EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-214-RC1, 2022

Comment on egusphere-2022-214

Jack Landy (Referee)

Referee comment on "Arctic sea ice radar freeboard retrieval from ERS-2 using altimetry: Toward sea ice thickness observation from 1995 to 2021" by Marion Bocquet et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-214-RC1, 2022

The authors construct a 25-year record of Arctic sea ice radar freeboard by reconciling the measurements from three radar altimetry missions, one ongoing and two historic. Their primary motivation is to generate the first step towards a long-term sea ice thickness record for the Arctic Ocean. This would be the first observational sea ice thickness record spanning such a long period and would offer valuable comparison to existing proxy sea ice thickness (SIT) records based on ice age and models. In my view, a robust 25+ year time series of Arctic sea ice thickness would represent a major scientific breakthrough with implications for understanding global climate changes in the modern era and validating and improving sea ice models, among other potential applications.

Generally, I find the approach and methods to be scientifically sound. I have some minor comments but nothing that questions the rigour of the generated time series. The validation against existing SIT data from satellites, airborne and in situ sensors is comprehensive and convincing.

Excellent work on a really valuable study – it was a pleasure to read! Feel free to get in touch if you have any questions, Jack Landy

Minor comments/edits:

Line 2. Sea ice volume**'s**..?

L14-15. I would suggest including other statistics of the variability on the bias within the abstract. Given the ML algorithm aims to remove the bias I would argue the stats on variability are more interesting for the reader.

L28. Technically past radar altimeters have not allowed basin scale, so altimetry doesn't offer a 'global approach' over the long term. But this is nit-picky.

L51. Explain 'heuristic retracker TFMRA50'.

L65. Check Appendix Table 1. Does this tally?

L103-104. How is it aggregated? Bit vague.

L134-135. Requires citations.

Figure 1. Could you add here a map of the satellite coverage and the locations of different validation datasets? This would be useful for the reader to understand limits of the record and interpret differences to specific validation data.

L166-167. Which version of the IS-1 data was used?

Figure 2. I would suggest to add histograms to one side for each of the three elevation profiles, so it is easier to visualize any differences/biases.

L215. Sure, but how much are they improved quantitatively if we are using Envisat as the reference?

L226-227. Can you add a table of the thresholds after they have been calculated to keep lead/floe proportions the same during overlap periods? This would aid repeatability of the study.

Also is this based on SIC from an external dataset? Otherwise which mission do you use as a reference to calibrate the other to?

Since snagging to off-nadir leads is more likely with the LRM mode missions, is there a chance these missions will include more leads accidentally classified as floes, using this approach for an equal proportion of floe/lead?

L236-237. More information is required on the interpolation method and procedure.

L237-238. Do you discard rFBs above a max distance to the nearest lead? If so what limit do you use?

L250. Can you explain a little more about this constant SLA bias in LRM? Why does it appear and what could be done, in theory, to remove it?

Figure 4. Add the sensing mode to the plot. The CS2 data here is SAR mode right, not calculated from pLRM?

L278. Explain these terms.

L283-284. What does this mean? Retrained again or just some sort of tuning? Might it be very different from the training with 90-10 split?

L301-303. Needs more info. Why do you calculate uncertainty differently between leads and floes? The uncertainty at floes is governed by the variability in height measurements at proximal leads.

What distance is used to calculate an along-track mean elevation? Is the variability in individual floe height obs around this not just a measure of the topography? It will be higher over MYI but does this realistically mean the uncertainty is higher?

L307. How are the uncertainties reduced during gridding? Speckle noise should drop as a function of N observations, but SLA uncertainty should only drop as a function of N tracks (because SSH error is highly correlated along track).

L312. Systematic uncertainty due to roughness is 20-30% of the freeboard as well as of the thickness.

L316-318. Based on the schematic in figure 6 everything you've done seems fine, but it is still confusing to follow all the steps. What are these 'other inputs'? And which variables do you divide by the sqrt of the number of observations vs the sqrt of the number of tracks when gridding?

L325-329. How do you estimate the gaussian noise distribution statistics? is this the sigma = 2*sigma_omega in Figure 6? The output from a monte carlo error budget depends closely on the assumptions taken for the error distributions so this is important.

L340. For which months in Fig 8?

L357. What are these numbers as a % of the mean rFB?

L359. Again what are these as a % of the mean rFB?

L363. Can you do the same for the ERS2-Envisat comparison?

Figure 7b. It looks like you may have some spurious tracks in Hudson Bay, Baffin Bay and Bering Strait that could contaminate the comparisons?

Figure 7 caption. Emphasize the distributions include CS2 data only for the coinciding region south of 81.5N.

L384. ‘static data’?

L392. I think it is reasonable to discount IMBs because they represent only the single floe they are deployed on (usually a thicker floe) and not their surrounding 12x12 km grid cell area. The authors could remove these comparisons so they don’t draw reader’s attention and they come to the wrong conclusions about the satellite data validity; but that is up to the authors.

L397. Could be attributed to, but not definitely.

Figure 9 and elsewhere. Define the acronyms of statistical tests in the caption.

L406-407. Can this say anything about the calibration? Are the BGEP ice conditions more representative of average sea ice conditions in the Arctic and the other ULS datasets more of thin ice conditions?

Was the calibration not slightly overestimating thin ice thickness for Envisat?

L415-416. How do these numbers compare to your estimated uncertainties for the same regions?

L420-422. What are the statistics like for CS2 data processed with this method? You don’t necessarily need to show a plot, but some idea of biases would be useful. Do you also see generally negative biases for CS2? Especially over FYI?

L425-426. Could you try them also with the adapted warren climatology and see if biases get any smaller? Would help to clarify the impact of snow loading.

L427. Is Section 2.3 correct?

L441. It is worth making it a bit clearer on Fig 13 and throughout this section that these volumes miss out everything >81.5N.

Figure 13. I think readers would find it interesting to see more of your rFB dataset. I'd suggest an additional figure showing trends in rFB as a map for the overlap region, highlighting where the trends are significant or not.

Table A1. SAR you mean? or is this actually the CS2 LRM mode? I think SAR was used here right so state SAR parameters?