Manuscript ID: egusphere-2022-214

Title: Arctic sea ice radar freeboard retrieval from ERS-2 using altimetry : Toward sea ice thickness observation from 1995 to 2021

Marion Bocquet, Sara Fleury, Fanny Piras, Eero Rinne, Heidi Sallila, Florent Garnier, and Frédérique Rémy

Jack Landy (referee n°1) - global comment

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

Answer to Jack Landy (referee n°1) - global comment

We would like to thank the reviewer for his careful reading of the manuscript, for this positive feedback and for the relevant and constructive remarks that have helped to improve the quality of the manuscript. In order to fit with your comments, we have made a revision of the manuscript that should have corrected the textual issues and well improved the readability of the document. We hope that these modifications will meet your requirements. Please find below the details on how your specific comments have been taken into account. *In this document, the referee's comments are in bold type, the answers are in italic type, and the corrections to the revised manuscript are in normal type.*

Answers to Jack Landy (referee n°1) : specific comments

Line 2. Sea ice volume's..?

This correction has been done.

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.

These statistics have been added to the abstract. The following modification has been done :

L 14-15: Comparisons of corrected radar freeboards during overlap periods reveal good consistencies between missions, with a mean bias of 3 mm for Envisat/CryoSat-2 and 2mm for ERS-2/Envisat.
replaced by
Comparisons of corrected radar freeboards during overlap periods reveal good consistencies between missions, with a mean bias of 3 mm and a standard deviation of 9.7 cm for Envisat/CryoSat-2, and respectively 2 mm and a 3.8 cm for ERS-2/Envisat.

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.

That's true, we have modified a little the sentence.

L28. A global approach is possible through satellite altimetry, especially with radar altimetry, which is not impacted by the cloud cover and whose missions are continuous since 1991.
replaced by
A quasi-global approach is possible through satellite altimetry, especially with radar altimetry, which is not impacted by the cloud cover and whose missions are continuous since 1991.

L51. Explain 'heuristic retracker TFMRA50'.

Indeed, the sentence was not clear. It has been corrected. The purpose of this sentence was to explain that before any calibration (LRM/SAR) all the waveforms for both missions were processed with the same algorithm and the retracker that has been used is a TFMRA retracker (which is categorized as an empirical retracker based on a heuristic approach).

L51 : The consistency between missions is preserved by using the same processing chain regardless of the mission, with the heuristic retracker TFMRA50 (Helm et al., 2014)
replaced by
The consistency between missions is preserved by using the same processing chain regardless of the mission (before calibration), starting by the retracking algorithm : the empirical threshold first-maximum retracker algorithm (Helm et al., 2014) with a threshold of 50 % (TFMRA50).

L65 : Check Appendix Table 1. Does this tally?

The appendix indicates the RA characteristics. The table is supposed to help the reader to understand the estimation of uncertainties from the speckle noise for Envisat and ERS-2. Wingham et al. (2006) estimates the uncertainty due to speckle noise for CryoSat-2 in SAR, SARIn Mode as well as for LRM. CryoSat-2 LRM mode data are not used (neither pLRM) in this paper, but the CS-2 LRM speckle noise error on range is used to compute ERS-2 and Envisat uncertainties that come from the speckle

noise (L297-302). The legend of table A1 (now A2) has been developed, and a reference has been added in the corresponding section 3.5.

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

The following modification has been done to try to improve the clarity of the manuscript:

L103-104 : This information comes from the NSIDC 0061 sea ice age product (Tschudi et al., 2019) that is aggregated into two classes (MYI and FYI)

replaced by :

The study also requires a sea ice type product, this information is derived from the NSIDC 0061 sea ice age product (Tschudi et al., 2019) that is aggregated into two classes (MYI and FYI) according to the age of the ice (FYI : ice age between 0 and 1 year, MYI : ice age of at least one-year) at a weekly frequency. Data are respectively available as daily and weekly map with a 12,5 km grid resolution. The fraction of MYI is derived from the ice type information during the gridding processing step.

L134-135. Requires citations.

The citations were a few lines after, but for clarity we have added one in the first line.

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.

Figure 1. has been completed with a sub figure to represent the locations of different validation data sets with satellite coverage limitations.

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

ICESat-1 version used is the one from NASA Goddard not from NASA JPL. This precision has been added in the description paragraph.

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

As suggested a sub figure with probability density function, especially density probability function has been added in the following figure 2:

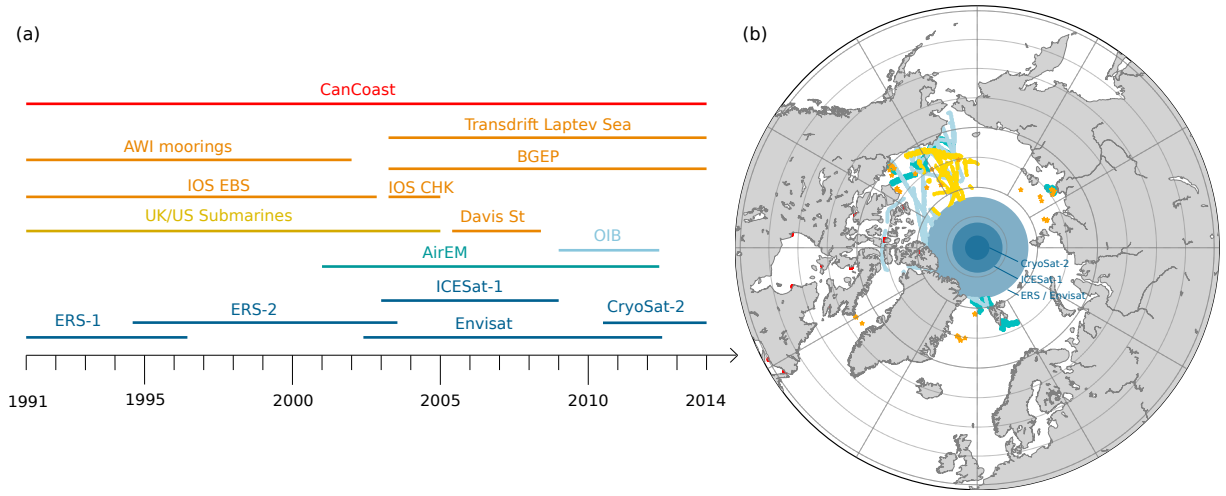


Figure 1. Summary of various available dataset for Envisat and ERS validation. Colors distinguish the different types of data. Dark blue for satellite products, light blue for airborne data, yellow for submarines, orange for anchored moorings, green for buoys and red for direct measurements. (a) Temporal availability (b) Spatial availability and extent of missions data gaps. Blue rounds represent altimeters coverage limitation due to their orbit inclination (81.5°N for Envisat, 86°N for ICESat-1 and 88°N for CryoSat-2)

