Response to review from Veit Helm

Firstly, the authors would like to thank the reviewer for taking the time to review this paper and for the kind words!

Here is the point-by-point response to the reviewer's questions in blue text:

Review: Elevation Change of the Antarctic Ice Sheet: 1985 to 2020

Nilsson et. al.

The paper provides a new data set of elevation change of Antarctica based on analysis of satellite altimetry.

The authors put in a massive amount of work to provide this comprehensive dataset. Thanks to the authors.

The authors used for most of the time L2 data from earlier mission for their study. They applied a slope correction (similar to Schröder et.al.) and which is new, included ICESAT(2) data in their study. For CryoSat-2 they applied an independent processing using L1B waveform data.

A new approach to derive elevation change is the decoupling of time-variable and static topography. Then both spatial and temporal patterns of changes in the scattering horizon are estimated and corrected for using waveform parameters and sigma in a separate step Combination of the different mission, which is very challenging was done using a least square approach which also differs from other studies. Another new step is the amplitude normalization to further reduce seasonal amplitudes explained as residuals of not fully corrected radar scattering effects. Finally, an extrapolation method is suggested to fill the pole hole south of 81.5°/88° to provide a complete dataset for Antarctica to be used e.g. for model input. The new data set shows slightly higher accuracy then the TUD and CPOM dataset when compared to ATM or the 2003-2019 ICESat(2) data set of Smith et.al.

To my opinion the paper as well as the data set is of high quality.

The dataset could be accessed through the given link.

I think the data set is significant, unique, useful, and complete and of value for the community and worth to be published.

Language and figures are excellent.

Methods applied are described in detail. Validation as well as comparison to similar datasets are thought-out and well explained.

Results and the development of elevation change of Antarctica over the last 3 decades are well presented.

I have some open question in regard of the uncertainty estimate - see below.

and some minor points:

1.

I was not comfortable with the data handling:

I have the feeling that the chunk dimension of the data in the netcdf file is not applied or is too small or applied to the wrong dimension (I'm not sure). This makes it extremely difficult to read in the data in an acceptable amount of time.

E.g. using ncview takes ages to scroll through the layers. I also used IDL to readin the data and it took very long. Finally, I converted the data myself to be able to check the data quality in a reasonable amount of time. I used a chunk dimension of (x_dim,1,1) when writing my own netcdf file. This improved the access time.

Yes, thank you for that it has been under discussion inside our group as well and we took action to improve the reading time of the product by modifying the current chucking scheme.

2. Error estimate

A very important point is your error estimation. To me it's not clear if you used uncertainties estimated in 4.1 in the integrated error estimation of 4.3. It also not clear what is provided in the product.

Thank you for pointing this out we have made this clearer in the text of what is applied and at what time. Further we have also removed one of the error components (sigma-x) as it was in error. The new total error assigned to each epoch in each time series is now only dependent on the cross-calibration error (sigma-c) and the spatiotemporal variability inside each monthly time interval (sigma-m). Table 1. now reflects the overall quality of the different missions and modes and is mapped directly into monthly uncertainty via the variability. The text has been rewritten to reflect this and that the error varies both by location and time. L497-512

sigma_x I guess are fixed numbers for each mission or mode. Can you provide a table?

Is sigma_c spatially varyiing? I don't get how sigma_m is computed, which variability is used? Please give more details.

Is sigma_m computet for each grid cell?

In table 1 sigma_rand and slope are shown. How is this used in equation 5?

We agree that is somewhat confusing and we have now made clear that this shows the overall mission/mode error, but is not used specifically in the time-series uncertainty. We have re-written a large part of the initial section of 4.1. L497-512

In section 4.3 you estimate sigma_s and sigma_r by comparing to the Smith product. This might make sense for the same time period, but is this valid for the period before 2003?

We totally agree with the reviewer that sigma-s and sigma-r are not appropriate as a stand-alone error for data before the 2003 time period, as it would provide the same error for previous missions as the missions after 2003. We can also see from the crossover-analysis that the relative precision of the different missions decreases as you go back in time. This is the reason that we included the third-term in the error budget to better account for the inherent error for each mission over the different time intervals. Our reasoning is that the systematic bias estimated from the IS1/IS2 validation over the 16-the period provides a baseline. This baseline is then modulated with the inclusion of the random error sources estimated from our analysis. We believe that this provides a conservative estimate of an absolute error. Previous studies have only used relative variability in their estimates and/or errors estimated over short time-spans or regions from airborne data.

