Response to Reviewer 1

We thank the reviewer for his comments and suggestions, which clearly improve the paper. Please find our point-to-point response below.

The manuscript "Drift-aware sea ice thickness maps from satellite remote sensing" by Ricker et al. describes a new algorithm for aggregation of along-track sea ice thickness measurements using corrections from satellite-derived sea ice drift. The produced sea ice thickness dataset is thoroughly validated and supplied with an uncertainty estimate. The manuscript presents new results that are important for climate and cryosphere research. It is well structured and contains all the necessary algorithm details and dataset description details. Nevertheless, a few open questions need to be addressed in the manuscript. I believe it can be recommended for publication after a major revision.

Major comments

The text contains grammar mistakes, and a thorough check by a native English speaker is recommended. I marked as much as I could, but I'm not a native English speaker either.

We did another check of the grammar and fixed mistakes accordingly.

It is good that a separate validation of ice drift trajectories used for tracking the parcels is included. However, the presented algorithm (a data-driven model) contains one more module that can potentially significantly affect thickness distribution: accounting for thermal growth. How accurate is that one alone? It seems important to show the RMSEs computed between the input SIT in the parcels (Line 163), and the fitted values (fH, Line 165). It will probably be proportional to the presented growth uncertainty, but RMSE is a more robust method.

The growth correction includes both thermodynamic growth and dynamic changes, as pointed out in the paper. We acknowledge that the applied linear model is coarse, but more complex models may be susceptible to overfitting and reduce the usefulness for estimating sea ice thickness growth. The quality of the estimate basically improves with data density, i.e. spatially coincident parcels from different data acquisition times. The purpose of applying the growth correction is to avoid biases. For a given 25 km grid cell, we typically only find a handful of parcels from different days, and they are usually not symmetrically distributed around the target day (n = 0); for example, we could find SIT estimates from +14 days, +10 days and -2 days. Averaging this set of SIT values likely results in an overestimation of the mean SIT without growth correction. However, the growth correction is rather moderate, as we use a smoothed field (see Fig.4). And the maximum correction period is 15 days. Therefore, we do not expect large changes to the SIT distribution. In the following, we show an example of the difference between corrected and non-corrected SIT for April 15, 2019 (Fig. R1).

The authors used arithmetic mean (Line 207) to compute SIT from advected parcels. Given that the uncertainty of SIT increases with the advection time (both due to growth and drift), would it be more reasonable to use a weighted average with weights inversely proportional to the time of advection? Alternatively, a limit on trajectory length can be added to prevent points from drifting too far (see also comment below on thick ice blocks in MIZ).

In fact, we also considered using a weighted mean. However, we found that this approach does not lead to better results. On the contrary, grid cells that contain only a few parcels are more prone to biases. The weighting increases the risk of primarily relying on one or very few SIT estimates.

It is not clear from the description how the ice drift is used for advecting parcels, i.e., how the ice velocity is integrated to obtain displacement? Is it a linear Euler method or a more accurate Runge-Kutta 4th order, typically used for computing trajectories in inhomogeneous flows? What is the impact on the accuracy of the advection of the selected method (especially in areas with strong ice drift, e.g., Fram Strait).

At this stage, we only use a 1st order Runge-Kutta (RK1) scheme. We acknowledge that a 4th order approach (RK4) is overall preferable. However, given that the drift fields are fairly coarse (75 km resolution) and smooth, daily displacements between roughly 5 and 20 km/day, and a maximum advection period of 15 days, we argue that RK1 is sufficient. This means that the daily displacement is significantly shorter than the size of one grid cell of the drift product. The advantage of the RK1 approach is considerably lower

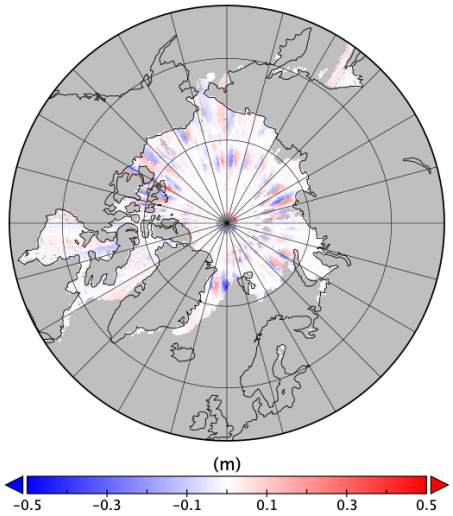

Figure R1: Growth-corrected SIT minus uncorrected SIT for April 15, 2019.

computational costs. But we aim to implement RK4 in the future, especially when using high-resolution SAR-drift products. We have added more details to the description of the drift correction for clarification.

