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

Robbie Mallet (referee n°2) - global comment

Global comment as a community comment :

I was really pleased to see this paper come up in TCD. I think that generating radar freeboard data from ERS1/2 is one of the most pressing tasks for the sea ice community, and so I agree with Jack Landy's review. Overall I think the paper is well written and addresses what is a very significant gap in our knowledge of the Arctic Ocean. In particular I think the figures are well-designed. I do have a couple of concerns, questions and suggestions over wordings, citations etc. I hope the authors will take these in the spirit of discussion, rather than as negative criticism. I really do think that this research is high-quality and useful.

Global comment as a referee :

I left a community comment on this manuscript (https://doi.org/10.5194/egusphere-2022-214-CC1) before being nominated as a referee. I have therefore read and considered the manuscript again. As part of this, I investigated the data that was made available to me as a nominated reviewer. I wanted to see the size of the correction/calibration applied by the neural network presented in this paper. This has led me to question the nature of the 'correction' being applied, and whether it is reasonable to present this data product as a series of 'corrected' radar freeboard values at all. I would like to review this manuscript again once the queries raised here have been addressed.

Answer to Robbie Mallet (referee n°2) - global comment

We would like to thank the reviewer for his careful reading of the manuscript and for the relevant 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.

In our understanding, the main concern of the reviewer is : "To what extent can we claim that the resulting product is a corrected or calibrated retrieval when it doesn't reflect the variability in the raw, retracked values ? " Expressed in other word, the referee states that "nothing of the original radar freeboard measurement remains in the corrected value" and that is an issue.

First, we would like to indicate that your detailed analysis on the correlations between raw freeboard and calibrated radar freeboard have pointed out difficulties that have motivated the calibrations (past studies) and thereafter the use of a neural network.
Using the exact same processing chain as for CryoSat-2 (with a TFMRA-50 retracker), the Envisat, and ERS radar freeboard estimates are very different to what we are supposed to observe in terms of magnitude and spatial patterns. Indeed, LRM

waveforms are strongly impacted by the size of the footprint which is much larger in LRM than in SAR Mode ($\sim$ 180 km^2 to $\sim$ 5 km^2 [Stammer et al, 2018]). This is the main cause for the misfit between Envisat and CS-2 radar freeboards. In order to deal with this issue, differences between CS-2 and Envisat have been analyzed, please see Guerreiro et al. 2017, Paul et al. 2018 and Tilling et al. 2019 for a complete overview. As mentioned in the manuscript, the first two studies point out that the radar freeboard differences between the two altimeters are correlated (not especially linearly correlated) to the sea ice roughness, characterized by the waveform backscatter, the leading edge width or the pulse peakiness. The third study identified a link between the misfit and the distance between floes and leads.
The optimal solution would be to find a theoretical model, such as the Brown's model over open ocean, to represent the radar response over sea ice in order to correctly retrack the waveform. Despite significant progress in SAR mode (SAMOSA+, LARM, etc.), these models are not yet able to represent all the complexity of this response even in SARM. For instance, they are not able to represent snow penetration effects (i.e. volume backscatter effects). Moreover, in LRM, no study reports relevant retracked height over sea ice floes with a physical retracker and the complexity of the response is still poorly understood. The objective of this paper is neither to model any effect of the ice surface condition, nor to understand its influence on the FBr but rather to reconstruct the best possible ERS-2 radar freeboard with our actual knowledge consistently with Envisat and CS-2 ones. To do so, roughness or more globally sea ice surface state proxies are used to post-correct the estimated radar freeboard using as a reference Envisat previously calibrated on CS-2. Our study is based on the principle that the radar freeboard computed with a TFMRA50 from LRM waveforms is strongly polluted by the surface roughness. Then, we propose to calibrate LRM radar freeboard on CS-2 using some parameters characterizing the sea ice surface roughness. The same methodology is applied to calibrate ERS-2 radar freeboard on a CryoSat-2 like radar freeboard from Envisat. Thereafter, some other parameters such as the ice concentration or the sea ice age were added to improve and consolidate the learning of the NN so to reach a better match with CS-2 (in the case of Envisat and Envisat calibrated for ERS-2).
Unlike the review suggests, we would like to specify that the sea ice age is not directly used in the regression, we use a MYI fraction. The way this fraction is calculated has been developed in the manuscript, but it is not discrete values, as it is considered by the reviewer. Also, we would like to specify that correlations calculated with a variable that takes only two values can not be relevant.

To illustrate that the calibration is based on the PP and the LES, Figure 1 shows radar freeboard for April 2011 for CS-2, for Envisat with the calibration presented in the manuscript and one with another model trained only with the raw freeboard, the Pulse Peakiness and the Leading edge slope. It shows that these three parameters are sufficient to represent the magnitude and the patterns we are supposed to see in the Arctic. The other parameters help the calibration to get closer to CS-2 radar freeboard and bring more spatial variability.

