Response to referee comments – Marion Bocquet

We thank the referee for their useful comments. We appreciate the time and effort dedicated to providing feedback on our manuscript and are grateful for the comments. We believe we have been able to address each of them.

General comments:

1. The first one, relatively minor, concerning the representativity of the dataset, to what extent are the data used to train the NN over the period with CS-2 representative of the whole area you are trying to estimate, in other words, is there a bias in the sampling of the dataset that can lead to a bias in the predicted SIT. For instance, is the backscatter of CS-2 'area' is representative of CS-2+channels area ? A simple plot of the backscatter distribution in the region used for training and the global region would give a good idea, same for the parameters used from ice charts that allow to train the NN (floe width etc)

Removing the CAA channels from the training dataset removes around 1/3 of the datapoints (depending on the month). Below is a plot (Fig 1) of the distribution of backscatter from C-band and Ku-band in November and April for the training region and the full region. As you can see the distribution is very similar.

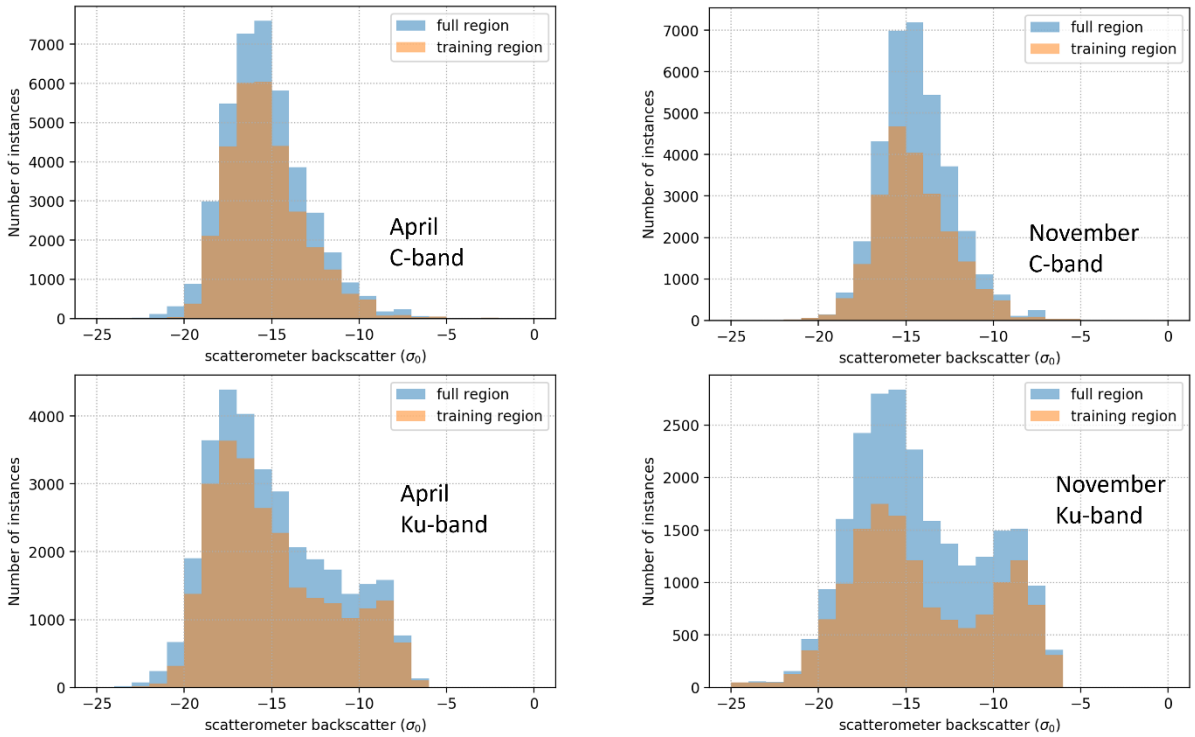


Fig 1. Distribution of scatterometer backscatter

The same is true for the features used from the ice charts. Figure 2 shows an overview of some of them for April. The one feature where we would expect the difference between the full region and the training region to be strongest is for fast ice, as this is more common in the channels in the CAA, and less so in the open seas. However, even for this feature there is a full spread from 0 to 1 in the training region. The fact that the distribution of features in the training region is representative for the full region has been made clear with a few added sentences to the text.

