Replies to comments by reviewer 1

We thank the reviewer for their constructive comments, and we have addressed the points raised as described in following. The reviewer comments are shown in blue, and our point-to-point replies are shown in black.

This is my second review of this paper. Compared to the original version, the authors have made significant improvements, particularly in terms of data sources and methods. Most of the issues I previously raised have been addressed satisfactorily. I recommend that this paper be accepted after minor revisions.

Here are my several comments:

1. The authors emphasize that only in summer can the two types of altimeters obtain consistent elevation observations at the edge of the ice sheet, and they also generated the DEM based on this assumption. While this approach is feasible, why do we need to do this? What is the practical use of obtaining a DEM for the edge of the ice sheet in summer?

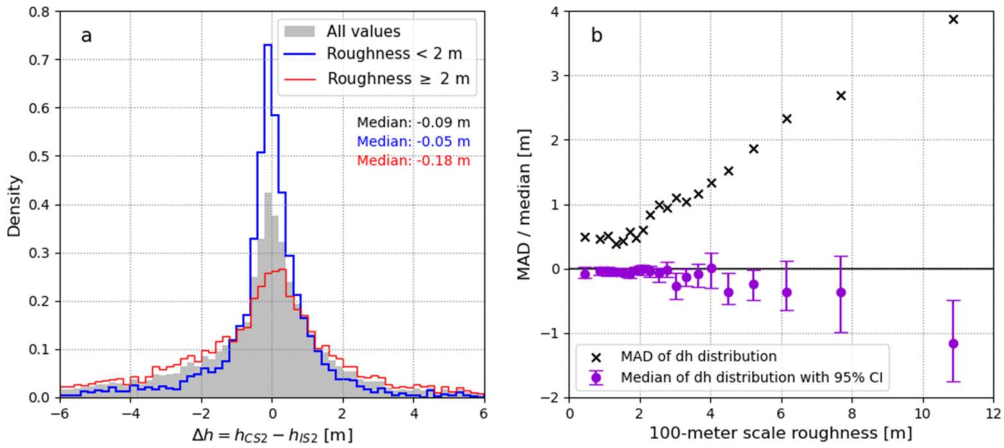
While it would indeed be valuable to have annual DEMs for the entire ice sheet, an annually-updated marginal DEM has several use cases. First of all, the margins experience the largest changes, and hence these areas are most important to monitor on short timescales. One specific use case for the PRODEMs is to improve the ice sheet mass balance estimates using the mass budget method, for which the ice sheet mass loss is calculated from the mass flux through gates located across all outlet glaciers. This requires knowledge of ice sheet thicknesses across the fluxgates. Accounting for the annual changes in surface elevation will therefore improve the mass balance estimates, and, not least, its interannual variability. For this purpose, summer DEMs have the added advantage of lesser complexity during density conversions, since we assume that there is no snow or firn.

We have expanded the mentioning of this in the paper to make this use case apparent to the reader. It now reads (additions/changes in red):

L. 73: The PRODEMs will be valuable for marginal mass balance assessments, as they provide crucial information on the inter-annual variability in surface elevation across these pivotal areas for understanding the complex interplay between ice sheet, oceanic processes and climate dynamics. Further, one specific use case of the PRODEM series is to improve mass balance assessments for the entire ice sheet based on the mass budget method (Mankoff et al., 2021; Mouginot et al., 2019; Van Den Broeke et al., 2009)*, in which the solid ice discharge is calculated by summing the contribution from all individual outlet glaciers based on measured ice velocity and ice thickness* (Mankoff et al., 2020)*. The annually-changing PRODEM surface elevations will support the ongoing assessment of solid ice discharge, as carried out e.g. within the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) project* (Ahlstrøm & the PROMICE project team, 2008).

2. The authors have analyzed the elevation differences obtained by the two types of altimeters. Although the authors believe that the elevation differences between the two types of altimeters can be ignored, I still see a meter-level deviation between them in Figure 4. Can this deviation be considered negligible? The author needs to further clarify this point or find a way to achieve some consistency between the two datasets. The differences are greater on rougher surfaces, and the edge of the ice sheet is also rough.

In the roughest parts of the PRODEM area, it is not possible to obtain very good statistics when comparing the elevations obtained from the two altimeters near intersecting tracks: Indeed, the median of the distribution of elevation differences is ~-1 m, but the spread of the distribution is substantial (MAD = 4m), indicating that the median difference is not well determined. To quantify this point, we have computed bootstrap confidence intervals for the median of the distribution of elevation differences, and these are added to Figure 4 (see updated figure below).



