Response to referee comments – Robert Ricker

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.

Major comments:

1) I find there is a lack of information regarding the data and methods, especially scatterometer data, but also ice charts, partly "hidden" in the supplements. For example, I recommend being more specific on the used ASCAT data. Which data product have you used? The sigma-0 at 40 deg incidence? Moreover, information like the workflow diagram is key to understand the study and should be included in the paper. In general, I think information, crucial for the paper should be present in the actual paper, while Supplements only support the paper, e.g. additional plots that show something in more detail.

> We agree with the reviewer that having this information in the Supplementary Materials is not as accessible as it should be. Therefore, we have added the workflow diagram into the main paper. However, we also think that having the full technical details on the scatterometer data is not critical to support the conclusion of the paper and it would disrupt the flow of description. Therefore, we think it's best to move the description of the scatterometer data and its consistency across missions to the appendices (which are printed in the same document as the manuscript).

2) I wonder if the output can be improved by using different training data. The Beaufort Sea is known for being very challenging for satellite altimetry, because of the mixed ice types, low SIT correlation lengths, and sometimes high drift speeds. I suppose that this leads to some misfit between altimetry, scatterometry, and ice chart data. What about using only areas, where confidence in CS2 SIT is better?

> The locations where training of the model is possible relies on the availability of the CIS ice charts, which are available in the Beaufort Sea, Baffin Bay, and the channels of the CAA. We believe the reliability of the CS2 dataset in the CAA is low, mainly owing to the lack of leads, and thus don't use this region in the training. If we were to remove the Beaufort Sea from the training dataset, we would only be left with Baffin Bay. This would give a rather small region to train on, with a mix of ice types lacking in diversity and therefore unrepresentative of the larger Canadian Arctic region we apply the model to. There is very little MYI ice in this region for example, so the training dataset would not have the same variability in the model features as the full region would have.
> We are aware of the fact that including the Beaufort Sea in the training dataset brings in uncertainties, but think that the addition of this region with a mix of ice types, sea ice thicknesses, and scatterometer results, will make the model better as it will learn more from the diverse inputs. We also think having the low spatial resolution we have (50 km), might decrease some of the uncertainty caused by high sea ice drift.
> We have added a discussion of these uncertainties in section 4.3.

There are also CS2 SIT products that provide data in the Canadian Archipelago, for example the AWI product. It would be interesting to compare the proxy product with such estimates, too.

> We have made a comparison to the AWI CS2 SIT product and included this in the revised supplementary materials and referred to it in the revised manuscript.

3) I would expect a certain bias between scatterometry data sets from different sensors (but same frequency bands). For example, between ASCAT and ERS, because the sensors are different (as pointed out in the supplements). Can you rule out a bias that might affect derived trends in the proxy product? Or should it be at leased discussed in the limitations section?

> We have looked at whether there was a bias when combining different scatterometers, including OSCAT-1 and OSCAT-2, OSCAT-1 and QuickScat, and ERS-2 and ASCAT. This was included in the Supplementary Materials but is now moved to Appendix A. For ERS-2 and ASCAT there is unfortunately no overlap in time, instead we looked at the PDFs of backscatter for both satellites for different ice ages. We see a similar backscatter in ERS-2 and ASCAT data for the same ice age and for this reason conclude that we did not need to apply a correction before merging the products in a C-band scatterometer product. We have now included a few more lines on this in the main text as well.

4) A major concern is the correction to thinning of ice types (3.4). I understand the problem, but PIOMAS also comes with considerable uncertainties in the study regions. And it is a model, too. I wonder if it makes sense to correct the proxy product (based on a ML model) with trends from a sea ice model, while one could then also just take PIOMAS to look at trends. Figure 11 actually shows (except Baffin Bay may be) that the trend in the corrected product is basically coming from PIOMAS? I think it is feasible to compare with PIOMAS trends to discuss the problem. But I am not sure if this can be sold it as a separate product. I would argue that it is less confusing for potential users if there is just one product, where limitations and uncertainties are clear.

> Following a comment from the other referee, we have now applied the validation to the independent SIT datasets on the corrected proxy SIT product as well (we did it on the non-corrected product before). We find that the correlation (for most of the locations and datasets) and anomaly correlation coefficient (for all datasets and locations) is higher after the ice type-thickness trend correction is applied. This makes us believe that despite the uncertainties present in the PIOMAS product, adding the correction is better than not adding the correction. We therefore choose to keep the ice type-sea ice thickness correction in the manuscript. We do agree that presenting the two different products was confusing, and therefore now focus on just the corrected version, as we've found this gave better results in the validation. We have included a more thorough discussion of the uncertainties of this correction in section 4.3.
>
> We don't think the adding of the ice type-SIT correction using PIOMAS results in the trends being so similar to PIOMAS that you could just take PIOMAS trends instead. The ice charts and scatterometer data used by the proxy SIT product contribute significantly to the trends and spatial variability. For example, in April over the full study area, PIOMAS gives an overall trend of -0.5 cm/yr for 1996-2020, whereas the corrected proxy product results in a trend of -1.5 cm/yr. Moreover, the spatial variability in this trend is large, with PIOMAS showing stronger trends than the proxy SIT in the Beaufort Sea and Parry Channel, and weaker trends compared to the proxy in Baffin Bay and the Arctic Ocean Periphery.