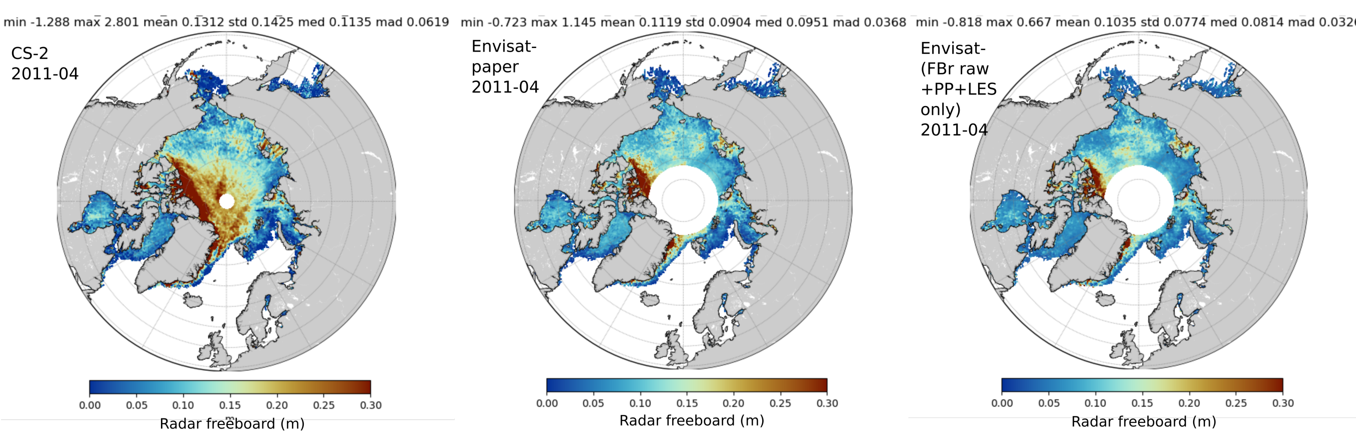

Figure 1. Radar freeboard from CS-2 (left), Envisat calibrated presented in the paper (middle) and Envisat using only the raw freeboard+LES+PP (right)

The correction we have to process is strongly non-linear, so this is the reason why we have chosen a neural network approach, which has the specificity to handle well with non linearities. Then, correlation between parameters based on linear approximation are not representative of the dependencies between parameters (inputs/output) in the neural network. Indeed, it is much more complicated to estimate the relative importance of each parameter in a regression, and it is not given by the correlations between the inputs and the predicted value. As it has been already mentioned, the main reason is that the relations that have been established by the neural network are not linear, while the correlation only evaluate whether the variables are related by a linear relation. Figure 2 shows the "partial dependencies" which refers to an illustration (statistically computed) of the relations between each input parameters (x-axis) and the predicted value (y-axis). It also illustrates the relative importance of each parameter: a parameter with no influence would have a horizontal curve as a mean state but it is not a quantitative approach. Partial dependency plots should be interpreted with caution, it refers to the mean state of statistical computation, depends on a discretization choice and values of input parameters have been standardized (mean = 0 and Standard deviation=1). The Figure 2 presents 2 panels, one for Envisat calibration (left) and one for ERS-2 calibration (right). Nevertheless, we can say that curves are not linear, no input is unused or with a very low influence, we can also note that LES and PP have the largest influence on the predicted radar freeboard.

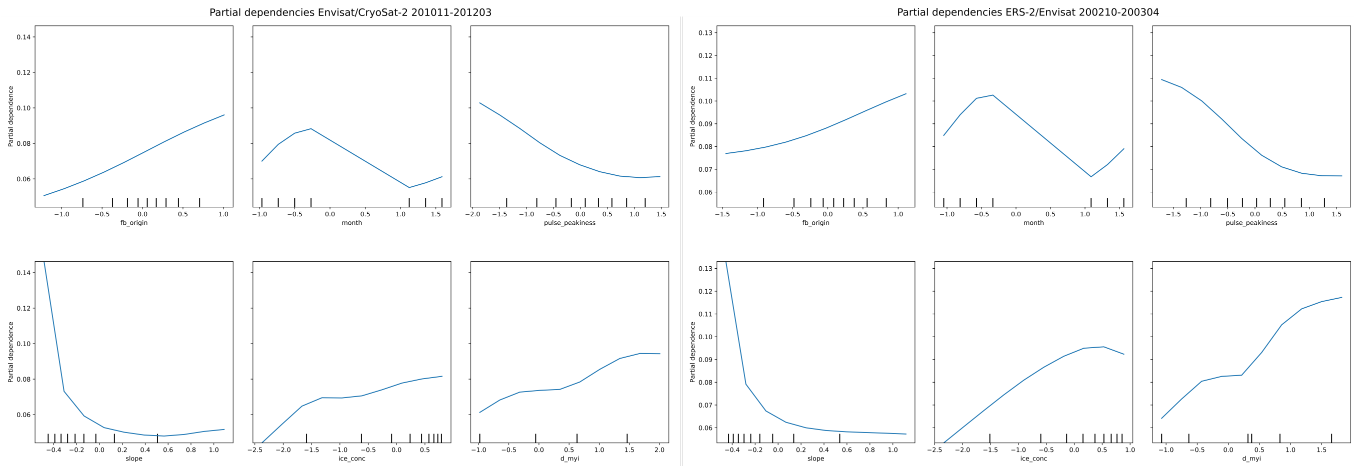

Figure 2. Partial dependencies plots for Envisat and ERS-2 calibration, from top left to top right, inputs are, the raw radar freeboard, the month, the pulse peakiness, the Leading edge slope, the concentration and the f_{MYI}.

The question of the non-linearity is central in our study but also in your analysis. But, since the correction is not linear at all, the largest raw radar freeboard is not necessarily the largest corrected freeboard, just as it is not the largest raw freeboard that will benefit from the largest correction. The raw radar freeboard, even noisy, still gives information on how the altimeter perceives the surface and how much it should be corrected, which remains an important information. We expect that the raw freeboard define the space and time variability of the calibrated radar freeboard over the whole period but this is hard to show since we don't have any reference of the expected variability of the SIT/FBr/FB during 1995-2010. To enhance the fact that the raw radar freeboard impacts the corrected radar freeboard, figure 3 shows the relative difference between the predicted FBr of the NN presented in the paper and one from a NN trained without the raw FBr for April 2011. It shows that for a large part of the basin, the difference of FBr is up to 25% of the predicted radar freeboard.

Finally, it's important to keep in mind that we have trained the neural network to reach the best score i.e. the best coefficient of determination (compared to CS-2 for Envisat and to Envisat corrected for ERS-2). Choosing the best NN, means choosing the combination of hyperparameters and even the choice of input parameters that gives the best scores. This means that the fraction of MYI allows to better fit CS-2 radar freeboard, that's why we keep it. However, it's even expected to find a good correlation between the sea ice age and the sea ice freeboard because, in average, older ice will be thicker.