We note that the median of the dh distribution with 95% confidence is less than 0 m only for areas of extreme roughness. Within these areas, the spatial uncertainty associated with especially the CS2 elevations is large ($\sigma_{cs2}^{spatial} \approx 1.5m$, $\sigma_{is2}^{spatial} \approx 0.3m$; computed based on the equations in Fig. 2b,d). The detected bias is therefore within the margin of the measurement uncertainty, supporting our claim that, even within these areas, the observed bias is not practically significant.

From a physical view, we would also expect the elevation differences to be largest in areas of potential snow cover, e.g. at high altitudes, where the surface tends to be relatively flat. Within these areas, we observe very small elevation differences (5 cm), leading to our statement (L 337) that we consider this value to be an upper limit for the potential bias between the two altimeters over bare ice.

The text has been updated as follows:

L 330: The median of the distribution tends to be smallest in magnitude in areas of low roughness (-0.05 m; Fig. 4a), where the elevation differences are best determined and the dispersion of the elevation difference distribution is small (Fig. 4b). Only in areas of extreme roughness does the median elevation difference fall outside the 95% confidence interval. This deviation from zero is only a fraction of the large (meter scale) observation uncertainties within these areas (Fig. 2b,d), suggesting that this apparent bias may not reflect a true, systematic difference between the two elevation estimates.

3. The authors mention in the paper that Kriging interpolation was not used to interpolate the surface elevations but rather the difference between the altimeter elevations and ArcticDEM. What is the physical basis for this?

We consider this approach to have several advantages, as also detailed in the paper. To sum up the two most important points:

- Removing a background field is a standard kriging preprocessing step. Kriging requires the field to be stationary, meaning that it cannot be applied to directly to the elevation field, which displays a clear trend (low elevations in the margin; high elevations in the central

area). By subtracting ArcticDEM we obtain a first-order stationary field, to which we may robustly apply the kriging equations (L 395). We now also mention this earlier in the text where the elevation anomalies are introduced:
L358: The resulting data set is labelled elevation anomalies. This is a standard preprocessing step prior to kriging (e.g., Dodd et al, 2015, Loonat et al. 2020) to ensure statistical robustness, as the method assumes stationarity of the mean field (see also Section 5.3)

- The approach allows us to take advantage of the high spatial resolution of ArcticDEM, and to also produce reasonable elevations in areas with limited data (Section 9.1).

4. The author only provides a general explanation for choosing the 500 m resolution. What is the specific reason for choosing this resolution? Has there been a comparison of different resolutions to select the optimal one?

We do not have much choice regarding which resolution to produce the map of elevation anomalies, as we cannot go beyond or below the resolution of the input data: For the applied approach ("Point Kriging"), we must align the grid resolution with the scale of variability of the observations (L197). The resolution of the CS2 data, which is a few hundred meters, is therefore the limiting factor, leading to an appropriate resolution for the elevation anomaly field to be on the order of 500 m (L198).

Reply to reviewer comments by Romain Hugonnet

We thank Romain Hugonnet for their constructive comments, and we have addressed the points raised as described in following. The reviewer comments are shown in blue, and our point-to-point replies are shown in black.

General comments

The authors have satisfactorily addressed all of my comments, even exploring potential improvements that I had not envisioned, selecting the most performant ones and revising their manuscript accordingly.

In my opinion, the manuscript is almost ready for publication. I just have a minor comment on further refining the text.

Minor comment

A lot of text has been added (in the Introduction, Section 4 and Section 7, in particular), with statements that are slightly less polished than the rest of the manuscript. I picked up a few errors in my line-by-line comments below, but probably missed things and didn't want to propose rewordings. Be sure to revise this text thoroughly.

We have thoroughly gone through the text (with primary focus on the introduction, section 4, section 7) and revised the language and structure accordingly.

Additionally, I was not convinced by the large expansion of the introduction (though it followed a comment by referee 2), especially moving the datasets descriptions there. I found it a bit too long to grasp the topic rapidly and getting into the paper. I think it would benefit from being more concise and having the long dataset descriptions separate as originially.

Upon re-reading the current version of the manuscript, we agree with the reviewer's comment – the introduction has become too long. Yet, we also consider it a valid point previously raised by referee 2 that the introduction of the paper would benefit from setting the PRODEMs in context of previously constructed DEMs.

We have therefore revised the introduction as follows: we have removed most of the "brief history of DEMs", and now instead describe the existing range of DEMs in more general terms, with particular focus on how they differ in terms of temporal span and spatial resolution and coverage. The data set descriptions have been moved back to their original place (Section 2).

We hope that both reviewers will be satisfied with this change.

Several sections (for instance the one introducing kriging) are also a bit "educational", and could be cut short when the statistical information is available online (equations or definitions) and the general concept is already introduced.