5) The uncertainty estimate of 30-50 cm is based on the estimated model uncertainty, verified by the ULS comparison, if I understand correct. How do the OIB data compare to the proxy product in numbers? Section 4.2.2 is rather descriptive. It would be good if some numbers to verify the uncertainties can be presented here as well, like RMSD values etc.

> We agree that this section is mostly based on descriptions and the visuals. We have now determined the correlation and RMSE between the OIB tracks and coinciding proxy SIT

values, and included these numbers in Figure 8. These numbers have also been added to section 4.2.2.

Specific comments:

L81: I think the justification should be formulated here.

We have included a few more lines on this justification in the main text.

L122: Perhaps mention that scikit-learn is a python library.

This will be added to the revised manuscript.

L123: Any reasoning why 95?, and the value for the maximum depth? Is it an empirical choice, after trying different setups?

These parameters were selected using the hyperparameter tuning function GridSearchCV in the python package scikit-learn. This function optimizes the parameters by using a cross-validated grid-search over a parameter grid.

L136: May be consider introducing sub-sections for each validation data set.

This was added to the revised manuscript.

L147: Is it the same ice density as used for the CS2 SIT product? Why do you not distinguish between FYI and MYI density, see Alexandrov et al. (2010), Jutila et al. (2021)?

Thank you for pointing this out. We have changed this to make the sea ice density different for FYI and MYI according to Alexandrov et al. (2010), in line with the CS2 SIT product.

L149: I think it should be clarified that OIB does not measure SIT but uses the ATM laser and the snow radar to measure snow freeboard and snow depth and convert freeboard into thickness. This conversion goes along with several uncertainties as well. So, I suggest to rather write that OIB provides SIT "retrievals" or something similar.

We have changed the wording to say SIT retrievals.

L178: I suggest shortening the lower case terms here to improve readability of this formula. And what is the difference between the "i" and "icecategory"? Please clarify.

We have changed the formulation of this formula to make it more intuitive.

L191: May be shortly explain the "10-fold cross validation RMSE", what does it mean? And is the testing error directly related to the uncertainty given in the Abstract (30 to 50 cm)?

An explanation of the cross validation was added to the revised manuscript. Yes, the testing error was directly related to the uncertainty in the abstract. We have now clarified that this is the testing error, not the full uncertainty.

L199: How have you chosen the 20% CS2 data? Are they randomly picked points? Or did you cut out a certain area? This distinction can be important as the correlation between both chunks of CS2 data might be different. Please clarify in the text.

Yes, these points were randomly selected. We will clarify this in the revised manuscript.

L217: The usage of "proxy product" and "not corrected proxy product" is sometimes confusing. May be use a more consistent nomenclature, e.g. proxy_corr product and proxy_nocorr (or only proxy) product, throughout the paper.

We have made this more consistent.

Figure 2: I do not understand Fig. 2b) - what does "form of ice" mean here? And why does it have "km" as unit? Shall it relate to floe size? Or shall it show different forms of ice more in the sense of ice type, but then the colormap should be discrete?

The form of ice comes from the ice chart's WMO egg code, and relates to floe size or kind of ice (e.g., iceberg, fast ice). The forms of ice are defined using a minimum width, described in Table 'Coding for Forms of Ice' on https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations/publications/interpreting-charts/chapter-1.html. What we did for this figure was to take these widths for each of the present forms of ice in a polygon, and weight them according to their relative partial concentrations, into a final number that described the weighted mean for present form of ice. This number is a bit subjective and not used anywhere else, only to depict the spatial variability in forms of ice. We do agree that this is confusing, and have changed Figure 2b to state forms of ice – small to large, so it is more consistent with Figure 2a.

Figure 8: The histograms are very difficult to separate. Maybe you can have the columns next to each other (like in the supplements) and/or use colors that are easier to separate.

This was changed in the revised manuscript to having the columns next to each other.

Figure 9: It is very difficult to compare the in-situ values, as the time axis is very coarse, while the interannual gradient is quite strong.

We have added a subfigure for each location showing the mean seasonal cycle; the subfigure is wide enough to compare the two.

Figure 11/12: For the line plots, it would help to have legends. Why are you using colors for the different regions if you present them in different boxes? I am also slightly confused with the "solid and weaker circles". "Colours show the trend for the not corrected version" -> But they are all colored? This figure not so easy to read.

We have removed the different colours and made all the line plots black. We have also added a legend to the first plot and updated the caption. We hope this makes the figure easier to read.

Figure S1: I suggest including this figure in the paper, as it shows the workflow to derive the proxy product - too crucial for the Supplements from my point of view.

This figure is now included in a revised version of the manuscript.