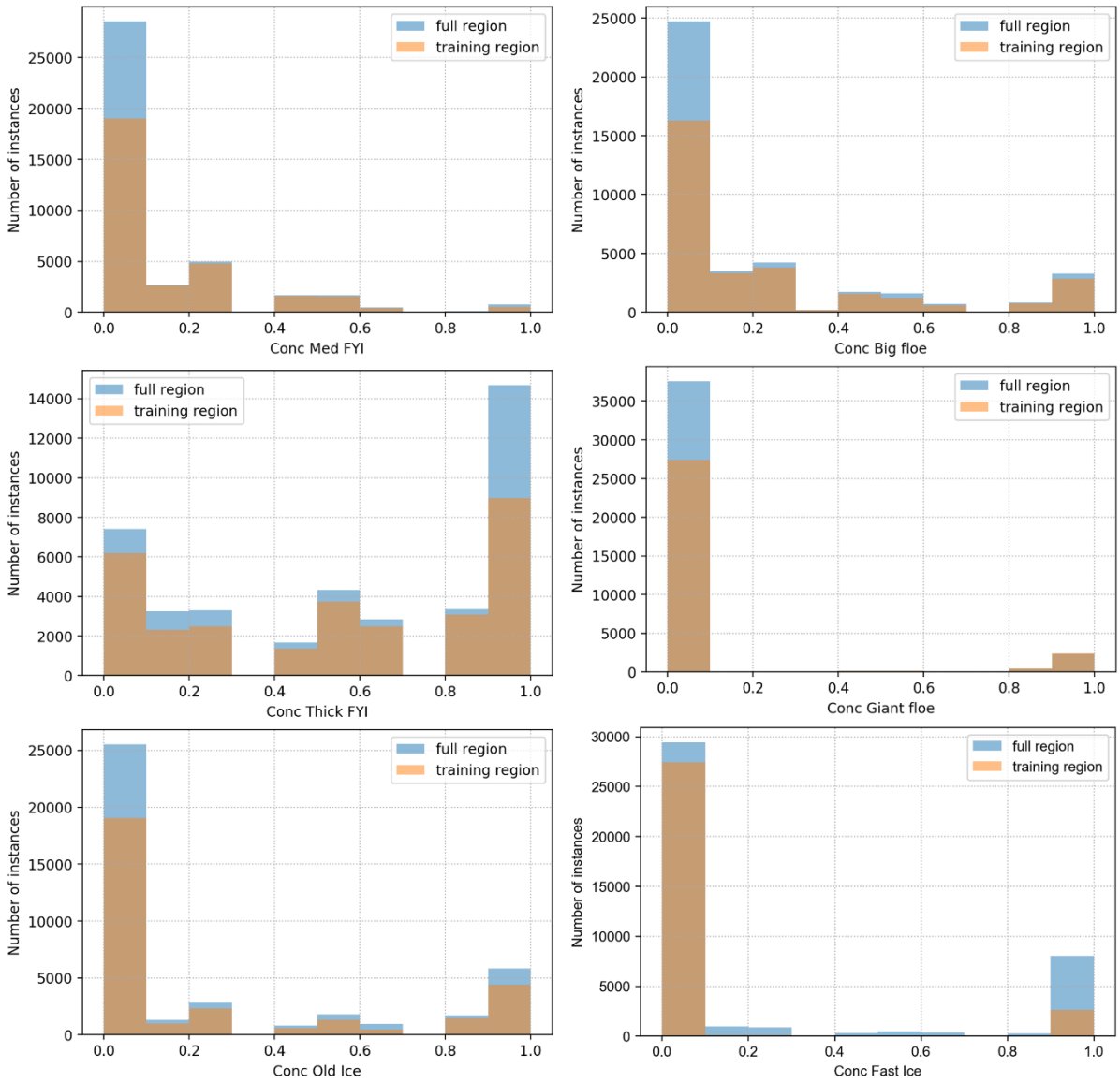


Fig 2. Distribution in training region and full region of key features used from the ice charts in April.