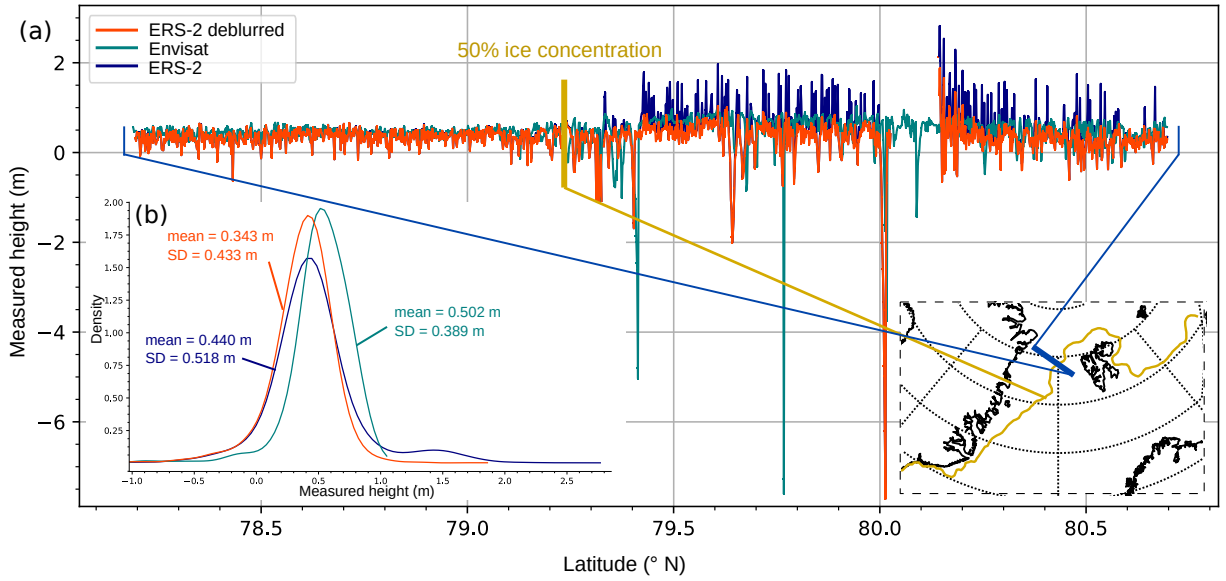


Figure 2. Profiles of surface height anomaly over sea ice and ocean for pass 25 between 78° N and 81° N for Envisat in blue-green (cycle 12), ERS-2 in blue and ERS-2 deblurred in orange (cycle 80). The red line represents the limit of 50 % concentration of sea ice, so as the limit between open ocean and an ice-covered area. The dark blue line shows the location of the pass between Svalbard Island and Greenland. (a) The surface height along the latitude and (b) the probability density function of surface height for the three passes with the associated statistics, the average and the standard deviation (SD). The color legend is identical for both sub-figures.

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

Pulse Blurring has to be seen as an asymmetric noise. As the FB or SLA processing are mainly form by statistical operations, succession of smoothing or interpolations, this noise will be reduced, some outliers due to the blurring will be removed. The resulting impact of blurring will be a positive non-constant bias on the SLA or on the FB, nevertheless the comparison with Envisat is not as easy because both missions (even if the Radar altimeter instrument is identical) are biased, so ERS ASA averaged over the basin with or without blurring compared to Envisat won't be so much relevant. However, we can see the impact of the deblurring on the noise of the data. We thus suggest to compare the standard deviation of ASA within each grid cell of a 12.5km resolution grid between Envisat (cycle 12) and ERS (cycle 80) before and after deblurring as an appendice (A1), see Fig.3.

The deblurred surface anomalies of ERS-2 now appear similar to Envisat.

replaced by :

The deblurred surface height anomalies of ERS-2 now appear more similar to Envisat in terms of noise and amplitude of variation. For this particular track, the standard deviation has been reduced by 16 % and get closer to Envisat's one. Figure A.1, shows more results on the impact of deblurring on ASA noise reduction comparing to Envisat during a whole cycle.

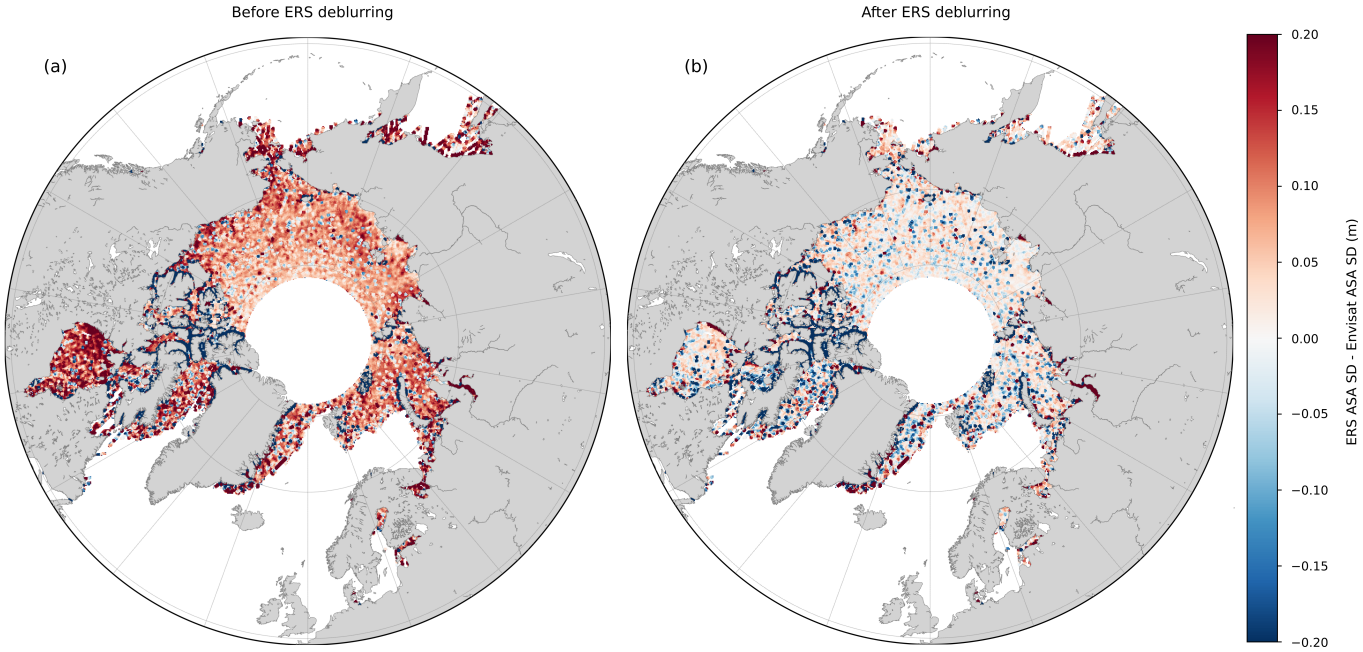


Figure 3. Comparison of the standard deviation of ASA between Envisat (cycle 12) and ERS (cycle 80) before (a) and after deblurring (b) within each grid cell of a 12.5km resolution grid. Median of the standard deviation difference for (a) is 0.075m and 0.008 m

---

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 the repeatability of the study.

We recompute the threshold only for ERS-2, considering Envisat as a reference. Thank you for this suggestion, the clarification has been made in section.The following table : Tab.1 has been added in appendices (A1) in the manuscript.

Table 1. Pulse peakiness thresholds for lead/floe classification

Mission (RA mode)	PP lead threshold	PP floe threshold
CryoSat-2 (SAR)	0.3*	0.1*
Envisat (LRM)	0.3*	0.1*
ERS-2 (LRM)	0.2839	0.1328

* Guerreiro et al 2017

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

As suggested, we have added some precision to the interpolation procedure.

L236-237 : Outliers are filtered out with a three standard deviation threshold along 25 km sliding windows. ILA and SLA are interpolated respectively over leads and floes and smoothed with a 25 km rolling mean. The difference between the measured height over floes and over leads is finally made to retrieve the radar freeboard. For the remains of the study, we will only use the FBr measurements made above the floes because the characteristics of the waveforms are used.

replaced by:

ILA and SLA outliers are removed by filtering data that are outside the interval : rolling mean $\pm$ 3 rolling Standard Deviation, with a 60 km large sliding window. After filtering, ILA and SLA are smoothed using a rolling mean at 12.5 km, then SLA and ILA are linearly interpolated (including bellows the floes for SLA and above the leads for ILA) and are again smoothed using a rolling mean at 12.5 km. No limit of distance is used to discard radar freeboard, but the interpolation as well as smoothing and filtering is not done with values separated by land. Indeed, the processing is done within ocean segments, separated by land, in order to isolate statistics between segments. In this study, we will only use the FBr measurements that are made over the floes, indeed, the LRM data calibration, explained in section 3.4, is based on floes characteristics.

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

No, we don't discard freeboard above a max distance to the nearest lead. This information has been added commonly within the previous comment. To do so, we have applied the usual geophysical corrections in altimetry, taking into consideration the choice of these data so that they are particularly appropriate for polar oceans.

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?

The SLA bias comes from the choice to use an empirical retracker with the same fixed threshold for both leads and floes. Poisson et al 2018, explains that over rough surfaces such as ocean, the usual retracked point is close to the position of the half power of the waveforms. Specular waveform that can be found over leads should have retracked point higher, nevertheless, measurements are more stable with lower threshold (Poisson et al 2018 fig.9). This explains why we have a bias for the SLA.

Laforge et al 2020 shows that over leads comparing to physical retracker, the SLA bias is constant for altimeters in SARM, nevertheless this conclusion is also relevant for LRM as peaky waveform are all the same over leads. To remove this bias, another threshold should be use e.g. 80 or 90 or a calibration over another mission such as CS-2. The choice over a higher threshold can introduce noise due to the power sampling of the waveform. It could be noticed that this bias also occurs for mission in SAR mode, but it is much smaller due to the fact that SAR waveforms on leads are more peaky than for LRM.