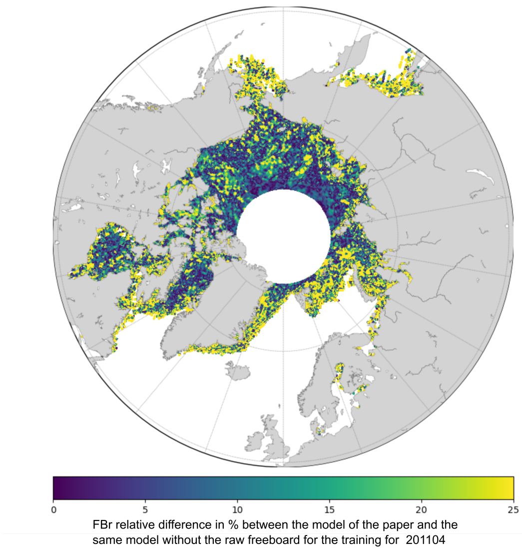

Figure 3. Relative difference in % between the corrected freeboard of our study and one with the same NN trained without the raw freeboard

To sum up, the purpose of this paper is to retrieve a consistent radar freeboard estimation for ERS-2 using the current knowledge on LRM waveforms over sea ice. Because LRM waveforms are highly impacted by the surface state and poorly understood over sea ice, raw freeboard have to be calibrated. Two calibrations need to be implemented to get consistent ERS-2 radar freeboard, first Envisat against CS-2 and then ERS-2 against Envisat calibrated radar freeboard. The calibration is first based on surface roughness proxy because evident link have been emphasized with the size of the correction (previous studies) and secondly on auxiliary data that were used to reach better fit with CS-2. The calibrated radar freeboard is partly driven by the raw radar freeboard, both parameters are not linear correlated as it would say that the calibration did not perform well. The "age" or in our case the MYI fraction is not the key input for the NN training. Furthermore, 'ice with a higher-than-average raw FBr in a given month" can not necessarily "end up with a higher than average corrected FBr value" as the calibration is not linear.

Please find below the details on how your specific comments have been taken into account. We have split in two part the specific comments, as the referee gives two detailed comments, one as a community and one as a referee. *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 Robbie Mallet (referee n°2) : specific comments

Specific comments - Referee comment

L280: I think you should by convention use the coefficient of determination rather than Pearson-r as a test score. Otherwise you'll end up with highly correlated relationships that have the wrong slope?

This is an error in the manuscript, we do use the coefficient of determination as the score for our regression. The correction has been made in the manuscript.

You need to explain quite a lot more about what's going on in Figure 6. The manuscript should not feature undefined letters and symbols, and there are many in this figure.

Caption has been largely developped as following :

Summary diagram of the uncertainty budget from along track to the propagation by the neural network.
replaced by:
Summary diagram of uncertainty budget during along track, gridding and calibration steps. Top left panel corresponds to the along track to grid uncertainty budget. Top right panel defines the notations, for the Monte Carlo procedure : Ω for the Neural Network input parameters, Γ for the Neural Network output parameter (radar freeboard) with σ_Ω and σ_Γ, the corresponding uncertainties. The middle panel corresponds to the training of M models with noisy inputs and outputs. Bottom panel show the predictions of the N noisy input with the M neural network trained. γ is the predicted radar freeboard estimation for one pixel of the MxN predictions. M=100, N=200.

Similar to above, you should explain much more about what's going on between lines 277 285. Papers in The Cryosphere should be accessible to scientists without extensive experience in machine learning. Don't be afraid to use the supplement for this, as I appreciate it's wordy. For instance, why did you choose 5 hidden layers and 100 neurons, and what are the implications of your choice? Why a sigmoid? There are noticeably no references to support your choices, and there's no element of later discussion about the impacts.

We have chosen a MLP because it has a very simple architecture but can deal with non-linear problem, which is the case of this study. The choice of the architecture (number of layer and neurons per layer) as explained in the manuscript as been fixed by testing a large amount of setup (called gridding) and choosing ones that have give the best score on the validation sample, with a reasonable time of learning. Concerning the activation function, the sigmoid was chosen to allow negative FBr as it is for target FBr and not to drop the value and bias the statistics of the predicted values. The sigmoid activation function was chosen so that it could allow negative FBr values in order not to artificially drop the negative predicted FBr values.

Machine Learning is largely used for various application even for geosciences but settings architecture and hyperparameters resides in testing testing and testing to get the best model with the best score. Citing study that use a MLP for geosciences will not be relevant as the hyperparameters highly depends on the issue we want to deal with so it could even be wrong. The paragraph you refer in your comment have been detailed to make it clearer. The implication of all choice is that by choosing the Neural Network type, MLP, with trained the best model possible so to have the best prediction possible comparing to a reference (CS-2 or Envisat calibrated).

The neural network is a multilayer perceptron regressor (MLP) composed of 5 hidden layers, each composed of 100 neurons. The activation function used is a sigmoid. 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. To avoid overfitting, we use early stopping to interrupt the training when the score is not improving anymore. Once the hyper-parameter combination is set, the MLP is trained with the whole dataset. The NN trained is then applied to the LRM monthly grids to obtain a monthly LRM-corrected radar freeboard.
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.

I also have the view that ‘radar freeboard’ is not a geophysical quantity to be measured with an uncertainty. Instead it is precisely the retracked elevation of a waveform returning from sea ice, and is specific to a given radar’s geometry and the chosen retracking algorithm. See the original definition in the supplement of Armitage Ridout 2015, and Tilling et al. 2019 for how different radars will generate different Rfbs even if they could ‘look at’ the same ice. Similarly, different retrackers will generate different Rfbs when ‘looking’ at the same waveform, all of them valid and precise.