In equation 7 sigma_m is mentioned. Is this the same sigma_m as in equation 5 and is this the rmse in your product?

Yes, we apologize that was not real clear and this has now been changed to reflects this better. The change was a part of the re-write of Section 4.1 and you can find specific changes on L508-512 and a change in eq.5. As we also said to the other reviewer there was a small mistake in the equation (double counting one of the error sources) which has now been corrected.

E.g. As a user of your product I'm interested in Pine Island drainage basin for the period 2011 to 2015. How is the procedure to estimate the uncertainty for this given time period for this specific basin and for total Antarctica? Can this be done with the information given in the data product?

This can be done in the following way: Select the basin (ROI) and the time interval from the cube (slice the data in time and mask out areas you are not interested in). Then take the monthly spatial error fields and compute the integrated error for each month. Once, that's has been accomplished compute either the mean or RSS of the errors over the time interval. This will provide the random error (sigma-m) for your ROI over the selected time interval. This can be mapped into an error rate either by dividing by the time interval or by using equation 7. To provide the standard error you can divide the rate-error by the number of un-correlated grid-cells using the correlation length provided in Table 2 (sigma-h^dot). The absolute error can be derived by following equation 7 using the bias (sigma-s), error (sigma-r), area and correlation lengths from Table 2. This follows the approach outlined in this study to provide the tabulated errors in Table 3. Further, if needed sigma-r and sigma-s can be replaced by values from the error model in Figure 7(g,f), which are then integrated over the ROI and substituted into equation 7 using an appropriate correlation length.

It's also not clear if interpolated, extrapolated and observed gid cells are handled in a different way. Furthermore, the uncertainty given for the pole hole looks pretty strange. The whole area is extrapolated but shows very large spatial differences in the RMSE, can you explain why?

The extrapolated grid-cells are handled a bit differently and the predicted errors from the algorithm are multiplied by three to make sure they are not too small and to downweigh them if they are used in time-series as weights. Further, as the only the closest 200 points of the 20 km averaged data are used, it can create a funnel pattern. This has only data from one side of the pole-hole is used in the extrapolation. This can be seen in the error-fields where large errors from the transantarctic mountains are funneled to the pole due to data location. However, this can just be removed by replacing it with a mean-error using the provided pole-mask. It's just our "best" try to provide something consistent and hopefully usable based on actual "local" data. We have revised this section of text in the manuscript to detail this behavior. L456 – 476.

3.

In the text you mention that the bedmap2 elevation model was used. However, the data set itself provide a different elevation model. This is not consistent.

Furthermore, I'm wondering why the old Bedmap2 data set is used instead of using the Tandem-X or REMA DEM's. They are much more accurate and provide reliable data in areas south of 86°.

The bedmap2 DEM was only used for the slope-correction of the older missions as the data used to make the DEM overlap in time with the missions (bedmap2 contains both ERS and ICESat data). For the slope correction we found that a resolution of about 2-3 km provide the best results, thus a high-resolution model was not needed. In general, the only mission that reaches above 86 S that needs a slope correction is CryoSat-2 LRM. The slopes are very gentle in this region and in combination with resolution of the DEM the magnitude of the correction is small. The model provided is of auxiliary use if the user wants to look at time-evolving topography. We have made this clearer in the text. L480-485

4.

Why is it not possible to use Envisat data after the orbit change? Did you try with and without or weren't you able to get good dhdt estimates due to data coverage issues?

Line 130: what exactly is a segmentation filter. How does it work? Please explain in more detail.
With the new reference orbit, the tracks are not on the older repeat tracks anymore and thus when removing topography, we found that the noise levels went up due to the need for either a larger search radius and/or it produced an offset in-between the two Envisat orbits. This provided an issue when trying to cross-calibrate the RA2 vs CS2, at least for our study. We are currently working on a way to mitigate this and have made some headway on this and this will be included in the next version of the product.

The segmentation filter is just a difference filter for the IS2 point data to remove data points above a specific threshold (it comes from Smith et al 2020). We compare point “i” with point “i+1 “and if the difference is larger than a threshold its removed/flagged. Its currently set to 2 m which has been derived empirically. We added some text to explain this (L134-135).

5.

To me it’s not clear how you handled the different ERS modes. How is the coverage of ICE and OCEAN modes? Do you have data in both modes for each month covering the whole ice sheet or

separate months with ICE or OCEAN or specific areas with one mode for the whole period?

I miss in Fig.3 the ERS1/2 OCEAN mode.

