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.

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

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.

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).

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.

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.

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"

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

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

Page 2, line 29: "..retrack*ing*…"

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

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

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.

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.

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).

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.

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 25km grid for the analysis. Also, why use a 50km search radius for the interpolation when the data are posted on 25km grids? Surely this will act to smooth the data, also improving the correlation?

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

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."

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.

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.

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

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?

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.