So I think you should change the phrase ‘radar freeboard correction’ to ‘radar freeboard calibration’, as you’re not correcting some uncertain value. Instead you’re calibrating the Rfb from one instrument so that it’s consistent with another instrumental geometry. The same with ‘radar freeboard estimation’ - you’re not estimating it: it’s a precise value resulting from the radar geometry and choice of retracker. I have a lot more to add on this issue, but it’s quite philosophical/ subjective and I think we need to first focus on the issue concerning the representation of the TFMRA50 Rfbs in the ‘corrected’ product.

If we understood your comment correctly, we don’t have the same point of view. The geophysical quantity we want to estimate is the freeboard. For that purpose, we measure the height over the floes and the height over the leads that are extrapolated below the floes. The so-called radar freeboard is the difference between both heights. These two heights are subjects to uncertainties (like for all measurements) which are propagated to the difference. Among these uncertainties, we have the speckle noise, the interpretation of the retracking to estimate the range e.g. the retracking step. I can’t find any definition of radar freeboard on Armitage and Ridout 2015’s paper supplement more than ’The radar freeboard is then simply the retrieved elevation of the sea ice floe relative to this interpolated sea level’ but it is also the definition we consider with an uncertain SLA and uncertain sea ice floe elevation anomaly or what we call ILA (Ice Level Anomaly) in reference to SLA.

Note that "error" is used in the manuscript while dealing with speckle noise because Wingham et al 2006 used that terminology, but it refers to uncertainties, both words were often mistaken to qualify uncertainties until a few years ago. This has been clarified in the manuscript.

We do agree that the word ’corrected’ is a bit confusing, as we also deal with uncertainties. Even though, an uncertain value can not be corrected, at least the uncertainty can be reduced contrary to an error that is known and could be corrected. These two words have a different signification.

In order to clarify the reading, we suggest replacing corrected by calibrated while dealing with the predicted radar freeboard from the NN or the surface state bias corrected radar freeboard.

Specific comments - Community comment

L25) I would question whether “thin ice is more sensitive to climatic hazards”. Bitz and Roe (2004) argued the opposite: that thick ice is thinning faster, because areas of thin ice grow more quickly in winter. Age products also show that thicker, older ice is disappearing from the Arctic and being replaced by thinner, seasonal ice (e.g. Nghiem et al., 2007). So I’m not sure it makes sense to say that thin ice is more sensitive to climatic hazards, when thin ice is coming to dominate the Arctic and is more robust to temperature perturbations.

We do agree that because of the global warming, multi-year ice have started to disappear and be replaced (in area) the next winter by FYI. Thin ice thickness will be recovered more easily than thick ice if it melts. Nevetherless, the thinner the ice is, the faster it undergoes melting and breakup when temperature rise in late spring. It will be more supposed to break while occuring climate hazard such as cyclones or strong winds [Rheinlænder et al 2022]. During all seasons, thin ice will ridge, raft, diverge easier than thick ice [Stroeve et al 2018] so can highly affect sea ice area (and of course volume). We suggest the following modification :

Thin ice is indeed more sensitive to climatic hazards than thicker ice but it especially enables to compute the volume.

replaced by:

Thick and old ice is disappearing and being replaced by younger, thin ice that has a higher mechanical sensitivity. Thin ice is more prone to deformation [Stroeve et al 2018] that induce area changes, and is more sensitive to climate hazards such as cyclones or strong winds [Rheinlænder et al 2022]. Thickness is a key parameter for sea ice study, it varies a lot according to the regions and it modulates the sea ice volume evolution in the Arctic ocean [Landy et al 2022].

L32) I don't agree that it's "commonly accepted" that Ku-band radar waves penetrate the snow layer when it is sufficiently cold. I think that assumption is still up for discussion, and I would argue the opposite. I'm not aware of any in-situ or airborne CryoSat evaluation ever done over sea ice that has produced evidence that Ku-band radar waves consistently return from the snow-ice interface. For instance, neither of the airborne CryoVex 2006 and 2008 campaigns (Willatt et al. 2011) indicated that this was consistently the case over FYI. Results from a different radar system in Antarctica (Willatt et al. 2010) also showed that radar waves do not always return from the ice surface. Results from a third radar system deployed on MOSAiC (on SYI) indicate that more Ku- band power comes back from the snow surface than from the ice surface (Stroeve et al., 2020 Fig. 7; Nandan et al., 2022 Fig. 8). Garnier et al. (2022; Figure 9) shows results from CryoVex 2017 where the difference between Ka and Ku band ranging is at times negative, further casting doubt on the assumption. Moving to satellite-based evidence, Armitage and Ridout (2015) calculated CryoSat-2's penetration factor as 82%. Ricker et al. (2015) used buoys to show that snow accumulation caused increases in Rfb, not decreases (implying that the radar waves are not penetrating fully). This agrees with the work of Gregory et al., (2022; Figure 9) that shows that snowfall is correlated (not anti-correlated) with Rfb over both ice types. I would also argue that the often-cited work of Beaven et al. (1995) was not realistic – it featured snow that was shovelled, sifted through a screen, and then artificially smoothed at the surface by the weight of a metal plate before measurement. It is also striking that what the authors identify as the snow-ice interface appears at 20 cm range when it was 21 cm away in free space. Since it was 21 cm away in free space it should have appeared further away, at something like 25 cm in range due to the wave-propagation delay. There's no need to mention all this in your paper, but I wanted to briefly state my evidence before making the point that full Ku-band penetration is not a settled consensus, even for cold, dry snow. I think it would be fair to say that full penetration is "commonly assumed in satellite-based sea ice thickness products". But just because we're forced to assume it in our products doesn't mean the we should actually believe or accept the assumption.

We do agree that the knowledge of how far into the snow layer Ku-band radar waves can penetrate is still under deep discussion in the community. There is no possible consensus on the fact that signal penetration will depend on salinity, temperature, humidity, snow age and other parameters... However, it is important not to confuse the results of studies done in Antarctica with those done in the Arctic, just as it is important not to confuse the SAR results with the results of field studies, since the SAR treatment impacts the waveforms and does not only reflect the behavior of the ku-wave in snow. We suggest the following modification :

It is commonly accepted that the Ku frequency penetrates the snow layer when it is sufficiently cold, in other situations this assumption can be questioned (Ricker et al., 2014; Nandan et al., 2017).

replaced by:

Implementing this method requires the assumption that the Ku-band radar wave completely penetrates the snow layer, which is still widely discussed and is not the subject of a definitive consensus (Ricker et al., 2014; Nandan et al., 2017).