2. The second one concerns the data « trending » that is made for MYI thinning correction. To what extent PIOMAS trend for MYI SIT is supposed to represent the real trend especially before 2010 as we usually see higher discrepancies while going back in time with PIOMAS ? This can maybe be discussed a bit more in the conclusion or in a discussion section ? Especially that, as explained by the authors, PIOMAS thick ice (MYI ?) is underestimated, linearly for the past 25 years ? How confident are the authors by « retrending » using PIOMAS ? I was also wondering it which extent does it make sens to trend the data and then to explain changes in the trend with the data as we could directly study PIOMAS dataset to get these trends ? I am a bit confused with the corrected and not corrected trends products. Why the authors detail trends and explain changes for the not corrected sea ice thickness product, as it is supposed not to reflect the trends ? Especially regarding the differences between not corrected and corrected trends for all regions and all months. Maybe in a way the study of the trends with the corrected dataset will be more relevant if it would have been validated before as the SIT changed, but I am not sure that is the purpose of this study. What bothers me is probably L 170, and

that SIT are consistent but not to estimate trends… so the dataset is not so consistent as correcting the SIT by the trend will change the SIT values. Same for fig 11 and 12, the not corrected trend values are shown but they are supposed not to be so relevant. Could you make Fig 11/12 (maps) with the trends for corrected SIT ? Are the spatial patterns similar ? (Not necessarily in the manuscript for now.)

We do agree with the reviewer that the way the two products were presented was confusing. We have followed the reviewers advise to do the validation on the corrected SIT proxy dataset, and we found that this gave better results than the validation on the not-corrected proxy SIT gave. This gives us more trust in our decision to apply this correction, despite the uncertainties that PIOMAS can bring. We have included a more thorough discussion of the uncertainties in PIOMAS, how the correction is exactly applied, and what the difference between the corrected and not-corrected product is and which one to use in which situation.

We have changed the maps of Figure 11 and 12 in the revised manuscript to show the SIT trends from the corrected SIT proxy. The spatial patterns are the same, the magnitude of the trend is stronger, and now negative almost everywhere.

Specific comments

L2/3: You should be mentioned in the abstract that the estimation is based on Ice chart/scatterometers as it is one of the main elements of the thickness estimation.

This has been added to a revised version of the manuscript.

L6: 'mean trend' is a bit confusing. If I am not mistaking, this is not the mean trend this is the trend for the non corrected product for all the period and the whole area.

There will be clarification added to the new manuscript that this mean represents the whole area.

L128: why to chose the mean of the two product as you have seen that ku seems to be more more uncertain (higher RMSE)?

Neither the Ku-band nor C-band scatterometer data are available over the whole time period. Because of this reason we cannot choose one of the two. Another possibility would have been to select C-band where available and fill in the gaps using Ku-band. However, as the two have different wavelengths, penetrate the snow and ice to a different depth and show different properties of the sea ice, we think this would have led to small inconsistencies. To remove some of this, we decided to take the mean where possible, and when only one of the two was available, select that one. Moreover, the different scatterometer sensor wavelengths have their own strengths and weaknesses, by combining both we get the full spread of their capabilities.

L176: Just to be sure, are you trending the SIT as following :

$$SIT_{corr_m} = SIT \cdot t \cdot \left(trend_{MYI_m} \cdot C_{MYI} + trend_{FYI_m} \cdot C_{FYI}\left(+trend_{YI_m} \cdot C_{YI}\right)\right)$$

With SIT the sea ice thickness, m the month you are correcting and C the partial concentration of each category. (Young ice trending is between parentheses as you not correct YI thickness) In other

words do you make a weight average of the the SIT with the values of the trends and the partial concentration?

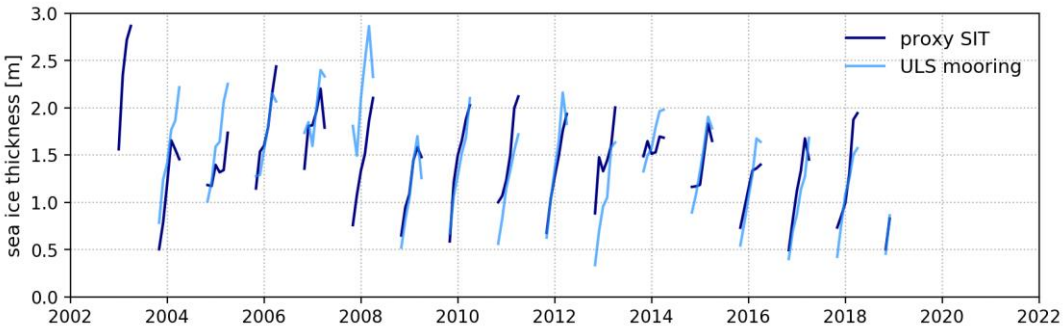
> Yes, this is a much clearer way of stating this function, thank you. We will be using this phrasing of the correction function in the revised manuscript.