ERS-1/2 are divided into ocean and ice data sets (like in Schroder et al 2019) and treated as different missions and processed independently (topography, scatter correction, bias estimation etc.). However, the ocean mode does not contain that much data so that’s why we didn’t include any figures as they do not provide that much information. We have updated the text to make the treated clearer. (L314-319)

6.

The combination of Enviat and ICESAT is not fully explained. At which point you combine these two products? The same for ERS1/2 ocean and ERS1/2 ice.

Do you combine it before the multi-mission cross calibration is applied? Maybe it’s worth to show two examples’ figures of certain grid cells how such a combination works (e.g., an grid cell with data gaps and without).

All missions or modes (ocean, ice, lrm, sarin etc) are initially adjusted using the least squares approach. The nifty thing with this approach is that at each local-grid cell all data are collected and a mean-offset relative to a target date is estimated for all the different datasets. These offsets are then subtracted providing a cross-calibrated time series (step-1). In the first step independent offsets are estimated for LRM, SIN, E1-OCEAN, E1-ICE, E2-ICE, E2-OCEAN RA2, IS1 and IS2. Once the first cross-calibration has been applied the missions are grouped and a secondary adjustment is performed. This is done to account for more non-linear behavior using the residuals to the model. We group the missions as follows: ERS-1 (Ice+Ocean), ERS-2 (Ice+Ocean), RA2/IS1, CS2 (LRM+SIN) and IS2. Then we compute offsets in-between these grouped using the data where they overlapping in time. Once, the data has been adjusted for the second time we integrated into a consistent time series using a weighted average. We have added some more text to make this clearer and a figure of the cross-calibration is provided in the supplementary material (Figure S1). (L314-319, L332-335)

7.

How much does the normalization of seasonal amplitudes change the trend estimates?

I'm not sure if this kind of normalization should be applied. The point is that you apply correction based on correlations with sigma, LE and TE for each mission. Those correlations make sense and reduce the seasonal amplitude. However, the normalization has no physical explanation. Maybe it’s' worth to check and show the seasonal amplitude of the CryoSat L2 product. If the amplitude is similar to your own processed product than the normalization is questionable.

Otherwise, you have an argument that due to your low-level threshold retracking the time varying signal penetration is strongly reduced. Maybe it's also worth to show in an Appendix for each mission the Antarctic wide reduction of seasonal amplitude. I think this can help to understand where the corrections have largest impact and where largest amplitudes are observed and if this is mission specific.

It's also not clear at which point this normalization is applied - before or after the mission cross calibration?

If it is applied before, then you should change the order of the sections in the text.

We knew that using the normalization would be a subject of discussion which is very welcomed. To answerer your first question: No, there is no major effect on the trend as the trend is removed before the correction is applied and then added back. The normalization effect is based on previous work where we did show the difference in seasonal amplitude (Nilsson et al. 2016) between different processing approached for CryoSat-2. This using our own CS2 product compared to the standard ESA L2 product. The results showed that there was clear difference in the magnitude of the seasonal amplitude depending on the choice of retracking technique. Both in that study and in the one presented here we see a much better agreement in amplitude in-between CS2, IS1 and IS2. This holds for both ice sheets when performing the same analysis in Greenland. CS2, IS1 and IS2 are almost perfectly aligned in seasonal amplitude compared to the older mission (RA2, ERS-1 and ERS-2). So, we believe that our justification of applying the correction is valid for the older missions. This as it is obvious that the scattering correction is not fully capable of removing all the artificial signals. Further, as an independent comparison we did compare the amplitude against the RACO FDM product and found a remarkable agreement. This further gave us confidence for the justification of the correction. We have added tables in the supplementary material (Table S1/S2), where we compare the amplitude of the RACMO FDM product to the amplitude estimated form our product over different time periods (missions and total) to help illustrate this in more detail. The normalization correction is applied after the calibration step. Data quality-wise it made little difference if it was applied before or after.

8.

Line 450: Here the reference to the figure 2 is not correct. It should be Figure 6.

Line 517: You mention a correlation length of 100km, however Table 2 list different values. Which one was used?

We used 100 km for the correlation length. (L575 and L462)

9.

Line 608: Do you have any idea why your product is not closer to LA in WAIS? For EAIS and AP they are and this is what I would suppose, as you also used ICESAT in your approach?

This is in my view a classic example of the difficulty of radar to measure in high slope areas. The radar can't capture the same amount of signal as laser, as also seen by other products. IS1, due to its lower temporal and spatial sampling, is kind of over-shadowed by RA2 here. Maybe in the future version of the product we might be able to better account for that?