L381: Year of this citation is 1986.

We have taken into account this semantic shade in the manuscript.

L65: I think we're not really measuring sea ice thickness, but instead estimating it based on freeboard measurements (or radar-altimetry measurements). This might seem like a semantic point, but I think users of sea ice thickness products do benefit from this distinction. "estimates" rather than "measurements" is more commonly used by convention (e.g. Tilling et al., 2018, Kurtz et al., 2014; Landy et al., 2017).

Increasing the along-track resolution of the aperture radar has led to considerable advances in the measurement of sea ice thickness.

replaced by:

Increasing the along-track resolution of the aperture radar has led to considerable advances in sea ice thickness estimation.

L75: I think readers like me who aren't expert in roughness would benefit from a citation here. Is LRM definitely more impacted by a given roughness than SARM? I can believe it, but would like to read some evidence.

Kurtz et al 2014 pointed out that ice roughness or more generally, sea ice surface properties impact the waveform of return echoes. Such as a lot of remote sensing instruments, the illuminated area will impact your measurements. Concerning the difference of roughness impact between SARM and LRM range measurement, it's due to the acquisition processing itself. Knowing how both work (see https://www.aviso.altimetry.fr/en/techniques/altimetry.html), the theoretical return power will be the same for both nadir and off-nadir in LRM whereas in SAR most of the return echo power will be concentrated to nadir, which reduces the impact of off-nadir and give the peaky shape to SAR waveforms and a lower impact of surface roughness [Raney et al 1998]. It was a bit more explicated in section 3.4. We propose to add this citation in the sentence you are mentioning and add a reference to the section where it is more explained.

Contrary to SARM, LRM altimetry measurements are strongly impacted by the surface roughness of the surface illuminated by the radar, also affecting the freeboard measurement.

replaced by:

Because LRM altimetry has a larger footprint than SARM altimetry (by a factor 30), LRM range retrieval are significantly more impacted by surfaces roughness of the [Raney et al 1998] than the more nadir-focused measurement (SAR technologies).

L103: I think you mean NSIDC 0611? This product gives the maximum of the ice age distribution in a grid cell at each timestep (see quote below). So I'm not sure how you've used these max values to generate an MYI fraction product? I think it could be done if you had access to the Lagrangian data, which is out there. But if you've used this I think you should state that. (Tschudi et al. (2020) states "This approach does not consider new ice that may form within a grid cell because it retains only the oldest ice in its accounting. Thus, the product is effectively an estimate of the oldest ice in a given grid cell.")

The type is attributed to the 20hz along track measurements from the NSIDC age nested in two categories (whether the age is greater than 1 year). During data gridding, the type is also gridded and gives us an idea of the fraction of MYI by averaging the ice type into cells.

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.

L115: I think at some point you should direct the reader to Kwok and Haas (2015), which discusses some key issues in the product that you've chosen.

This section aims to present the dataset not to discuss it, however, we added the reference to section results as following:

The bias between OIB and Envisat estimation could also be attributed to the OIB snow depth which estimation seems sensitive to the algorithm used (Kwok and Haas, 2015; Kwok et al., 2017).

L310: "Surface roughness is identified as the largest source of uncertainty" - I didn't really understand how you made it to this conclusion. I think this is specifically a reference to Fig. 8 of Landy et al. (2020). The error in the sea ice roughness over FYI is 4cm, and the error from the snow basal salinity (just part of the "penetration bias") is 7 cm, and the uncertainty due to snow depth is 6 cm. So over FYI the roughness uncertainty is smaller than either the snow depth or the snow salinity. As such I don't think roughness can be reasonably characterised as "the largest source of uncertainty" over FYI based on Landy et al. 2020 Fig. 8. Over MYI the sea ice roughness uncertainty is equal to the snow depth uncertainty, and admittedly larger than "partial snow penetration" uncertainty. So the statement is narrowly true if you only consider MYI and don't factor in the (highly related) uncertainty in snow depth in the comparison. But I think that only considering the largest source of uncertainty and ignoring the other uncertainties is a pretty risky strategy, given the other sources are comparable and perhaps actually larger in magnitude? If you are wedded to this approach, I think you should state that this will induce a pretty serious underestimate in your uncertainty values (which is important info for product end-users).

This sentence is inexact, it has to be shaded to "surface roughness is identified as one of the largest sources of uncertainty". Nevertheless, the other sources of uncertainty while measuring the FBr, as summed up in Landy et al 2020, is the uncertainty due to SLA (off nadir and low density) and the limited Ku-band penetration in the snowpack (caused for instance by snow basal salinity for FYI or metamorphic snow for MYI) not the snow depth. The point that was not explained in the manuscript, incorrectly, is that the uncertainties due to partial signal penetration in the snow are only indirectly taken into account, we don't ignore it. Indeed, it is not so trivial when comparing freeboards from different retrackers, to differentiate between roughness and penetration [Ricker et al 2014]. We believe that a "significant" part of the uncertainty on penetration is included in the uncertainty on roughness presented in [Landy et al 2020]. For this reason, we made the choice not to add values for the undefined limited penetration of the signal in the snowpack in this uncertainty budget. The problem of penetration is not ignored, but the manuscript lack of information on this point. As contribution of sources are not defined, yes, it is possible that the final uncertainties are underestimated. We suggest the following modifications:

Landy et al. (2020) decomposed it in two, the FBr systematic uncertainty budget, on the one hand, the uncertainties due to the penetration of the signal in the snow (depending on its salinity or if it is composed of metamorphic snow, according to the type of ice) and 310 in the other hand, the surface roughness. Surface roughness is identified as the largest source of uncertainty and we, therefore, choose to consider only this source in our systematic uncertainty evaluation. Roughness is estimated to be respectively about 20 % and 30% of the sea ice thickness for FYI and MYI (Landy et al., 2020). Note that this systematic uncertainty budget only concern CS-2 mission which are afterward propagated to Envisat and ERS-2, indeed other mission will be "corrected" from surface roughness effect during the calibration procedure.

replaced by :

In Landy et al 2020, the FBr systematic uncertainty budget is decomposed in two parts, on the one hand, the uncertainties due to the penetration of the signal in the snow (depending on its salinity or if it is composed of metamorphic snow, according to the type of ice) and in the other hand, the surface roughness. We assume, as in Ricker et al 2014, that the comparison of the freeboard from different retrackers does not enable to separate the contribution of the roughness from the signal partial penetration. We therefore assume to consider both sources as one mixed contribution, estimated to be respectively about 20 % and 30% of the sea ice thickness for FYI and MYI (Landy et al 2020). The systematic uncertainties can be underestimated as the penetration of the radar waves in the snow uncertainty may be poorly handled. Note that this systematic uncertainty budget only concerns CS-2 mission which is afterward propagated to Envisat and ERS-2, indeed other missions will be "calibrated" from surface roughness effect during the calibration procedure.

Fig. 6: I see in the top panel that you've "summed the squares", which has the implicit assumption that uncertainties that you have considered are uncorrelated. It may be that you have good evidence to support this that I'm ignorant of, but it seems, for instance, that speckle noise may well be (anti?)correlated with surface roughness? Just as an example. I think that the omitted snow uncertainties involving penetration & depth are more likely than not to be correlated in some way. I think you should state that you've assumed the uncertainties are uncorrelated in your analysis, and give the reader some information as to what the results of that assumption may be.

Yes, we assumed the uncertainty due to the speckle noise and the SLA uncertainty to be uncorrelated, as it is precised l 304 in the manuscript. The speckle noise will not be correlated to the surface roughness but it is attributed to the surface asperities which are of the order of magnitude of the wavelength of the signal, about 2 cm, that causes interference in the signal. But a surface can have several roughness scales, for instance MYI highly rough can have asperities of about 2cm as well as newly formed sea ice, both will present speckle noise that induce the same uncertainty on the range.

By construction, systematic and random uncertainties are not correlated, so this is not really an assumption. Concerting the uncertainties due to snow penetration, we redirect you to the previous comment (and here the snow depth is not considered, so its uncertainty either).

Figs. 7 & 8: These are really well designed and presented

Thank you for this comment.

L381: 4) Why take snow density as constant? SnowModel-LG outputs depth and density, and includes some physics of densification/settling over time. So I think it's odd to use one of its variables and not the other, since they're so linked in the model. Snow impacts thickness retrievals by weighing the floe down and slowing radar waves: both of these effects are proportional to the mass of overlying snow – not the depth (see Mallett et al., 2021). So I think it makes a lot more sense to use both the depth and density (the SWE) in your thickness retrievals rather than just the depth. Here's a plot of the seasonal densification of SnowModel-LG snow north of 88N for the period 1995-2018. You'll see that as well as being more dense than your assumption, it also evolves over the season.

As suggested, we have done the scatter plot using the snow density from the model, figures have been updated in the manuscript. Nevertheless, it was not possible for CanCoast due to SnowModel-LG output coverage, and as we take OIB snow depth for OIB/Envisat SIT conversion, we have chosen to keep a constant density to keep consistency. It is interesting to see that the comparisons look really similar than with a constant snow density. Comparisons with moorings even give worse results with higher biases than using constant density. You can find the statistics in figure 4, 5, 6 and 7.

Fig 13: I'm a little unclear what the radar freeboard timeseries is supposed to represent. I imagine it mostly reflects the trend and variability in sea ice extent, and I think you should point this out to the reader. A simple correlation with SIE would quantify this relationship and reveal if the quantity is useful. For the part of it that doesn't represent SIE, would decreasing volume reflect a thinning of sea ice? Thinning snow (Webster et al., 2014) will mask the effect of thinning ice on the Rfb. In areas where the snow is really thinning quickly, the Rfb could potentially even increase even if the ice is thinning. I guess I would like to see a little interpretation of this quantity figure 13 rather than being left to do it as the reader.

The time series present the radar freeboard volume we have computed during the ERS-2, Envisat and CryoSat-2 period. We chose to show the evolution of the volume of radar freeboard instead of mean radar freeboard as it is more difficult to interpret as the mean freeboard depend on the number of pixel covered by ice and is not necessary representative of the global ice state because low concentration area have the same weight as compact ice area. The motivation of computing the volume is to represent better this evolution. according to [Landy et al 2022], figure 8, the anomaly in volume are mainly driven by thickness anomalies and not area for the Arctic, so this would not reflect the variability of sea ice area. Concerning the impact of snow on trend, you are right, it would change the trend as snow load have changed during the past 30 years, but it will give the same trend as using a snow depth climatology to derive the volume as it has been done in the majority of the previous studies. We suggest adding the following precision:

The evolution of the snow load is not taken into account in Figure 13, which means that the evolution of the volume is not fully represented, in the same way as if the total volume were derived with a snow depth climatology. Indeed, a decrease in FBr volume may merely indicate that the snow depth is greater and the ice thickness unchanged.

L450: I think you should state the limitations in your uncertainties here. In particular (and I think this is key), do the "observed" thicknesses fall within your uncertainty bounds? If not, then either your uncertainty bounds are wrong or the validation data is wrong. I think uncertainty bounds on retrievals are not useful unless you can show that observed data fall within them.

Thank you for this suggestion, uncertainties of satellite estimates have been added to validation plots for the review. Nevertheless, as explained in the manuscript section results, conclusion are not as simple to draw out, validation or retrieval are not necessarily wrong if validation dataset don't fall within uncertainties bounds. First, because validation data also present uncertainties but also because validation procedure of monthly satellite estimation assume that we observe in average the same sea ice surface than the validation do and that can be questioned for airborne or submarines dataset for instance.