L 182-184: I would also put this two lines within the conclusion maybe juste clarifying L 375 as you suggest that the proxy SIT need to be trended to take into account MYI thinning.

> We have clarified this in the revised manuscript.

Figure 7: I'm not sure that this figure is so convincing and enhance your dataset. The new series is consistent with CS2 which is very good. Altimetry products show big discrepancies with BGEP for the winters 2006/2007 and 2007/2008, so it doesn't surprise me that there are quite big differences for these years in this product too. Nevertheless it might be more meaningful to show a time series (still with CS-2), we would see that the seasonality is well represented too for other winter.

> We have created a time series of the proxy SIT and the ULS moorings, included here:



> We do think this figure shows that the seasonal cycle is captured. However, we think this is already shown by figure 9 in the manuscript (fast ice thickness comparison). It is good to see that the seasonal cycle is captured by the proxy SIT product, however, we think it is also important to see if the interannual variability is captured by the proxy product. We aim to show this in Figure 7. We agree that there are years where the comparison with the ULS BGEP moorings shows large differences, but think these differences are explainable and that this figure is of added value to the manuscript. We therefore suggest to not change this figure in the main text. We will add the figure above to the Supplementary Materials.

Section 4.3: Maybe it would be even more readable at some point to put this section as a part of the discussion section as it both discusses the time series and the method and not only the method. But this is more a detail.

> We have chosen to not follow the results-discussion set up but instead have a chapter that presents and discusses the 'Model performance' and a chapter that presents and discusses the 'sea ice thickness proxy-product', so the outcome of the model. We do believe that section 4.3 (Limitations and potential), only discusses the limitations and potentials of the model (and not the proxy SIT record itself) and should thus belong in the chapter on model performance.

L 223: I may have missed this information, but I didn't understood how you estimated the 30-50 cm uncertainties of the product.

The 30-50 cm is the testing error of the model as presented in Figure 4. I do agree that this may not be exactly the same thing as the uncertainty of the product, as the testing error treats CryoSat-2 SIT as 'correct', and thus only shows the error in the model compared to CryoSat-2 SIT and not actual SIT. We have clarified in the revised manuscript that this is the testing uncertainty of the model.

L 310-311: RSE characterized how the prediction fits the reference SIT (CS-2). Values are relative to CS-2 SIT not to the trend. Maybe I didn't get the point, why the variability between SIT proxy uncorrected and CS-2 SIT represent the variability to the trend, CS-2 SIT also get a variability to the trend isn't it?

The residual standard error (RSE) does not look at how the predictions fit the CS-2 SIT. As stated in line 311, it characterises the standard deviation on the residuals of the regression model on the proxy SIT product. It quantifies how far the data points are scattered around the fitted regression line, and thus the variability from the trend. This is done according to:

$$RSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$$

Where y_i is the observed value of mean SIT in the proxy SIT product for given year i, $\hat{y}_i$ is the expected value in the fitted linear regression model for the same year, and n is the number of observations (years). This is different from the root-mean-square-error (RMSE) used to compare two datasets (like the proxy SIT product and the CS-2 observed SIT), used in other parts of the manuscript, which might have been the source of the reviewer's confusion. We have included this formula in the revised manuscript to clarify this.

L 369: « mean trend »? This is still a bit confusing, I suppose it is the trend for the SIT for the whole studied area not the mean trend of each region.

We have clarified that this is the mean trend over the full area in a revised version of the manuscript.

L 375: Which one? Corrected or not, the not corrected will not provide consistent trends isn't it?

The trends and variability in SIT can be studied by either of the products, the choice which one needs to be made on the basis of what someone wants to study. The not-corrected product can be used to study SIT trends caused by changes in ice type. However, this product does not include changes in SIT caused by the thinning of a given ice type, like MYI, which the corrected product does. We have clarified this in the revised manuscript.

Which kind of regression are you using to compute trends?

We are assuming linear trends and thus using a linear regression. This was clarified in a revised version of the manuscript.