L248-251. In LRM, most of this error comes from a constant bias on the Sea Level Anomaly

replaced by

Negative radar freeboards are mainly due to the retracker choice. Indeed, a TFMRA50 is used to retrack height on both leads and floes, this introduces a bias on the height over the leads. The TFMRA threshold to retrack heights over leads should be closer to 80% and because we use a threshold of 50% that corresponds to the position of the retrack point for ocean surfaces, not specular ones (Poisson et al 2018), the leads are measured higher than they are and even higher than the floes. The SLA bias (in leads) is evaluated constant for SARM altimeter in the study of Laforge et al 2020, this conclusion is also relevant for LRM altimeters as waveforms over the leads are peaky and similar from a lead to an other. This positive constant bias over the leads results to a negative bias on the radar freeboard. To avoid this bias, the retracker threshold could be adapted for leads or the SLA could be calibrated on CryoSat-2 one. Nevertheless, a threshold of 50% ensure the stability of the range (Poisson et al 2018, Fig.9) contrary to higher threshold (80%-95%) that could lead up to 47 cm of random error on the SLA. A TFMRA at 50 % for both leads and floes is preferred in this study as a constant bias is easier to correct than an undetermined random error.

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

Yes, CS-2 data is in SAR+SARIn mode here. This precision has been added to the plot (cf 4)

L278. Explain these terms.

The activation function is a sigmoid.

replaced by:

The activation function for the hidden layers neurons is a sigmoid, motivated by possible negative radar freeboard values and the optimizer is and ADAM [Kingma et Ba, 2014].

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?

Once the hyper-parameter combination is set, the MLP is trained with the whole dataset" means that the chosen model (with the optimal combination) is trained 'again' with 100% of the data (again, but weight are reinitialized). The training on 90% comparing to the one on 100% are not slightly different at all, but it's supposed to be better because with have trained it with a larger dataset (larger the dataset is, better the model is supposed to be). For instance, as we have only one winter for Envisat-ERS-2 mission overlap period, 10% is not negligible as it represents a bit less than a month. We have clarified this part, hope it would help the reader to understand better this part of the methodology.

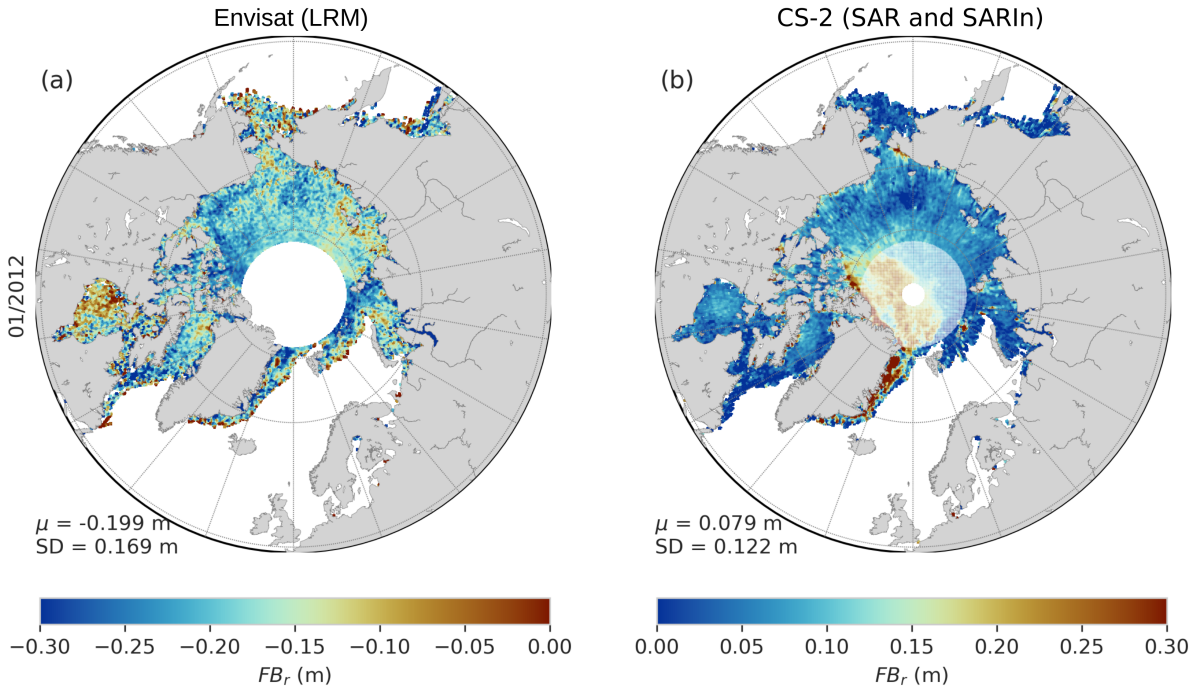


Figure 4. Pan-Artic radar freeboard maps for January 2012 for (a) Envisat uncorrected and (b) CryoSat-2

L283-284 : Hyper-parameters have been tuned by dichotomy by choosing at each step the hyper-parameter combination with the highest mean score (average score made on 5 models) on the test sample. The score used for this regression is the Pearson correlation coefficient. To determine the most suitable hyper-parameter combination, the dataset is randomly split into a training and a testing dataset, corresponding respectively to 90% and 10% of the initial dataset. The activation function used is a sigmoid. Once the hyper-parameter combination is set, the MLP is trained with the whole dataset.

replaced by:

The neural network used is a multilayer perceptron (MLP). Both calibrations have been processed with Scikit learn [Pedregosa et al, 2011]. The MLP is composed of 5 hidden layers, each composed of 100 neurons. The choice of hyperparameters : number of neurons, the learning rate, the regularization term, batch size, activation functions, solver for the weights optimization, have been done using gridding methodology, e.g. testing combinations and take the one that give best score. The evaluation criterion, called the score, is chosen as the determination coefficient. Models are trained on 90% of the dataset and tested on the remaining 10%, the splitting in random. During the tuning step, models are cross validated, it means that they are each trained 5 times with the same combination of hyperparameters but without the same train/test dataset, the 5 scores are then analyzed to determine the best combination. Cross validation give a better idea of the model performance as the dependence to the training dataset is limited. The activation function for the hidden layers neurons is a sigmoid, motivated by possible negative radar freeboard values and the optimizer is and ADAM [Kigma et Ba, 2014]. Moreover, in order to avoid over-fitting, an early stopping criterion is used to stop the model training as soon as the score is not improved during 10 consecutive iterations, with a defined tolerance.

Finally, once the hyperparameters combination is set, the MLP is trained on the whole dataset to provide the calibration function. The trained model is then applied to the LRM monthly grids to obtain a monthly LRM-corrected radar freeboard.

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?

We compute differently the SLA uncertainty for the floes (where we interpolate the SLA). Indeed, the rolling standard deviation of the interpolated SLA (where we do not have leads) does not really make sense to estimate the SLA uncertainty on floes. That's the reason why the uncertainty of the SLA where there are no leads is different and is assumed to be the difference between the estimated (interpolated and smoothed SLA) and the mean SLA, this method is taken from Ricker et al 2014. Concerning the along track mean elevation, no limit of distance was used, nevertheless, such as for the interpolation or the smoothing (section 3.2), they are computed per segments of water (so there is one mean SLA per segment of water). As we consider values on floes, we assume that the main part of the random uncertainty of the ILA (ice level anomayl, above the floes) comes from the speckle noise, even if ILA is smoothed, this could be a limitation of the random uncertainty budget. If we understood well the last part of your comment, this will not be the topography of the ice, but the difference between the SLA we have and the mean SLA, which is supposed to represent the SLA we should have. We suggest the following modification in the manuscript, with more details.

SLA uncertainty (σ_{SLA}) estimation depends on the surface type (leads or floes). For leads, we take the standard deviation of the measured height within a sliding 25 km window. Concerning floes, the uncertainty is estimated as the difference between the height measurement and the mean elevation along the track (Ricker et al., 2014).

replaced by :

SLA uncertainty (σ_{SLA}) is estimated to be the standard deviation of the SLA within a sliding window of 25 km if there are some leads within this window. If not, the SLA uncertainty is taken as the difference between the interpolated and smoothed SLA and the mean SLA computed as the mean values of measured SLA at leads within a segment of ocean (if the pass is over land, statistics are made segment of ocean by segment of ocean).

