The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2018-197-RC1, 2018

Interactive comment on "Assessment of Contemporary Satellite Sea Ice Thickness Products for Arctic Sea Ice" *by* Heidi Sallila et al.

J.C. Landy (Referee)

jack.landy@bristol.ac.uk

Received and published: 9 October 2018

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.

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:

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).

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)?

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.

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.

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?

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

P3. L17-18. This is confusing: you're not including the IECSat 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?

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

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

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?

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.

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

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

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

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

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.

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?

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

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?

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?

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?

P12. L3-4. For which season?

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.

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?

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

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?

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).

P17. L7. 'freeboards of less than approximately 0.05 m'.

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..?

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

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 intercomparison 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.

Table 1. This is great! Really useful compilation.

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.

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.

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

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2018-197, 2018.