Response to reviewer #2

The authors thank reviewer #2 for their comments.

The review focuses on the suitability of the CryoSat-2 SIT observations for assimilation, rather than on the new methods for assimilation of this data described in the paper. Questioning the quality of the assimilated data is certainly a sensible point, but is one we are able to counter as follows. (Reviewer comments are shown below in green).

A good data assimilation scheme takes into account errors and uncertainties in the observations. Any quality of data can be assimilated, as long as the observation uncertainty is properly accounted for. This allows the analysis to give more or less weight to the observations as required. An in-depth discussion of the observation uncertainty and potential areas for improvement is already included in the paper.

*"Thus, how are the authors confident that assimilation of Cryosat-2 observations does actually bring modeled SIT closer to the reality? Particularly, given the fact that degradation is observed when compared against ULS and EM induction data. There is some improvement with respect to Ice Bridge data, but the amount of data is limited."* *and* *"My main concern is that the paper does not contain a clear evidence that assimilation of CPOM Cryosat-2 retrievals helps to bring modeled SIT closer to the reality."* Assessment against the independent OIB (Operation IceBridge) observations clearly demonstrates that the assimilation leads to improvements in the model output. The reviewer bases their conclusions on the Air-EM (45 matchups) and ULS (BGEP; 54 matchups for assimilated data above 1 m thickness) validation, and dismisses the conclusions from the OIB validation on the grounds that the number of matchups is limited, despite there being far more data (547 matchups). It is therefore difficult to see the reviewer's argument here.

As noted by the reviewer, the results for Air-EM matchups are worse than for OIB, but various reasons are given in the paper including that actually these, and not the OIB matchups, are limited in number (45 matchups vs 547). The ULS (BGEP) results for thicknesses above 1 m (data below 1 m are given little weight in the assimilation) are fairly similar with or without the assimilation since the validation against the CryoSat-2 data and the model are similar at this location, as discussed in the paper. Both the BGEP and Air-EM matchups have additional sources of uncertainty: conversion from draft to thickness (BGEP) and the addition of snow (Air-EM). Further discussion on this is given in the paper.

Furthermore, there are several published papers which demonstrate the good performance of the assimilated CryoSat-2 data against independent observations, e.g. Tilling et al., 2015 (already cited in the paper, but I will explicitly add this to section 5).

The reviewer suggests that we work on improving the CryoSat-2 retrievals. However, this is well beyond the scope of this paper, which is on the topic of assimilating the data that is presently available. Since there is an increasing body of peer-reviewed literature on the assimilation of CryoSat-2 observations, e.g. Yang et al. (2014); Mu et al. (2018); Xie et al. (2018); Liang et al. (2020), including in this journal, the grounds for rejecting our manuscript for doing the same would be highly questionable.

Measurement uncertainties can be accounted for in the observation uncertainty estimate used in the assimilation, and so do not necessarily need to be addressed at the retrieval stage. It is not possible currently to account for particular uncertainties in the simple thickness-based observation uncertainty estimate generated for use in this study, but it should be possible to include them in an observation uncertainty estimate generated as part of the data processing chain. Indeed, this paper provides the evidence and motivation for data producers to work on this, with the knowledge that there is a direct application for it, and users who are ready and waiting.

The Landy et al. (2020) reference for surface roughness suggested by the reviewer was already cited in the paper as part of the discussion on random uncertainty in the retrievals. However, on reflection, this is actually a systematic and not a random uncertainty (can't be removed by averaging), so actually comes under the measurement uncertainty (section 2.4.1). Similarly, the Nandan et al. (2017) work on ice salinity would be covered by this too. We will add more information to explicitly list examples of the sources of error that are covered by the measurement uncertainty.

The effect of sea ice type is already accounted for at the retrieval stage (e.g. Tilling et al., 2018).

The stated aim of this paper is to demonstrate that the assimilation of along-track CryoSat-2 observations is feasible, i.e. a "proof of concept". The further work that needs to be done before the assimilation can be included in an operational system is fully acknowledged in the paper. An in-depth discussion of the weaknesses in the current observation uncertainty estimate is already included in the manuscript, but this does not mean that the dataset has no value for assimilation at all. It is clear that the modelled sea ice is much too thin; the assimilated data therefore does not have to be of high quality to contribute any improvement. Additionally, there is certainly the expectation that, like all the satellite observations currently assimilated in FOAM, CryoSat-2 freeboard retrievals will continue to undergo further improvements and corrections by the data producers. It is not necessary to wait until the data is of extremely high quality before attempting to assimilate it.