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).

The way the random uncertainties are computing during the gridding are slightly more complicated than that. We have done the same choice as in Ricker et al, 2014 and compute the grid-cell resulting radar freeboard with a weighted average, with the radar freeboard uncertainty as weights. For each grid-cell : $FB_r = \sqrt{\frac{\sum_1^N \frac{1}{\sigma_{R_i^2}} \cdot FB_{ri}}{\sum_1^N \frac{1}{\sigma_{R_i^2}}}}$*, so the resulting uncertainty computation should take it into account that weighted average as* $\sqrt{\frac{1}{\sum_1^N \frac{1}{\sigma_{R_i^2}}}}$*. Taken this method, it's difficult to draw a way to make uncertainties for the speckle noise and the SLA dropping differently, even we agree that theoretically the* σ_{SLA} *is correlated along the tracks, thus, the random uncertainty can be a little bit underestimated.*

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

Thank you for this confirmation. We adapted the sentence to the radar freeboard.

Roughness is estimated to be respectively about 20% and 30% of the sea ice thickness for FYI and MYI (Landy et al., 2020).
replaced by :
Roughness is estimated to be respectively about 20 % and 30 % of the sea ice thickness for FYI and MYI according to Landy et al., 2020, this results is also applicable for the freeboard

---

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?

We apologize for the confusion raised by this figure. The other inputs are the one in the grey box in the top right of Fig 6, and detailed in Fig.5, (except the date) : the concentration, the Leading edge slope, the Pulse Peakiness, and the Multi-year ice fraction. The uncertainty of these 4 variables is defined as : two times the standard deviation of the values of these variables within the grid cell divided by the sqrt of number of tracks within this grid cell. The uncertainty of the radar freeboard, as said in Sec 3.5, is neither multiplied by two nor divided by the number of tracks. In order to make this steps clearer, we have developed the legend of Fig.5 and Fig.6. and modified the following paragraph :
L316-317 : The uncertainty of the other inputs is considered to be, for each grid cell, the standard deviation of the measurements used to calculate the average value (grid cell value) divided by the number of tracks passing through the corresponding grid cell.

replaced by:
The uncertainty of the other inputs (LES, PP, sea ice concentration, MYI fraction), is considered to be, for each grid cell, two times the standard deviation of the measurements used to calculate the average value (grid cell value) divided by the number of tracks passing through the corresponding grid cell.

As well as :

L325-327 : The method consists of training a number M of NN with noisy inputs (noise has been added to all inputs according to a Gaussian distribution), and then to analyze the distribution of radar freeboard predictions from the M noisy NN applied on noisy N inputs for each considered grids. The whole uncertainty budget process is summarized in Fig. 6.

replaced by :

The method consists in training a number M of NN with noisy inputs. The noise has been added to all inputs (for each grid cell and each month) according to a Gaussian distribution centered on the estimated value and the corresponding uncertainty as standard deviation. The calibration processing (training and prediction) is done for all the noisy inputs/output, then the distribution of MxN radar freeboard predictions (from the M noisy NN models applied on N noisy inputs) has been analyzed of each grid cell and each month. The whole uncertainty budget process is summarized in Fig. 6.

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.

We hope we have clarified this information with the modifications made for the previous comment.

L340. For which months in Fig 8?

This missing information has been added to the manuscript.

Figure 8 presents the same feature for Envisat and ERS-2 radar freeboard during December 2002 and April 2003.

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

We have added this information in the manuscript.

The higher mean difference is 7 mm and concerns February 2011, the Envisat calibration. For the ERS-2 calibration, the mean freeboard difference with Envisat does not exceed 3 mm. Concerning all the overlap times, the mean difference is 3 mm for Env/CS-2 calibration, and -2 mm for ERS-2/Env one.

has been replaced by :

The highest mean difference reaches 7 mm in February 2011 for Envisat calibration, i.e. 9.5% of the mean Envisat radar freeboard. For ERS-2 calibration, the mean freeboard difference between ERS-2 and Envisat does not exceed 3 mm, 3.3% of ERS-2 mean radar freeboard. Concerning all the overlap times, the mean difference is 3 mm for Env/CS-2 calibration, 4.1% of Envisat mean radar freeboard and -2 mm for ERS-2/Env one, about 2.2% of ERS-2 mean radar freeboard.

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

This information has been added to the previous paragraph (see previous comment).

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

Yes, this information has been added as following :

Similarly, for the period 2002/2003, the mean and median uncertainties of ERS-2 are always larger than those of Envisat by about 6 cm over the median radar freeboard (see Tab.A5 and Tab.A3 for more statistics)

has been replaced by :

Concerning the period 2002/2003, the median uncertainty is 8 cm for ERS-2 radar freeboard and 7.3cm for Envisat, similarly, statistics on uncertainties are globally higher for ERS-2 estimates (see Tab.A3 and Tab.A4 for detailed statistics).

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

Unfortunately, in spite of the different filtering, spurious tracks remains, especially close to the coast. Nevertheless, few validation data are available in these areas, so comparisons with independent data sets would not be contaminated with these spurious tracks.

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

Figure 7 caption has been modified as followed :

Comparison of Envisat calibrated radar freeboard against CryoSat-2 reference for December 2010 in the upper half and April 2011 in the lower half. The maps (a) and (g) refers to Envisat aside with corresponding CryoSat-2 radar freeboard (b) and (h). Maps bellow (d), (e), (j) and (k) are the related uncertainties. The right column presents differences freeboard maps (Env-CS-2) ((c) and (i)). (f) and (l) are the distribution of Envisat F Br in red, CryoSat-2 F Br in blue and in grey.

has been replaced by

Comparison of Envisat calibrated radar freeboard against CryoSat-2 reference for December 2010 in the upper half and April 2011 in the lower half. The maps (a) and (g) refers to Envisat aside with the corresponding CryoSat-2 radar freeboard (b) and (h). Maps bellow (d), (e), (j) and (k) are the related uncertainties. The right column presents freeboard difference maps (Env-CS-2) ((c) and (i)). (f) and (l) are the distribution of Envisat FBr in red, CryoSat-2 FBr in blue and ΔFBr in grey. Histograms only include common data between Envisat and CryoSat-2, data north or 81.5 °N are excluded. μ refers to the average and SD to the Standard Deviation.

L384. 'static data'?

'Static data' refers to moorings, data set have fixed longitude and latitude for a given time. We have clarified this point as following:

Static data are monthly averaged to get one value per month.

replaced by :

Concerning moorings or coastal measurement stations, data are averaged to get one value per month.

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.

Thank you for this suggestion. We agree with this point of view. This comparison was a part of the validation for completeness, but also to get feedback from the community as we were struggling while using this data. To remove the confusion, we decided not to compares Envisat data with IMB as advised.

L397. Could be attributed to, but not definitely.

We have changed this sentence to stay more hypothetical.

The bias between OIB and Envisat estimation can also be attributed to the OIB snow depth estimation that remains slightly different from one algorithm to another (Kwok et al., 2017).

replaced by :
The bias between OIB and Envisat estimation could also be partly attributed to the OIB snow depth which estimation seems sensitive to the algorithm used (Kwok et al., 2017).

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

Captions of Figure 7,8,9,10,11,12 have been modified to add the statistic acronyms definitions.

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?

This could tell something about the calibration, but it's hard to conclude knowing that BGEP comparisons and other moorings comparisons draw different conclusions. Considering BGEP, calibration seems to overestimate thin ice which is not the case comparing to moorings within the Laptev sea.

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

We don't really know how to compare properly the Standard deviation and biases with our uncertainties as the question of the uncertainties on the SIT is another issue and for this paper the uncertainties are limited to the radar freeboard. We propose the Figures 10,11,12 and 13 that present the 95% confidence interval of the SIT but without taking into account uncertainties on snow depth, densities etc for the FBr to SIT conversion step.

The plots have been updated with the variable snow density as asked by Robbie Mallet so they are not similar to those in the previous version of the manuscript. For esthetical reason, bounds are not represented for comparisons with other satellite-based SIT estimation.

L420-422. What are the statistics like for CS-2 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?

