Response to Reviewer #5 comments

The authors thank Reviewer #5 for their comments. Responses to specific comments are below, with reviewer comments shown in purple.

1) In the assimilation experiment, the converting from the freeboard Fi to sea ice thickness is a key part to affect the SIT forecast performance in FOAM. Compared with the common converting tool like Round Robin Data Package (ESA, 2013), the authors use the model snow depth to replace the climatology. If possible, showing these two types of along-track SIT could be interesting and meaningful to understand the large contrast between the different SIT observation products in Fig. 12c and Fig. 13c.

It is acknowledged that snow depth is an important source of error in the conversion from freeboard to SIT. However, modelled snow depth has a greater spatial and temporal variability than can be obtained from a climatology, as demonstrated by Mallett et al. (2021) and illustrated on Fig. 1. Using this method also maintains consistency between SIT and snow depth within the FOAM model. It should be noted that large uncertainties in the snow depth may apply whether it has been modelled or taken from climatology.

Figure 12(c) shows the good relationship between CryoSat-2 SIT produced using the modelled snow depth and Operation IceBridge SIT observations, and demonstrates that there is no reason to believe that the modelled snow depth is poor, or adversely affecting the SIT observations. Uncertainty in the snow depth will be contributing to differences between Figs. 12(c) and 13(c), although the difference between the CryoSat-2 and Air-EM observations is greater than the snow depth itself and is therefore not the main source of this difference.

In order to quantify and reduce the uncertainty in the FOAM modelled snow depth, future plans will include the assimilation of satellite snow depth observations. It is agreed that a comparison of the effect of modelled and climatological snow depth on freeboard conversion to SIT would be interesting, and it should be included as part of that work. An in-depth comparison is therefore beyond the scope of the current paper, whose main focus is to demonstrate the assimilation of the along-track observations.

For information, discussion on this topic in the paper has been expanded, from line 147:

Currently, CPOM makes use of a modified snow depth climatology, based on Warren et al. (1999) and halved over first-year ice, for processing CryoSat-2 sea ice freeboard retrievals and conversion to SIT (Tilling et al., 2015). This approach is also used by other centres processing CryoSat-2 freeboard observations: Alfred Wegener Institute (AWI; Ricker et al., 2014) and NASA (Kwok and Cunningham, 2015). Instead, here the FOAM modelled snow depth is used. Modelled snow depth has a greater spatial and temporal variability than can be obtained from a climatology, as demonstrated by Mallett et al. (2021) and illustrated on Fig. 1. Using this method also maintains consistency between SIT and snow depth within the FOAM model. A preliminary validation indicates that the FOAM snow depth is somewhat thinner than the modified climatology of Warren et al. (1999), as shown on Fig. 1, particularly over multi-year ice. Tuning experiments demonstrate that simply increasing the snow depth in the model does not result in better evaluation of the SIT analysis against independent observations, owing to feedbacks in the model and between the SIT assimilation and the snow depth itself.

Snow depth uncertainty is a large source of error in radar altimetry sea ice measurements, both in the retrievals of freeboard and the subsequent conversion to SIT (e.g. Giles et al., 2007; Ricker et al., 2015). Due to the linear relationship between SIT and snow depth (Eqns. 1 and 2), an underestimation of the snow depth would lead to an underestimate in the SIT. Large uncertainties in the snow depth may apply whether it has been modelled or taken from climatology. Additional uncertainty is also introduced in Eqn. 1 through lack of knowledge of the snow and sea ice densities which, although constants in the CICE model used here, are spatially and temporally varying in reality (e.g. Alexandrov et al., 2010; Kern et al., 2015). Uncertainties due to variations in water density can be neglected (Ricker et al., 2014; Kurtz et al., 2014). In order to quantify and reduce the uncertainty in the FOAM modelled snow depth, future plans will include the assimilation of satellite snow depth observations.

Reference: Mallett et al., (2021): The Cryosphere, 15, 2429–2450. https://doi.org/10.5194/tc-15-2429-2021

Other references as already cited in the paper.

2) As Line 286 of “an estimate of 50 km for theminimum SIT correlation length scale.” It can infer the spatial scale for assimilation of SIT used 50 km. If that is true, it may be one of reasons why the increments shown in Fig. 14 (a) and (b) are still noise. So there are two related comments: 1) How about to compare the used horizontal scale for the SIC assimilation in this system? 2) Have you tried to increase this scale from 50 to 100 km which is a practical order of the spatial scale for averaging RA data.

