised our methodology to calculate the correlations between the satellite data products and the IceBridge data. We now place the IceBridge and satellite-derived thickness data on a 0.4° latitude by 4° longitude grid to conduct the comparisons. The approach helps to mitigate the uneven spatial and temporal sampling of the sea ice along the IceBridge flight-lines compared to the monthly means obtained from the satellite data products. Also, as we have adopted the same approach originally used in Laxon et al. (2013) and subsequent studies (e.g. Tilling et al., 2018), this allows the reader to place our extended results in the context of the existing literature.

P11. L4. Can you give the mean and +/- (between the techniques) of the fall 2011/2012 ice thickness, to provide the reader with a bit more context? Here and in other similar places within the paper?

Added values for the mean thickness and its variation on the suggested line as well as in other places within Section 4.1.

P11. L18-20. Is there a way you can quantify this? e.g. by calculating the spatial covariance of the gridded data between years, or by simply calculating the average pixel standard deviation in ice thickness across the period? i.e. how sensitive are each of the products to spatial features of ice thickness like the thick ice tail in the Beaufort Gyre, at what scales?

This was meant as a brief discussion of results in Figure 2, based on visual inspection. Here we did not emphasize exact quantitative differences, as there are plentiful analyses included in the later figures and tables. We agree though that studying the sensitivity of the products, or even one product, to spatial features, would be a great study on its own!

P12. L3-4. For which season?

Good catch, these are for the spring season. We added the season in the second paragraph of Section 4.2.

P13. L 14. And indeed of the thicker sea ice too. . . It's very strange that the technique based on a thermodynamic model and surface EB suggests highest ice growth rates at the end of winter, when you'd expect such a method to be better at capturing the expected reduction in thermodynamic growth.

Unfortunately we have not discovered a reason behind this behaviour in the available references describing this data product. It seems like the air temperature, and other factors that likely support ice growth, outweigh some of the other physical phenomena in the EB model.

P13. L15-34. It is very interesting that the CS2 only products are more similar in spring than in fall, when intuitively you'd expect them to diverge more when the ice is thicker. Can you provide any insight on why this might be? In what ways may the processing differences induce variations in fall that are not apparent in spring?

After applying revisions to the method used to calculate seasonal averages, and obtaining a new version of the CS2SMOS data, we find that the CS2-only products tend to agree quite well in both the spring and the fall. The exception is the GSFC data product which diverges from the other CS2 products in the fall. Figure 3a shows that while ice thickness in the fall is on average thicker in the GSFC product, it is most apparent in regions 5 and 6 (East Siberian and Laptev Seas) and over the thickest multiyear ice in the central Arctic. Recall that the major difference between the GSFC product and the other CS2-only products is that the GSFC retracker uses a waveform fitting method (Table 1).

P14. L 34. What do you mean by 'varying dependency to draft'?

We were attempting to describe the varying correlation values for the three different buoys (Fig. 8). The line in question in Section 4.4 has been deleted from the revised manuscript.

P15. L2-3. Since the thickness is relatively low in this region (<1.5 m) this would imply a higher contribution of SMOS data to the optimal interpolation, right? So does this give you any information about the reliability of SMOS data?

The major difference between the CS2SMOS and AWI products is the inclusion of SMOS data, and this suggests that there are occasions when SMOS estimates higher thickness, which could be due to the relatively thin ice in this region, as you state. However, we do not make statements about the reliability of SMOS data and refer the Reviewer to Ricker et al., 2017 regarding the CS2SMOS interpolation scheme.

P16. L1-15. It is also worth noting that GSFC freeboard for both fall and spring is generally lower than the AWI freeboard, whereas the GSFC ice thickness was determined as the highest in Figs 2-3 etc. So the positive bias (whether an error or not) in GSFC ice thickness compared to CPOM and AWI is not coming from the freeboard measurements, and must be from the conversion to ice thickness (snow depth, snow or ice density differences, or perhaps filtering/processing chain differences).

We agree that, on average, GSFC freeboard observations are 0.14 meters thinner than corresponding observations from AWI in spring, and 0.08 meters thinner in fall (Fig. 10), despite GSFC indicating higher sea ice thickness for both seasons compared to the AWI product, as shown in Figures 2 and 3. We have included this result in section 4.5, but can only speculate about the cause.

P17. L7. ‘freeboards of less than approximately 0.05 m’.

Added the word ‘approximately’ to the line in question, in Section 5, paragraph 3.

P17. L 23. You suggest the assumption is that SMOS data is too heavily weighted in the combined product, but where do you get this assumption from? Are you making this assumption, or have you got it from previous papers? (if so which ones?). Potentially the uneven thickness increase detected by SMOS is more realistic and the CS2 products cannot detect new thin ice..?

We based this statement on existing studies where it is stated that SMOS assumes 100% ice concentration in its thickness retrieval algorithm, which may cause underestimation of ice thickness in conditions with less than 100% ice concentration, as mentioned in Tian-Kunze et al. 2014 and Ricker et al. 2017. It is certainly possible that SMOS estimates are more realistic for some conditions, most notably thin ice in areas of high ice concentration. We have revised the text of Section 5 (discussion) to reflect the reprocessed CS2SMOS data set used in the revised manuscript.

P18. L 5-8. Re. variations in ice concentration thresholds, you could have defined your own conservative threshold (e.g. 75

The focus of this study is in the assessment of publicly available data products as they are provided to the community. Furthermore, a variety of ice concentration products are used in the product processing chains. Choosing one of those, and applying it routinely across all products, would likely favour some products and hinder others. Table 1 provides further details of the ice concentration data and thresholds applied to each product.

P19. Snow section. This is quite out of place here, as you do not pay much attention to variations in snow treatment between products, and the focus of the paper is on inter-comparison rather than the systematic issue of estimating snow (common between all products). I suggest removing this passage and if desired, reference the challenge in a single sentence or so.

We have added text in Section 2 about the treatment of snow in each data product, specifying the differences. Although all products use snow climatology estimates derived from Warren et. al. (1999), they employ different ways to implement the climatology. Based on this, the paragraph about snow in Section 6 has been modified and as suggested by Reviewer 2, it is now shorter and most of the text is moved to Section 5, the discussion.

Table 1. This is great! Really useful compilation.

Thank you.

Fig 5. Can you add the linear per winter ice growth rate for each product and year to the plot? i.e. X cm/month.

We expanded our results to include growth rates in a new table, Table 5.

Fig 7. Can you emphasize more clearly on the plot that the bold distribution is ice draft, as a quick glance at the figure gives one the instant conclusion that CS2SMOS data must be best.

Added bold text to the legends in Figure 7 so as to highlight "ULS Draft".

Fig 8. Can you add the x = 0.9y line to the plot as a reference?

The 0.9 factor is an estimate of the ratio between thickness and draft based on Rothrock et al. (2008), and therefore we do not want to emphasise it as the true ratio. Whereas for Figure 7, where the distribution shapes are compared, plotting also ULS draft is crucial.

We have modified the text in the third paragraph of Section 4.4 to emphasize the fact that 0.9 is an estimate of the draft to thickness ratio.