Just for information, Figure 9 show comparisons for CS-2 thickness with OIB, BGEP and Transdrift Laptev Sea. For BGEP bias is higher, CS-2 FYI thickness seems to be more overestimated than Envisat one but with find this same negative bias for Transdrift Laptev Sea moorings especially for thick ice.

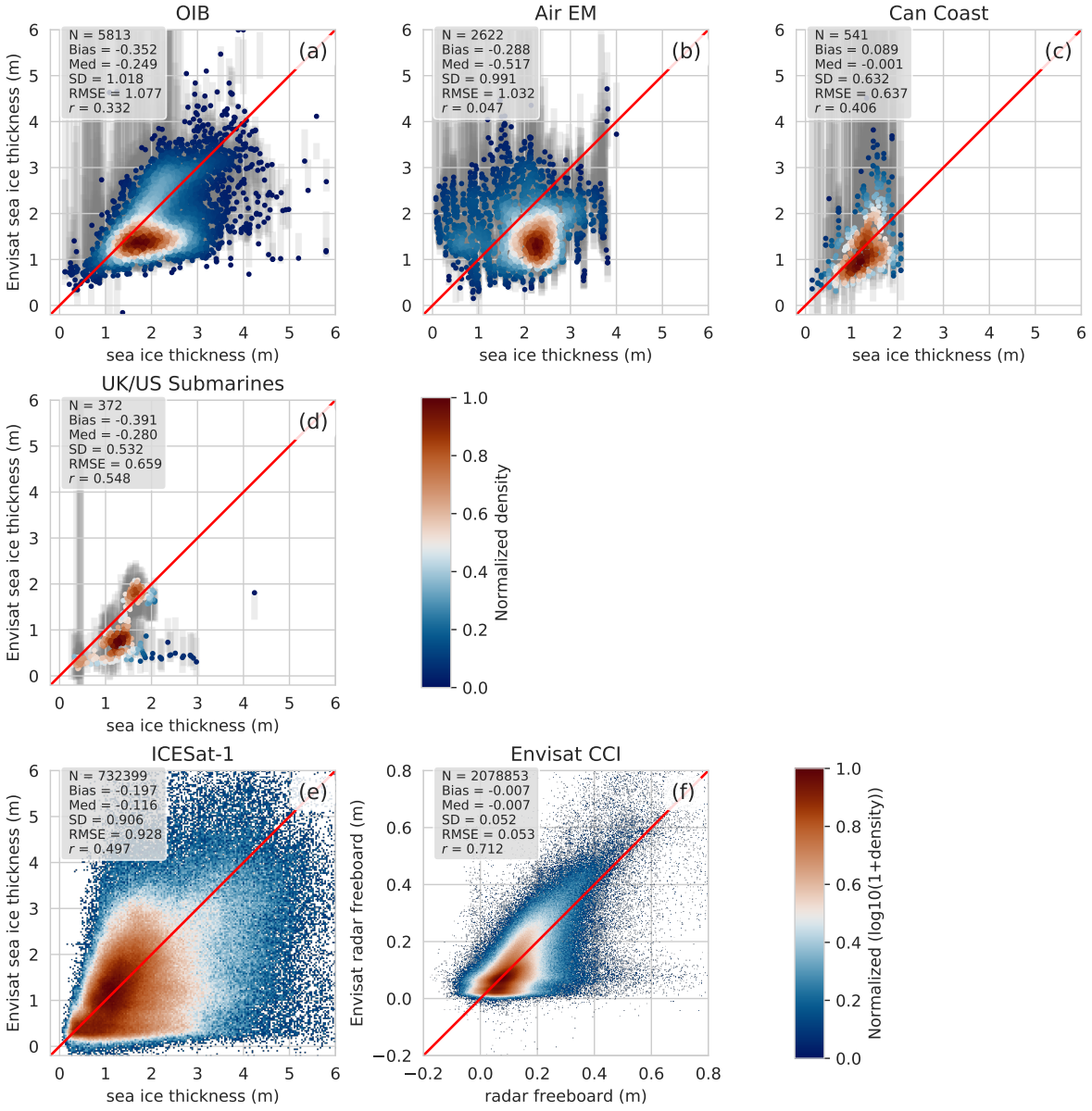


Figure 5. Comparative scatter-plots between Envisat sea ice thickness or radar freeboard estimations and other data sets. The x-axis indicates the sea ice thickness from (a) OIB total ice freeboard, (b) Air EM snow plus ice thickness, (c) Can Coast ice thickness, (d) UK/US submarines draft and (e) ICESat-1 total freeboard. (f) compares our Envisat radar freeboard with SI-CCI Envisat solution. Colorbars represent the normalized density. A log_{10} has been applied before the normalization for (e) and (f) due to the large number of data. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and r the correlation coefficient.

In order to help the reader, with have added the following sentences:

As a comparison, the bias between CryoSat-2 and OIB between 2010 and 2019 is about 16 cm and the RMSE 77 cm. Concerning BGEP (2010-2021) comparisons, the bias is 21 cm with the same overestimation of FYI thickness for CS-2 and with Transdrift Laptev Sea (2010-2016) comparisons show a negative bias of -38 cm.

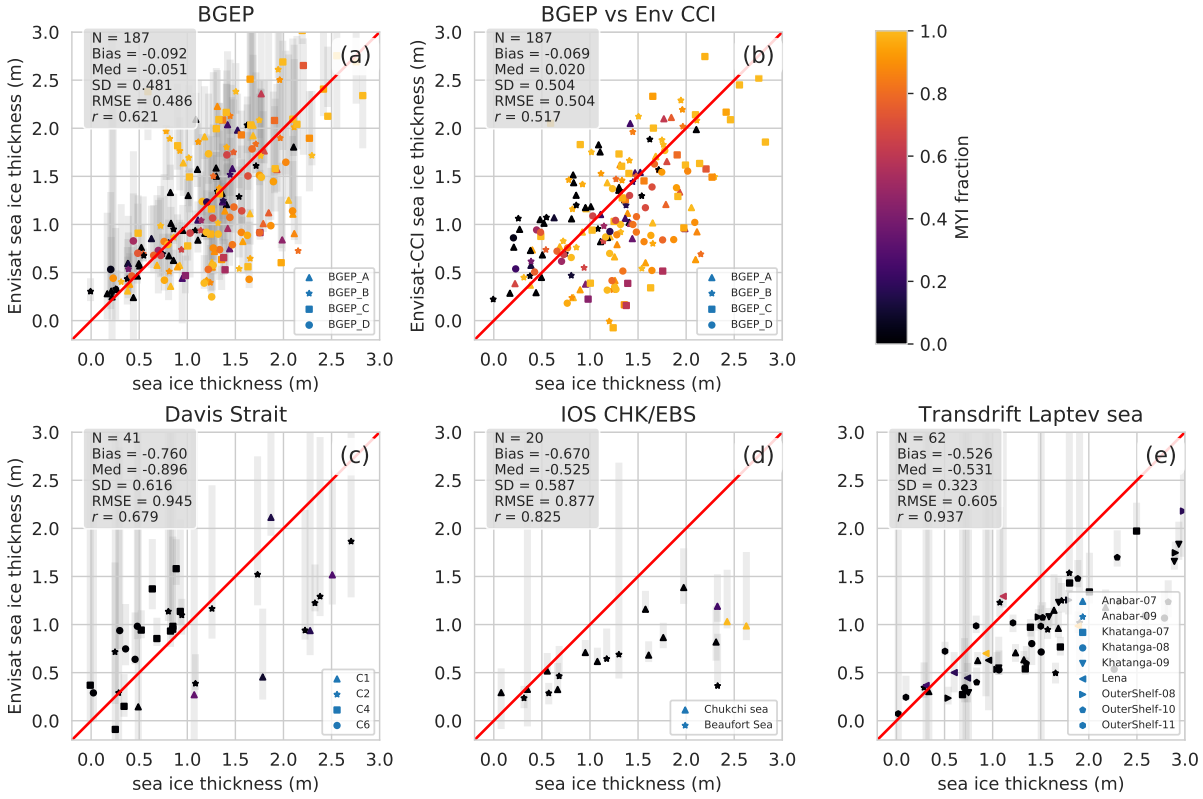


Figure 6. Comparative scatter-plots between Envisat sea ice thickness estimations and anchored moorings data sets. Each dot corresponds to a monthly averaged value. The x-axis indicates the sea ice thickness from (a) BGEP, (b) BGEP vs Env CCI, (c) Davis Strait, (d) IOS CHK/EBS and (e) Transdrift Laptev Sea ice draft. The colorbar shows the MYI fraction. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and r the correlation coefficient.