Yes, 50 km was used, see line 287: “This value was used as a constant length scale everywhere, except…”.

1) The horizontal scale for SIC assimilation in FOAM is 25 km. Preliminary testing was done using 25 km for SIT, but this resulted in spatial noise and defined track lines appearing in the SIT analysis field, indicating it was too short. As described in the paper, subsequent calculations for SIT put the minimum length scale at 50 km, which improved the analysis results.

2) Figure 4 shows that over the centre of the ice pack, where orbit track lines overlap, the SIT increments do not have gaps between them (unlike for 25 km, not shown in the paper). Choosing the length scale is a balance between not having distinct orbit tracks, but not "blurring" (smoothing) the small scale features captured by the observations. The SIT increments don't look especially noisier than the increments from other observation types assimilated in FOAM (not shown). We haven't tested the SIT assimilation using a 100 km length scale throughout (aside from in the pole hole). However, 100 km is much larger than the grid scale (~10 km in the Arctic) and this may smooth out small-scale features in the observations. However, it may well improve issues of noise in the analysis. Further refinement of the appropriate length scale for SIT (and potentially including a dual correlation length scale, as mentioned in the paper) will be carried out as part of future work. It is however beyond the scope of this initial study, which has used a recognised method (the "Canadian Quick") in order to make the first calculations.

3) To convert the sea ice draft from BGEP by dividing observations by 0.89 is not good for the validations in the ice brake-up and freeze-up months, due to it omits the snow existing. The big issue is that it will mix with the system bias of SIT and return to be detrimental of the analysis in Fig. 15. So I suggest to investigate the best fit lines in the scatterplot are divided into the three interesting periods by the months like MA, ON, MJJAS. It will be helpful to shed light on the MA evaluation to contrast with the validation in Fig. 13 in Beaufort Sea.

It is quite difficult to interpret four best fit lines (including DJF too) on one plot, and the number of observations for the two-month groupings (over all years) is too small to produce statistically significant results. Instead, we have retained the two best fit lines for above and below 1 m on Fig. 15 (which illustrate important issues in the assimilation, as described in the paper) but regrouped the coloured plotting to show MA, MJJAS, ON, DJF.

The BGEP validation in MA 2015-2017 is fairly similar for the SIT assimilation experiment and control (Fig. 15), and this is also the case for Air-EM matchups for April 2015 in the Beaufort Sea (Fig. 13). However, in ON there is an issue with ice being too thick in the SIT assimilation experiment compared to BGEP observations (Fig.15), and a comment on this has been added to the text, relating it to the poor performance for thicknesses under 1 m.

Also added to end of line 222: "…although note this does not take into account the presence of snow on the surface of the ice, which will be the case outside of the summer months." This would also presumably be detrimental to the assessment of the control as well as the SIT assimilation experiment.

4) In the section 2.4, the assimilation of SIT uses IAU as well, and the SIT increments how to feedback on the 5 categories ice in sub-grid, although there are some words about “in proportion to the initial volume distribution”. It should be paid more words or show one example for the reader to well understand how they can work together.

Text updated, from line 233:

“Following Blockley and Peterson (2018), SIT increments are added to each of the five sub-grid SIT categories, if the ice concentration within that category is above 1 %. The initial fraction of the contribution of each category to the gridbox mean ice volume is calculated, and that fraction of the SIT increment is then added to that category. This maintains the initial volume distribution of ice (and ice area) across each sub-grid SIT category.”

Other general comments:

1) Line 97 at P4: “No SIT observations are currently assimilatedoperationally". It is better to use “No SIT observations are currently assimilatedoperationally in the system.” Changed (to “in FOAM” rather than “in the system” for clarity.)

2) Line 193 at P7: “..for the SIT assimilation experiment period at the time of assessment.”

Replaced by “… for assessment of the SIT assimilation in the experimental period.” Not changed, as what is meant is that the V2 dataset for 2015-2017 wasn’t available when the work was being done (but may subsequently become available). This has a different meaning to the suggested replacement wording.