Validation of a non-DA-SIT should be included in section 3.2 Validation with Airborne EM, to be consistent with Section 3.3 Validation with ULS, and illustrate the impact of data awareness. Expectedly, the impact will not be significant, but it needs to be documented.
Yes, we agree. We have added the validation of the non-DA-SIT product and modified Fig. 8 accordingly.

It would be beneficial to homogenize presentation of the comparison – use either 2D histogram (like in Sec. 3.2), or a scatterplot with binning (like in Sec. 3.3). Additionally, the reference data (the independent variable) should be placed on X-axis, and the tested data – on Y-axis, following a common practice.
Agreed. We now use the binning also for Figure 8 and switched the axes.

It is important to add Pearson's correlation or coefficient of determination to the provided RMSE and bias. It's expected to be quite high for the CS2 DA-SIT vs AEM-sit, for example, but nearly zero for Envisat, indicating that it may have low predicting skills for SIT over 2 m.
Agreed. We added Pearson's correlation.

One of the applications of satellite-derived SIT is operational monitoring or assimilation into forecasting models. How efficient can the proposed algorithm be for such an application? Would advection over the last 15 days be sufficient (and important) for providing a more accurate SIT distribution in NRT? Can it be illustrated in the manuscript?
Yes, the product can be operationalized for such applications, i.e. advecting the last 30 or 15 days. If this will improve forecasting models also depends on the model itself and the assimilation. But it will provide a more accurate spatial distribution of SIT on the day of interest compared to traditional monthly means. We have added a sentence mentioning this potential application.

On the video supplement, in mid-January, I noticed some extremely high thicknesses in the MIZ in the Greenland Sea. What is the source of that? It should be either fixed or well presented in the paper with explanations.
Such phenomena might be caused by ocean swell influence on the elevation retrieval. We apply a marginal ice zone filter to avoid such artifacts, but sometimes, affected segments slip through unfortunately.

Figure4. a, left: There are lines with alternating very high and very low growth rates along the CS2 orbits.

Is that an error in the estimation of the growth coefficients or divergence/convergence of sea ice? More light needs to be shed on these cases (e.g., actual time series of advected SIT, divergence estimate from drift fields). It influences a lot the RBF-gridded SIT growth fields at a later stage but doesn't look as convincingly accurate data.
We checked the data particularly for this month. The high growth rate is caused by an individual satellite orbit close to the MIZ during the freeze-up. It might have been affected by ocean swell (see also comment above). This is an issue in the along-track SIT data, which we will try to address in the next version. It also affects the growth estimation, but the effect is small compared to the SIT magnitude of this feature (see also Fig. 5 upper right panel).

Specific comments

Line 2. Operational products are averaged over 2 weeks period, "one-month period" need to be rephrased.
It is true that there are products averaging over a two-week period, e.g., the CPOM product offers 14 days gridded SIT data (but also 28 days). However, this leaves significant gaps between the ground-tracks using a 25 km grid. To achieve a complete coverage, interpolation techniques need to be applied. Most of the well-known gridded products are provided as monthly products (e.g., ICESat-2 L4 Monthly Gridded Sea Ice Thickness, AWI CS-2 product, ESA CCI, etc.). Therefore we argue that the typical period is still one month, if a nearly complete coverage shall be achieved. We slightly changed the sentence for clarification.

Line 5. The acronym DA is very often used for Data Assimilation. May authors consider some other abbreviation for the product name to avoid confusing the users who have not read the paper. "Advected SIT" maybe?
We acknowledge the usage of DA for data assimilation. However, we think that within this paper the usage of "DA" should be clear. In presentations and handbooks, DA is typically spelled out, or it is made clear what it stands for.

Line 10. "enables" $->$ "facilitates"
Agreed.

Line 15. "to a state more closely to" $->$ "to a state similar to"
Agreed.

Line 17. "on a large scale"
Fixed.

Line 26. "one month" is not a standard for, e.g., operational products.
Please see our response to a similar comment above. We argue that the majority of the well-known products is provided as monthly grids. Of course, for certain purposes, a two-week coverage can make sense, but to achieve a nearly complete coverage of the entire Arctic, a period of one month is required (e.g., the CS-2 subcycle is one month).

Line 41. "might get lost" $->$ "is neglected"
We rephrased this sentence.

Line 44. "observer" $->$ "observe"
We rephrased the entire paragraph to improve the flow and fixed the typo.

Line 45. Either "single-synchronized orbits" or "a single synchronized orbit"
Agreed. Changed to "single-synchronized orbits".

Line 98. Improve sentence consistency, e.g.: "Location of parcel centres is selected using the NSIDC EASE-Grid 2.0 (Brodzik et al., 2012) with spacing of10 km in both x and y directions."

Thank you for the suggestion, we rephrased the sentence accordingly.