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.

The same plots with Warren 99 modified climatology are presented in Fig.10, Fig. 11, Fig. 12 and Fig. 13. Note that the scatters for OIB, CanCoast and EnvisatCCI are unchanged because no additional snow products are used. Biases with Air EM is reduced as well as for the submarines. Nevertheless, as explained in the manuscript, a bias is expected when comparing to Submarines of about 30 cm. The dispersion is augmented for the comparisons with moorings data. Thus, the snow load has an important impact, we suggest showing this figure in appendice, as it could impact the clarity of validation if it is shown in the validation section.

L427. Is Section 2.3 correct?

Sorry for this typo mistake, the reference was indeed "section 4.2". This has been corrected.

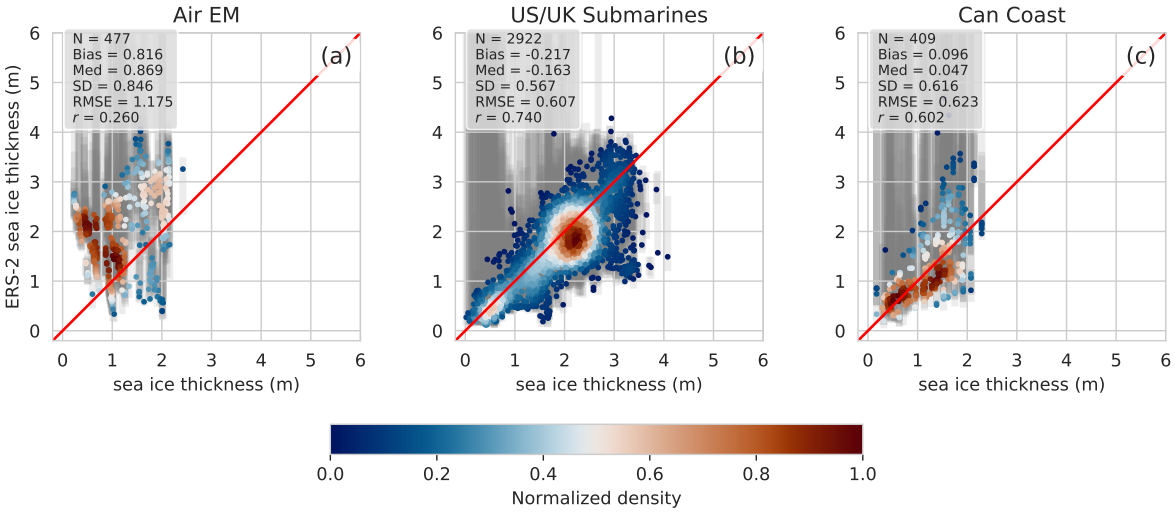


Figure 7. Comparative scatter-plots between ERS-2 sea ice thickness estimations and 3 in-situ data sets. The x-axis indicates the sea ice thickness from (a) AirEM total thickness, (b) UK/US Submarines draft and (c) Can Coast sea ice thickness. Colorbar indicates the normalized density. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and r the correlation coefficient.

415

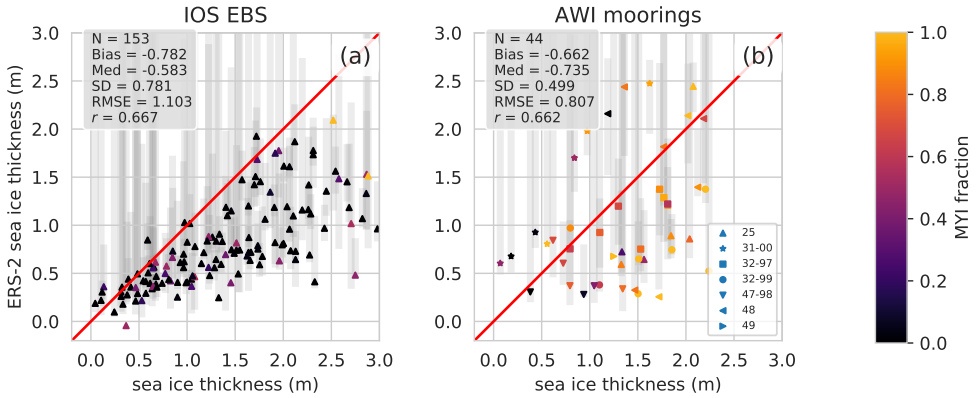
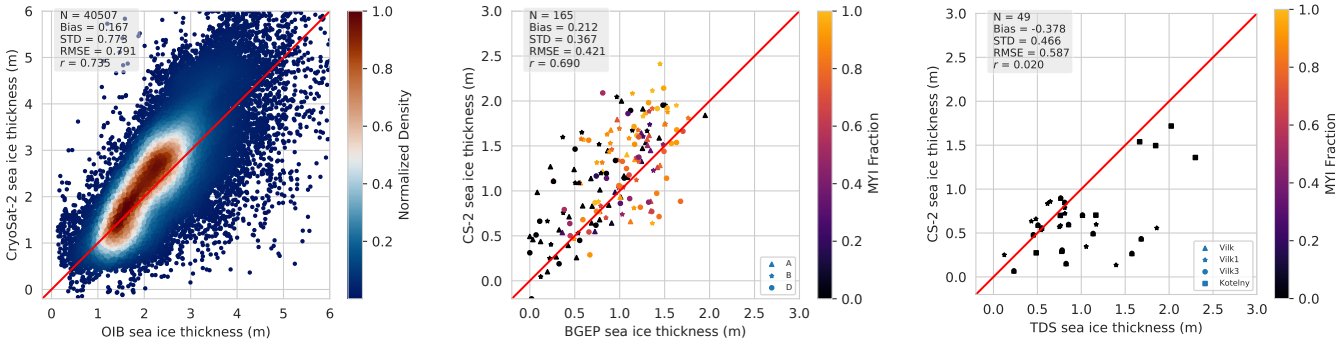


Figure 8. Comparative scatter-plots between ERS-2 sea ice thickness estimations and 2 anchored moorings data sets. The x-axis shows sea ice thickness estimations from (a) IOS Beaufort Sea and (b) AWI moorings sea ice draft. The color bar indicates the respective MYI fraction. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and r the correlation coefficient.

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

The caption has been clarified as following :

Fig 13 caption : Time series representing radar freeboard volume up to 81.5°N for each winter month for ERS-2 in orange, Envisat in teal and CS-2 in dark red. Blue triangles are winter mean volumes. Red lines are linear regressions of winter mean



(a) OIB (b) BGEP (c) Transdrift

Figure 9. Comparative scatter-plots between CryoSat-2 sea ice thickness estimations and OIB/BGEP. The x-axis indicates the sea ice thickness from (a) OIB total thickness, (b) BGEP draft and (c) Transdrift Laptev Sea draft. Colorbar indicates the normalized density and MYI fraction. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and r the correlation coefficient.

volume until 2002/2003 for dashed line and 1995/1996 for solid line.

replaced by:

Fig 13 caption : Time series representing radar freeboard volume up to 81.5°N for each winter month (no data between 81.5 °N and 90 °N, even for CS-2 for consistency). ERS-2 in orange, Envisat in teal and CS-2 in dark red. Blue triangles are winter mean volumes. Red lines are linear regressions of winter mean volume until 2002/2003 for dashed line and 1995/1996 for solid line.

addition of : It's important to note that the volumes presented in Fig. 13 are only considering values up to 81.5 °N.

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.

We apologize to show so limited data set, we feel that the paper is already rather long as it is and this will be the subject of another study.

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

It's actually the pLRM mode, this has been explained in a previous answer, we hope that the use of this table is now clearer.