3) Line 246 at P10: the representation uncertainty is set to 0.05 m. Does it mean the minimal observation error is about 0.1 m as the minimal value around the 3m SIT shown by the curve in Fig. 3a? If right, it could be better to be presented on this panel. And this setting could be too small, compared with previous studies and other observation platforms. Yes, that is correct - the measurement uncertainty and representation uncertainty are combined to produce the total OBE (observation error variance). However, we would prefer to keep it separate on Figure 3, to clearly illustrate the measurement uncertainty parameterisation. Figures 3(a,b) state in their titles that they are showing measurement uncertainty. However, the figure caption has been corrected, replacing “Observation uncertainty estimates for SIT.” with “Observation measurement uncertainty estimates for SIT.”

An initial sensitivity study was conducted to find the minimum SIT measurement uncertainty and the shape of the parameterisation curve which worked best with the assimilation system, that is, the optimum balance between the OBE and BGE (model background error covariance). The resulting measurement uncertainty curve is shown in Figure 3(a). However, the reviewer is correct that the minimum measurement uncertainty is rather small, but this is probably achieving a good balance since the BGE is also likely underestimated, as mentioned in the paper. The model BGE will need refining in the future, particularly if (most likely larger) measurement uncertainty estimates are produced alongside the observations. Have changed text from line 289:

"Sensitivity tests were conducted to produce the optimum SIT analysis by finding an appropriate balance between the OBE and the model background error variance (BGE; Sect. 2.4.2). This allowed the final form of the OBE function to be tuned and the minimum uncertainty assigned to the most reliable observations to be set. The minimum OBE is rather small compared to previous studies (e.g. Tilling et al. (2018) gives the accuracy of CryoSat-2 SIT as 13 cm) and this likely due to an underestimate in the model BGE (Sect. 2.4.2)."

4) Line 353 at P18: "The largest regional improvements are in March-April, where the mean difference is above 1.30 m and 1.22 m for the RMSD." Is it possible to specify where the regions are involved. Changed to "The largest regional improvements are seen in March-April (Fig. 8), with reductions in the mean difference of more than 1.30 m and 1.22 m in the RMSD, in the European sector and parts of the Canadian Arctic."

5) Line 396 at P 23, the 30-day periods were chose to cover the observation days. As shown in Table 1 of Section 2.3.1, the observations are located only in one or two days in the April so the 30-day window could be too wide and far away from the reality condition.

Calendar months were only used when producing matchups with the BGEP dataset (as stated on line 397). When producing the matchups with the OIB and Air-EM data, a 30-day period centred on the middle of the observation period was selected for averaging the CryoSat-2 data. The exception to this was OIB data for 2016, which was only available for 20 and 28th April. Since 30 days of CryoSat-2 observations are required, and these cease at the end of April each year, the 30 days was taken from 1-30th April 2016. However, this does not seem to have had a detrimental impact on results. The text has been updated to make this clear (from line 396):

"The 30-day periods were chosen to centre on the middle of the observation window of each yearly field campaign for the OIB and Air-EM validation datasets, and calendar months were used when producing matchups with the BGEP dataset. The necessary exception to this was for OIB in 2016, with fieldwork dates of 20, 28 April (Table 1). Here, the period of 1-30 April was used in order to acquire 30 days of CryoSat-2 data, since no observations are available after 30 April each year until production resumes in the autumn. This means that OIB matchups with CryoSat-2 in 2016 may not be representative of the true relationship between the datasets. Nevertheless, these matchups are not outliers of the OIB matchup group shown on Fig. 12(c), indicating a comparable level of accuracy."

6) Fig.13: The observations from Air-EM are clearly located into two regions: Beaufort Sea and north of Canadian. The related scatterplot separated into these two regions may be more helpful on physic to find something and compared with the result in Fig. 15.

Results for the Beaufort Sea have now been plotted as squares on Figs. 13(c-e), and the Canadian Arctic has been left as filled circles. However, since there are only 13 matchups in the Beaufort Sea from this dataset, the statistics have been left all together in Table 3 to avoid issues with statistical significance. There is already a discussion in the text separating the results for the Canadian Arctic and the Beaufort Sea, but the addition of the location data on the scatterplots backs up these points and so references to Fig. 13 have been updated in the text to refer to these subplots too.

On the comparison to Fig. 15, we have added the following to the end of Sect. 5.3 (BGEP assessment):

"The BGEP data for March-April 2015-2017 (Fig. 15a) compares better with the CryoSat-2 observations than the Air-EM data in the Beaufort Sea for April 2015 (Fig. 13a). This indicates that the Air-EM data may be less reliable, as discussed in Sect. 5.2."