We have therefore shown that, in forming their conclusion that assimilating winter SIT observations in sea ice numerical systems is not useful, the reviewer has not duly considered the impact of the data assimilation system on the analysis in addition to the observations. They have also not given proper consideration to the discussion already included in the paper on observation uncertainties. They have additionally discounted the strong evidence of improvement in the model performance validated using independent OIB observations, in favour of datasets with far fewer matchups, and have not taken into account the in-depth discussion in the paper of why these latter results should be considered less reliable. They have additionally not considered previously published validation of the CPOM CryoSat-2 observations.

It remains unclear why, even if the SIT improvements due to the assimilation were marginal (or even demonstrated a clear detriment), this would be grounds for rejection of the paper as that result would still be useful to know. Presumably the methodology and assessment in this study are sound, since the reviewer has made no comment on those. No suggestions have been offered on how to improve the assimilation, other than to upgrade the dataset prior to use, which is well beyond the scope of the project.

Addressing specific points in the review:

The reviewer states *"The large improvement in SIT analysis is observed when compared against Cryosat-2 biased data themselves (although not yet assimilated), and I think this improvement simply comes from the memory of the previously assimilated Cryosat-2 SIT observations. But such a result is expected regardless the quality of the assimilated Cryosat-2 observations."* This is correct. Assessment of observation-minus-background statistics is standard in data assimilation, and this demonstrates that the assimilation is working as expected. The paper does not claim that the results demonstrate a reduction in model bias, see discussion of e.g. reductions in "mean difference" and not "mean error". What it does show is that the model is brought into line with the observations, which have been independently verified (e.g. Tilling et al., 2018).

*"The authors could also consider using additional independent data sources for verification SIT analyses such as ice charts."* Ice charts show sea ice concentration, rather than thickness. However, in a follow-on paper, further validation of SIT assimilation in the FOAM system, including the effects on the modelled sea ice concentration, will be shown.

*"The authors report that the sea ice model retains improvements to the SIT field throughout the summer months, due to winter SIT assimilation. However, (Bushuk et al., 2020) investigated a so-called "Arctic spring predictability barrier", and they found that initializing sea ice models with SIT observations prior to May/June is not beneficial for predicting summer sea ice, and, therefore, summer ice thickness observations are strongly required. The authors should discuss in the paper how their results correspond to the previous findings in (Bushuk et al., 2020, Bonan et al., 2019)."*

The reviewer is referring to seasonal forecasting, whereas this is not what our model is doing over the summer. The system continues to run over the summer months, assimilating all other available data types daily, and using them to initialise only short-term (5-day) forecasts. It is agreed that summer ice thickness observations would be great, if only they were available! In any case, e.g. Blanchard-Wrigglesworth and Bitz (2014) found sea ice thickness anomalies in general circulation models to have a timescale of between 6 and 20 months, so seeing the impact of spring SIT assimilation in September is not unexpected. Seasonal forecasting is briefly discussed in the paper already, but further discussion as suggested would be beyond the scope of the paper.

Specific Comments:

*Line 1 and throughout the text. Usually, abbreviation is given in parentheses, i.e., "SIT*

*(sea ice thickness) - "sea ice thickness (SIT)".* Changed

*Line 51. "from retrievals of brightness temperature" - "from L-band brightness*

*temperature measurements".* Changed

*Line 100. “Temperature profiles are also obtained from marine mammals.” What does this actually mean?* Changed text to “Temperature profiles are also obtained from instrumented marine mammals (Carse et al., 2015).”

*Equation (3). Please define e* e is the mathematical constant (2.71828). Text updated to make this clear.

*Tables 2, 3, and 4. Please add dimension (m) where appropriate to the first column of the table.* Added

*Line 420. “may indicate spatial noise” - “may indicate that spatial noise”* Changed

Additional, as mentioned above:

Moved Landy et al. (2020) reference, as incorrect in the context of random uncertainty, and added to discussion of measurement uncertainties along with Nandan et al. (2017).

Added Tilling et al. (2015) reference to section 5 to illustrate independent validation of CryoSat-2 observations.