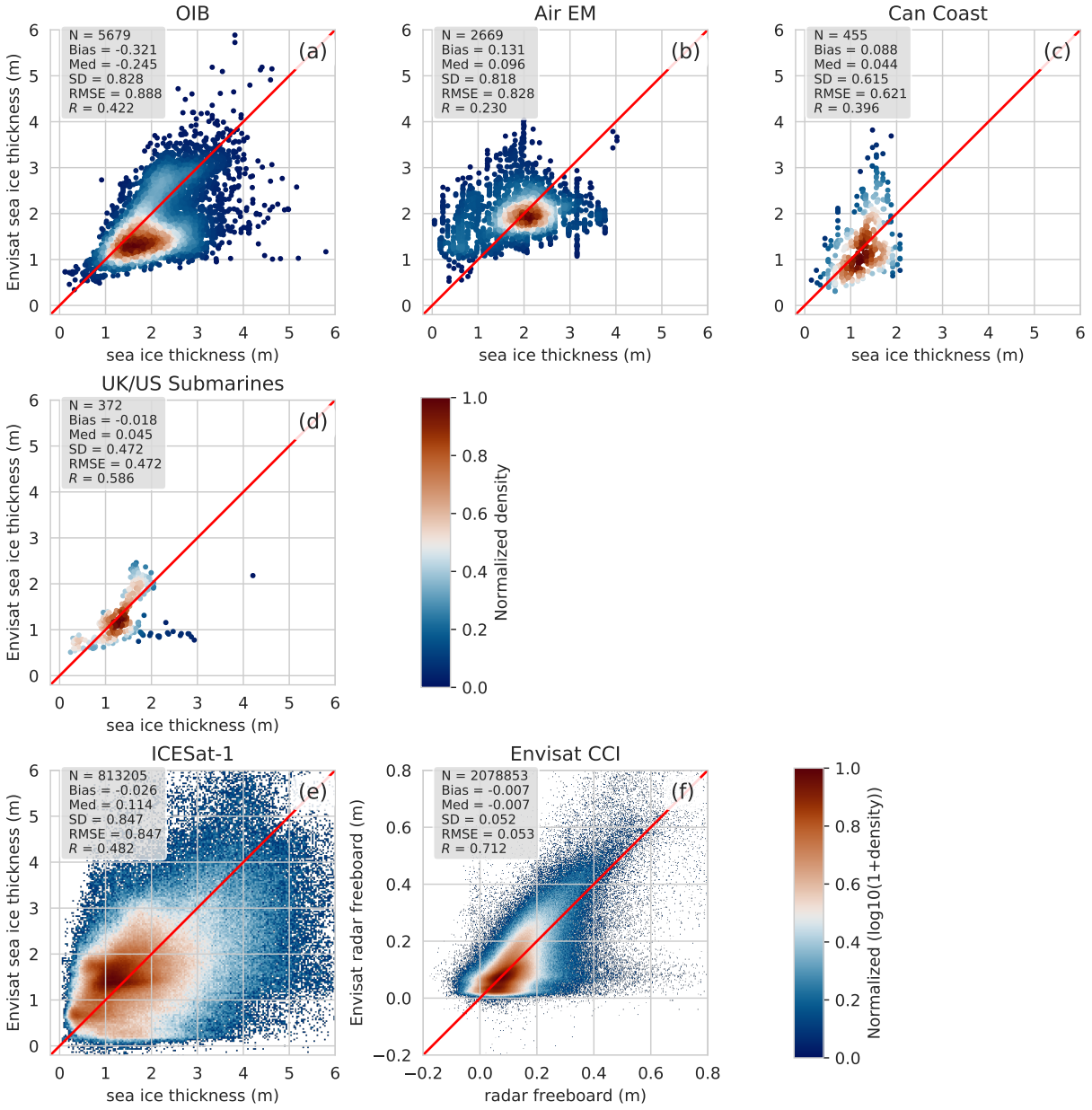


Figure 10. Comparative scatter-plots between Envisat sea ice thickness or radar freeboard estimations and other data sets. The x-axis indicates the sea ice thickness from (a) OIB total ice freeboard, (b) Air EM snow plus ice thickness, (c) Can Coast ice thickness, (d) UK/US submarines draft and (e) ICESat-1 total freeboard. (f) compares our Envisat radar freeboard with SI-CCI Envisat solution. Colorbars represent the normalized density. A log_{10} has been applied before the normalization for (e) and (f) due to the large number of data. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and R the correlation coefficient.

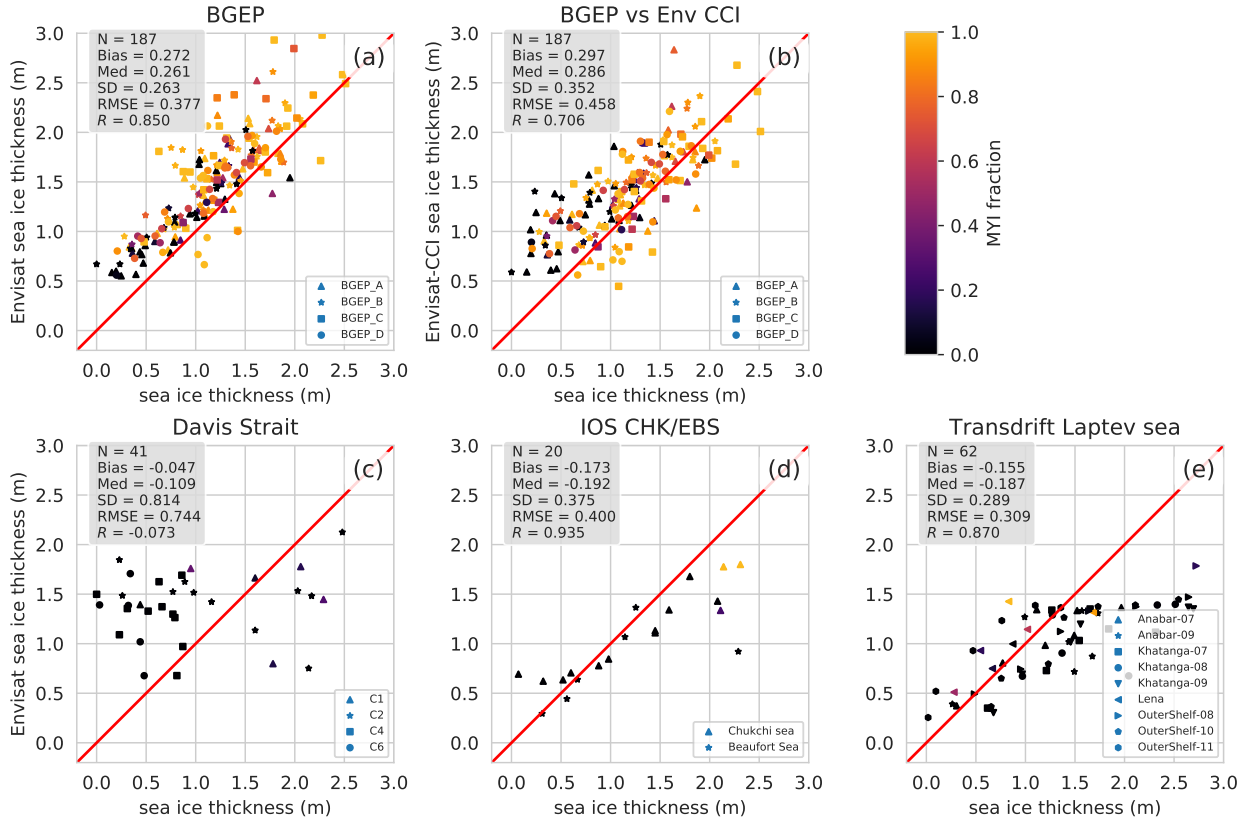


Figure 11. Comparative scatter-plots between Envisat sea ice thickness estimations and anchored moorings data sets. Each dot corresponds to a monthly averaged value. The x-axis indicates the sea ice thickness from (a) BGEP, (b) Davis Strait, (c) IOS CHK/EBS and (d) Transdrift Laptev Sea ice draft. The colorbar shows the MYI fraction. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and R the correlation coefficient.

References

Guerreiro, K., Fleury, S., Zakharova, E., Kouraev, A., Rémy, F., and Maisongrande, P.: Comparison of CryoSat-2 and ENVISAT radar freeboard over Arctic sea ice: toward an improved Envisat freeboard retrieval, The Cryosphere, 11, 2059–2073, https://doi.org/10.5194/tc-11-2059-2017, 2017.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, CoRR, 2014.

Laforge, A., Fleury, S., Dinardo, S., Garnier, F., Remy, F., Benveniste, J., Bouffard, J., and Verley, J.: Toward improved sea ice freeboard observation with SAR altimetry using the physical retracker SAMOSA+, Advances in Space Research, p. S0273117720300776, https://doi.org/10.1016/j.asr.2020.02.001, 2020.