These sections were expanded in the previous round of reviews to provide further clarity on kriging, as additional explanation was incorporated following feedback from reviewer 1. We consider it very important that the reader is aware of what kriging is, including the prerequisites of using the method, as it is the foundation of the DEM generation. We have therefore decided to keep this section as-is. Regarding the kriging equations: Several versions of the ordinary kriging equations exists in the literature, and we wish to specify which ones we have used.

Finally, and critically, the authors lack a discussion about the application of the methodologies that are at the core of this study: 1/ their data fusion scheme (regionally-varying kriging) and 2/ their uncertainty analysis (estimation of spatially variable errors, correlations). How does those compare to existing studies on the topic applying similar methods to DEMs, dh, or elevation anomalies? Right now, as a non-expert reader (user of PRODEM): either everything seems brand-new, or it is under-referenced. It would largely benefit the study to compare to the existing (=detail where it compares or builds upon existing work, and where it is applying new methods/considerations for this type of application).

We have added the following text to the manuscript at appropriate places to ensure that the reader is aware that some of these methods (or relatively similar ones) have indeed been used in the literature previously, although in a different context.

L. 204: …if uncertainties are underestimated, the interpolation may overfit the observations while inducing noise in the interpolated field. Nevertheless, the observation uncertainties have been largely disregarded during the construction of existing altimeter-derived ice sheet DEMs.

L. 228: To estimate the CS2 spatial uncertainty, we analyse the measured elevation differences at temporally-close intersecting CS2 satellite tracks within our study area, in an approach that lends from the cross-over methods used for deriving local longer-term elevation change rates (Khvorostovsky et al., 2003; Sørensen et al., 2018)

L. 357: Prior to interpolation, the satellite altimetry is detrended by subtracting a reference DEM (ArcticDEM v4.1) from each elevation measurement using linear interpolation. The resulting data set is labelled elevation anomalies. This is a standard preprocessing step prior to kriging (e.g., Dodd et al., 2015; Loonat et al., 2020) to ensure statistical robustness, as the method assumes stationarity of the mean field (see also Section 5.3). When using other methods to produce DEMs (e.g., Fan 22), such processing step may not be a requirement.

We do not know of any other studies applying regionally-varying kriging similar to ours, but we consider it likely to have been used in other fields – although probably in a slightly different implementation, and possibly under a different name. We have therefore added the following sentence to the paper:

L. 408: For a given location in the PRODEM grid, an appropriate set of variogram parameters is subsequently extracted and applied in the kriging routine (section 5.3.2) to produce an interpolated value for the elevation anomaly. We call this method "regionally-varying kriging".

L 883: The PRODEMs are interpolated using regionally varying kriging on elevation anomalies relative to ArcticDEM, in an approach that may also be applicable for interpolation of other semi-stationary fields.

Short reply to author response

On my comment about "dropping the nugget": The authors are right, in their case the nugget also represents the uncertainty, I went a bit too fast here. In the Gaussian Process side of kriging I am also familiar with (which I recommend the authors to dive into if they want to scale their analysis to more dimensions/parametrizations), short-scale variance and uncertainties are always separate parameters and not contained into the same one.

Line-by-line comments

232: Dartnell et al. (2000) or Wilson et al. (2007) are the reference for this calculation of the surface roughness

Wilson et al (2007) has been added as reference.

585: Reword colloquial term “cleaned”

Reworded to “filtered”

666: Remove “Click or tap here to enter text”

Removed.

Table 1: Add unit (or specify if it is normalized)

Unit has been added

923-924: “That errors in elevation estimate substantially larger than the prediction uncertainty are much more frequent than theoretically expected” → “elevation estimates”? And yes, but this is only true for a small percentage of the data, which is worth pointing out. Overall this whole sentence could be revised, it’s a bit convoluted.

The wording has been changed to the following:

L 640: The heavier tails, on the other hand, indicate that the reported uncertainties do not capture the full range of variability, and that, for a small percentage of the data, uncertainties are severely underestimated.

946: “robust” → you mean “consistent”?

Yes, the wording has been changed

Style:

- Quite surprised by first-letter upper-casing, such as for “Kriging” of “Block-LOOCV”: I don't think it requires any.

Casing has been changed.

- Defining an acronym also does not require upper-casing either, for example “We use leave-one-out cross-validation (LOOCV)” is correct, and can be continued along the text.

This has been changed

References

Wilson, M. F. J., O’Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. Marine Geodesy, 30(1–2), 3–35.

Dartnell, P. 2000. Applying Remote Sensing Techniques to map Seafloor Geology/Habitat Relationships. Masters Thesis, San Francisco State University, pp. 108.