Figures 4,5,6 and 7 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. For esthetical reason, bounds are not represented for comparisons with other satellite-based SIT estimation.

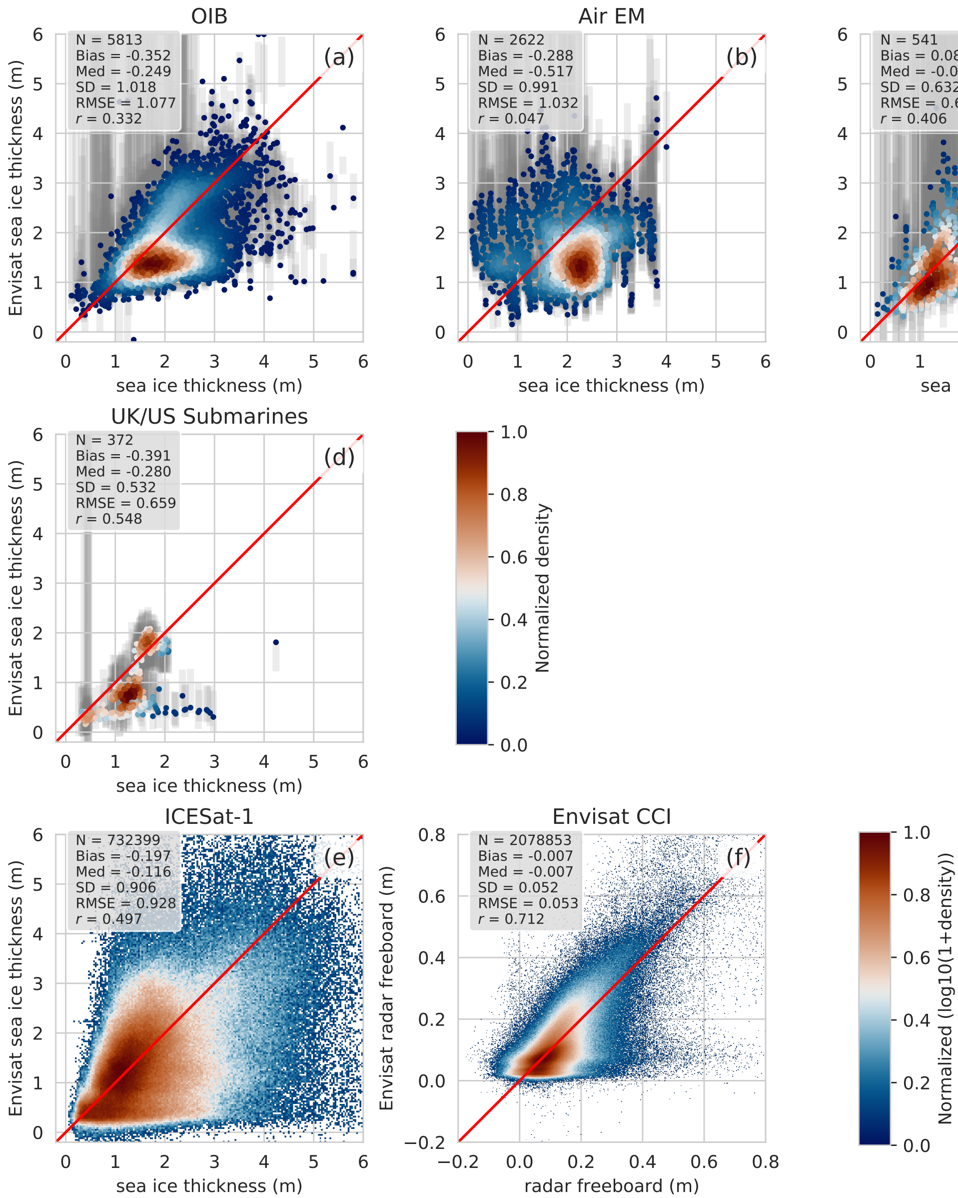

Figure 4. Comparative scatter-plots between Envisat sea ice thickness or radar freeboard estimations and other data sets. The x-axis indicates the sea ice thickness from (a) OIB total ice freeboard, (b) Air EM snow plus ice thickness, (c) Can Coast ice thickness, (d) UK/US submarines draft and (e) ICESat-1 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.