Line 108. What is the uncertainty of the interpolation? Can it be added to the total uncertainty budget in Sec. 2.4?
We neglect the error of the interpolation here as the spatial resolution of the drift product is much lower than the spacing and size of parcels. Therefore, we assume that the drift uncertainty includes any interpolation uncertainty.

Line 116. It is not clear how the position is adjusted. Please clarify in the text.
As explained in our response to the major comments, we used a first-order Runge-Kutta scheme. We have added a paragraph for clarification.

Line 117. "Consequently, subsequent" $->$ "Subsequent"
Agreed

Line 128. $->$ "some parcels have drifted for more than 200 km."
We re-phrased this sentence.

Line 193. "Python"
Fixed.

Line 211. "swapping" $->$ "sliding"
Fixed.

Line 223. How reasonable is the assumption that the uncertainties are uncorrelated? In principle, the same (or very similar) sea ice is sampled by an altimeter in one parcel. With the same roughness that largely contributes to the uncertainty.
We are aware of the fact that in reality the uncertainties are correlated to a certain extent. However, determining the covariances is very difficult as many factors play a role here, not only the roughness, but also the sea surface anomaly interpolation along track, densities, and other variables. The assumption of un-correlated SIT uncertainties is quite common, and also used in the ESA CCI Climate Data Record. For consistency reasons we do the same here.

Line 225. Is the uncertainty of the interpolation procedure taken into account?
No, as mentioned before, we neglect this uncertainty here. The distance between the interpolated and known value is usually smaller than the resolution of the displacements grid.

Line 261. Previously the product was called "drift-aware SIT", which sounds better than "Drift-Awareness SIT" (although, as mentioned earlier a better name can be found to avoid confusion with data assimilation abbreviation). Please check that the product and the algorithm are called consistently.
Thanks for pointing this out. It should indeed be Drift-Aware SIT in that case.

Line 341. The sample sizes for all bins on Fig. 9 should be provided.
Agreed and added to the figure.

Line 359. "Proves" seems to be an overstatement. Fig. 9b indeed illustrates that in some cases, there is a coincident increase of thickness in both products. However, it also shows that in some cases, it is the opposite (e.g., 10 – 20 Dec 2008). For proving, it is crucial to include r, or r2 scores for the comparison of ULS and DA-SIT, as mentioned earlier.
We agree that "Proves" is an overstatement. We rephrased this using the word "indicate".

Line 367. "latitudinal" $->$ "meridional"
Fixed.

Line 373. “on mainly three factors” $->$ “mainly on three factors”
Fixed.

Line 375. “been”
We fixed this sentence.

Line 383. It is not very clear. Does the conventional product use “parcel positions at the time of the satellite overflight”? Please rewrite.
Yes, exactly. We slightly rephrased this sentence.

Line 386. Quite often “mean difference” is called bias, and “standard deviation of differences” is called RMSD (or RMSE). I would suggest using these terms here and below.
Agreed. We will use the terms bias and RMSD consistently throughout.

Line 390. I would not agree with “In fact, the results for March 2020 appear to be representative overall.”, especially for the biases. Maybe this sentence can be simply excluded? As well as the next one, which doesn’t bring any useful information.
The statement is related to the RMSD/standard deviation, but we agree that the sentence is misleading. We deleted both sentences as suggested.

Line 392. The peak on Fig. 1.a looks quite sharp, more typical for a Laplace distribution. Is the actual type of the distribution important or just the fact that it is symmetrical? Maybe “Gaussian” can be avoided?
We agree that a Laplace distribution might be a good fit as well. However, the important fact here is that it is symmetric and rather narrow compared to other regions.

Figure 11. Please add bias and RMSE for the entire Arctic.
We have added the bias and RMSD for the entire Arctic.

Line 393. “, the higher the mean difference.” $->$ ““, the higher are the mean differences.”, or rather “, the higher are the biases.”
We changed it to ”the higher are the mean differences”.

Line 394. Do you mean the Barents Sea instead of the Beaufort Sea? The Barents Sea has the second largest bias, and the mean bias in the Beaufort is very little.
We actually meant Beaufort Gyre, but Barents Sea should be mentioned here as well. We added this accordingly.

Line 395. “longitudinal” $->$ “meridional”
Fixed.

Line 418. Not only “redistributed”. What about the ice growth in the advected parcels? I guess the standard product does not account for that. Does it add a bias in the new product?
No, the standard product does not account for ice growth. However, on a large scale the effect of ice growth is balanced to a large extent if we refer to the mid-month as the reference time in the standard product. With the symmetric $\pm$ 15 days scheme, ice growth effects from after and before the target day are opposing. On a regional scale, the effect can be more significant, depending on the distribution and timing of satellite overpasses in the respective areas.

Line 423. “Future altimetry missions”, as the authors also mention S1 SAR in this bullet point.
We changed the bullet caption to ”Future satellite missions”.

Line 429. “As an outlook” $->$ “In future”
Fixed.