Landy, J. C., Petty, A. A., Tsamados, M., and Stroeve, J. C.: Sea Ice Roughness Overlooked as a Key Source of Uncertainty in CryoSat-2 Ice Freeboard Retrievals, Journal of Geophysical Research: Oceans, 125, https://doi.org/10.1029/2019JC015820, 2020.

Landy, J. C., Dawson, G. J., Tsamados, M., Bushuk, M., Stroeve, J. C., Howell, S. E. L., Krumpen, T., Babb, D. G., Komarov, A. S., Heorton, H. D. B. S., Belter, H. J., and Aksenov, Y.: A year-round satellite sea-ice thickness record from CryoSat-2, Nature, 609, 517–522, https://doi.org/10.1038/s41586-022-05058-5, number: 7927 Publisher: Nature Publishing Group, 2022.

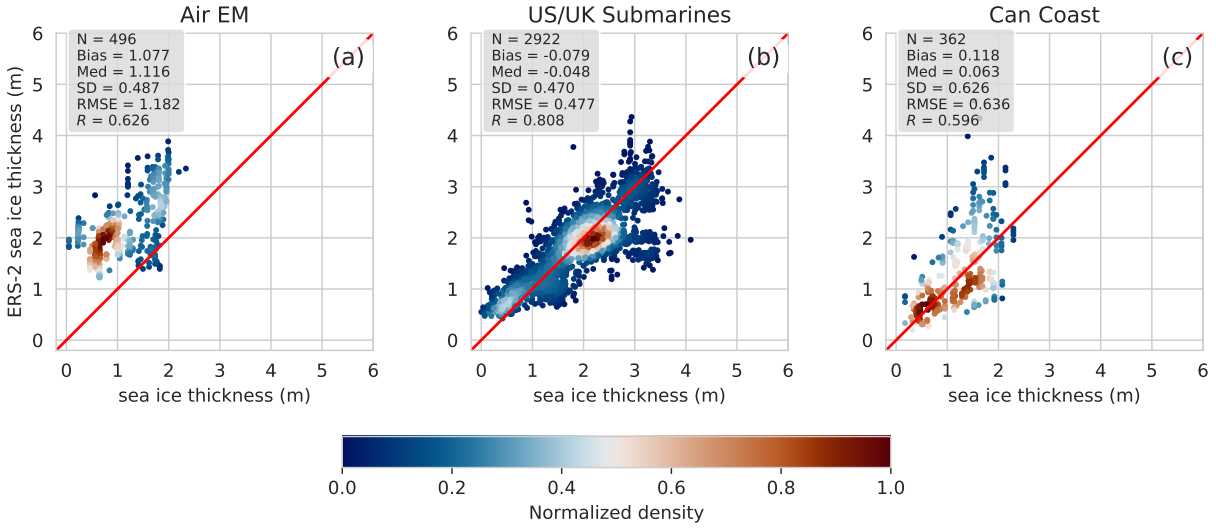


Figure 12. Comparative scatter-plots between ERS2 sea ice thickness estimations and 3 in-situ data sets. The x-axis indicates the sea ice thickness from (a) AirEM total thickness, (b) UK/US Submarines draft and (c) Can Coast sea ice thickness. Colorbar indicates the normalized density. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and R the correlation coefficient.

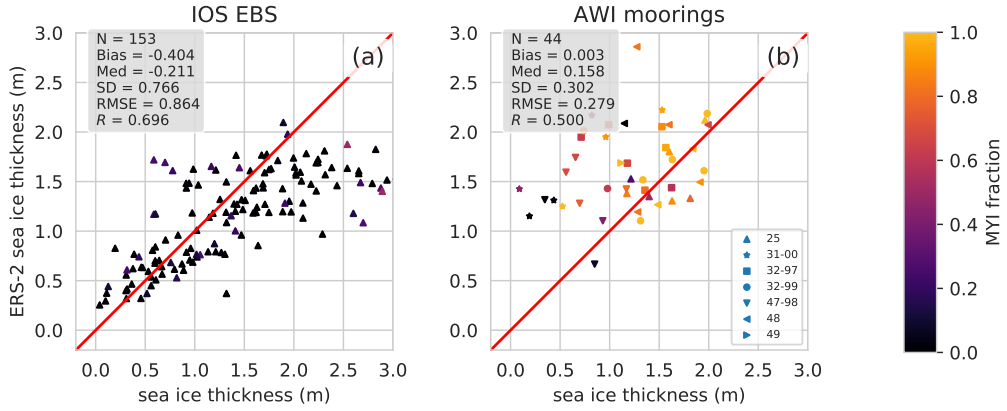


Figure 13. Comparative scatter-plots between ERS2 sea ice thickness estimations and 2 anchored moorings data sets. The x-axis shows the sea ice thickness measurements from (a) IOS Beaufort Sea and (b) AWI moorings sea ice draft. The color bar indicates the respective MYI fraction. N is the number of the couple of values that are compared, Med refers to the Median, SD the Standard deviation, RMSE the Root Mean Square Error and R the correlation coefficient.

Nandan, V., Geldsetzer, T., Yackel, J., Mahmud, M., Scharien, R., Howell, S., King, J., Ricker, R., and Else, B.: Effect of Snow Salinity on CryoSat-2 Arctic First-Year Sea Ice Freeboard Measurements: Sea Ice Brine-Snow Effect on CryoSat-2, Geophysical Research Letters, 44, 10,419–10,426, https://doi.org/10.1002/2017GL074506, 2017.

Paul, S., Hendricks, S., Ricker, R., Kern, S., and Rinne, E.: Empirical parametrization of Envisat freeboard retrieval of Arctic and Antarctic sea ice based on CryoSat-2: progress in the ESA Climate Change Initiative, The Cryosphere, 12, 2437–2460, https://doi.org/10.5194/tc-12-2437-2018, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Poisson, J.-C., Quartly, G. D., Kurekin, A. A., Thibaut, P., Hoang, D., and Nencioli, F.: Development of an ENVISAT Altimetry Processor Providing Sea Level Continuity Between Open Ocean and Arctic Leads, IEEE Transactions on Geoscience and Remote Sensing, 56, 5299–5319, https://doi.org/10.1109/TGRS.2018.2813061, 2018.

Raney, R.: A delay/Doppler radar altimeter for ice sheet monitoring, in: 1995 International Geoscience and Remote Sensing Symposium, IGARSS '95. Quantitative Remote Sensing for Science and Applications, vol. 2, pp. 862–864, IEEE, Firenze, Italy, https://doi.org/10.1109/IGARSS.1995.521080, 1995.

Rheinlænder, J. W., Davy, R., Ólason, E., Rampal, P., Spensberger, C., Williams, T. D., Korosov, A., and Spengler, T.: Driving Mechanisms of an Extreme Winter Sea Ice Breakup Event in the Beaufort Sea, Geophysical Research Letters, 49, https://doi.org/10.1029/2022GL099024, 2022.

Ricker, R., Hendricks, S., Helm, V., Skourup, H., and Davidson, M.: Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation, 8, 1607–1622, https://doi.org/10.5194/tc-8-1607-2014, 2014.

Stammer, D.: Satellite altimetry over oceans and land surfaces, Earth observation of global changes, 2018.

Stroeve, J. and Notz, D.: Changing state of Arctic sea ice across all seasons, Environmental Research Letters, 13, 103 001, https://doi.org/10.1088/1748-9326/aade56, publisher: IOP Publishing, 2018.

Tilling, R., Ridout, A., and Shepherd, A.: Assessing the Impact of Lead and Floe Sampling on Arctic Sea Ice Thickness Estimates from Envisat and CryoSat-2, Journal of Geophysical Research: Oceans, 124, 7473–7485, https://doi.org/https://doi.org/10.1029/2019 2019.

Tschudi, M., Meier, W. N., Stewart, J. S., Fowler, C., and Maslanikand, J.: EASE-Grid Sea Ice Age, https://doi.org/10.5067/UTAV7490FE type: dataset, 2019.

Wingham, D. J., Francis, C. R., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S. W., Mallow, U., Mavrocordatos, C., Phalippou, L., Ratier, G., Rey, L., Rostan, F., Viau, P., and Wallis, D. W.: CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields, 37, 841–871, https://doi.org/10.1016/j.asr.2005.07.027, 2006.