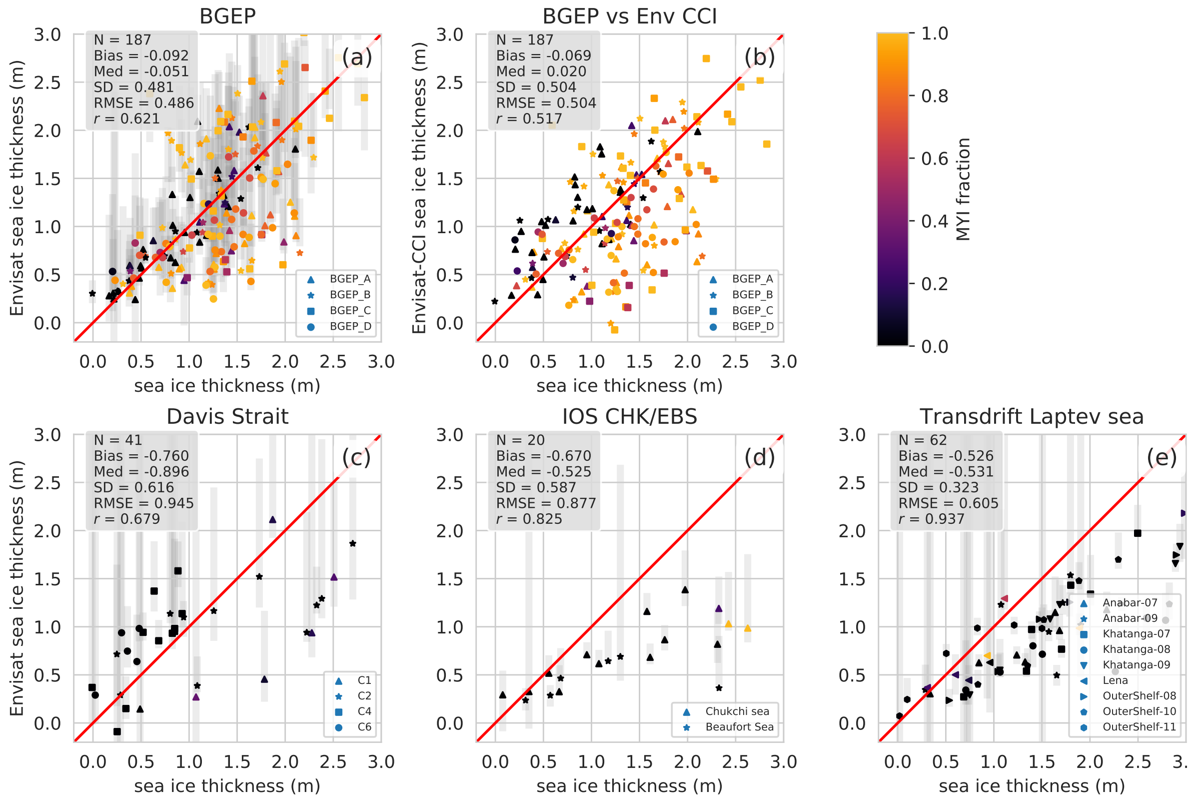

Figure 5. Comparative scatter-plots between Envisat sea ice thickness estimations and anchored moorings data sets. Each dot corresponds to a monthly averaged value. The x-axis indicates the sea ice thickness from (a) BGEP, (b) BGEP vs Env CCI, (c) Davis Strait, (d) IOS CHK/EBS and (e) Transdrift Laptev 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.

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.

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.

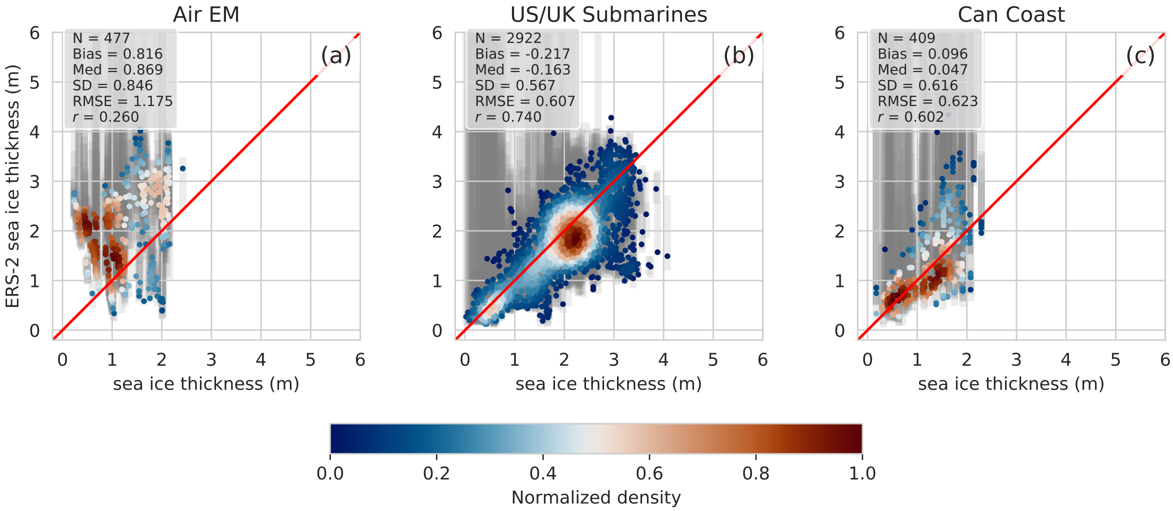

Figure 6. Comparative scatter-plots between ERS-2 sea ice thickness estimations and 3 in-situ data sets. The x-axis indicates the sea ice thickness from (a) AirEM total thickness, (b) UK/US Submarines draft and (c) Can Coast sea ice thickness. Colorbar indicates the normalized density. N is the number 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.

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.

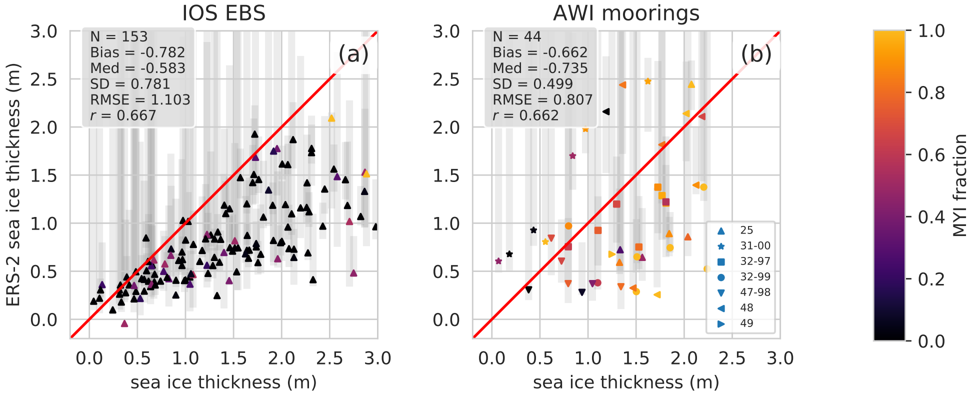

Figure 7. Comparative scatter-plots between ERS-2 sea ice thickness estimations and 2 anchored moorings data sets. The x-axis shows sea ice thickness estimations from (a) IOS Beaufort Sea and (b) AWI moorings sea ice draft. The color bar indicates the respective MYI fraction. N is the number of the couple 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.

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.