Replies to reviewer #2

We thank the reviewer for their constructive comments and have addressed the points raised as described in the following, with particular focus on clarifying the kriging procedure and underlying assumptions. The reviewer comments are shown below in blue, and our replies shown in black. Line numbers refer to the original version of the manuscript that has been reviewed.

Since the manuscript was submitted, several of the input datasets have been updated:

- We use CryoSat-2 Baseline E for all years, as it is now available for the entire period
- ICESat-2 ATL06 has been updated to version 6; the new version has corrected the geolocation error found to exist in version 5
- ArcticDEM mosaic has been updated to v4.1 (as also proposed by the reviewer)

We have produced a new version of the PRODEMs using the updated data sets, and the text and figures in the manuscript have been adjusted accordingly. This includes e.g., an updated version of figure 10 now displaying another area that shows a similar checker-board pattern in elevation anomalies relative to ArcticDEM v4.1 (albeit less pronounced than previously found when comparing to ArcticDEM v3), and removal of figure 11 (including the related discussion) which is no-longer relevant.

We hope that with these and the below changes, you will accept to review an updated version of our paper.

This paper introduces a set of four 500-meter resolution annual Digital Elevation Models (DEMs) representing the Greenland ice sheet marginal zone during the summers of 2019 to 2022, referred to as PRODEMs. Covering a 50km wide band from the ice edge, these DEMs meticulously encompass all outlet glaciers of the Greenland ice sheet. The integration of DEMs derived from ICESat-2 and CryoSat-2 constitutes a commendable initiative, providing an additional valuable source of topographic information for Greenland. While the paper is well-crafted overall, there are certain statements that require precision to ensure accuracy and clarity. It is crucial to address these points before considering the paper for publication.

General comments:

- The authors' decision to generate the DEM specifically for the summer and focus on the marginal zone of the ice sheet raises pertinent questions. What rationale underlies the choice of season, and what significance does the selected ice sheet margin hold in this context? Furthermore, a clarification regarding the specific function of the DEM under these conditions would enhance the paper's comprehensibility.

This is described in L114-120. In short, we wish to develop the DEM for areas (and periods) where the surface is minimally covered by snow since the two sensors (CS2, IS2) interact differently with the snow. This leaves the marginal zone during summer, with summer being the period with least snow cover, which we here take to be June through September. The selected area encompasses most of the summer bare ice zone, but to avoid a product with sparse and/or intermittent coverage, and to maintain the same coverage across years, it was decided to cover the entire 50km margin of the ice sheet.

To clarify this to the reader, we now also mention this earlier in the manuscript (L44):

Area and time period are selected to obtain areas minimally covered by snow, for which the two satellite sensors are expected to measure the same surface.

The resulting DEM is therefore the ice surface, without snow cover, which is e.g. the DEM appropriate for mass balance considerations in the ice sheet marginal areas. This has been clarified in the paper (L46):

The PRODEMs therefore represent the ice surface topography, which is e.g., the DEM appropriate for marginal mass balance considerations.

- The authors employ two types of altimeter data for DEM generation, a method requiring careful consideration of their consistency. While the authors assert that the disparity between the two datasets is negligible, it is crucial to acknowledge that applying these datasets without consistency correction may influence elevation estimations. Particularly noteworthy is the comparison between IS2 and CS2; the latter exhibits uneven spatial distribution and coarse resolution, introducing potential biases in the generated DEM. My concern centers on the relatively lower measurement accuracy and disparate spatial distribution of CS2, aspects that warrant careful attention.

We agree with the reviewer that it is most important to account for potential biases when combining information from the two types of altimeter data. In this respect, we emphasize that because of our choice of region and period (summer-time margin only), it is believed that there should be no elevation bias between the two altimeters, as supported by existing studies and well documented in the literature. We now refer to these studies in the text (L275):

The PRODEMs, however, are constructed based on elevation data from the ice-sheet marginal areas during summer, where the snow cover is limited. In the bare ice zone, it is well documented in the literature that the elevation bias of radar altimeters is negligible (e.g., (Dall et al., 2001; Otosaka et al., 2020; Davis and Moore, 1993)), and we therefore expect small biases only between the employed CS2 and IS2 elevations within the PRODEM area.

Nevertheless, we cautiously conducted a thorough analysis of potential biases by comparing elevations measured at crossing satellite tracks of the two satellites, and we found that (as expected) the disparity is negligible. Indeed, not even the sign of the median of the elevation differences is consistent among basins (L280-281).

As the reviewer correctly points out, the CS2 data is associated with lower measurement accuracy than IS2. During construction of the PRODEMs, we account for this by assigning uncertainties to the individual observations based on the sensor type (as well as location and acquisition time). We have in the updated version of the PRODEMs and manuscript thoroughly revised the uncertainty assessments to now fully consider the uncertainty contribution from spatial and temporal uncertainty for each observation. The updated distributions of uncertainties for the CS2 and IS2 observations are shown in the updated Figure 3, see below. As for the previous PRODEM uncertainty calculations, the uncertainties remain significantly larger for CS2 (median: 0.98 m) than for IS2 (median: 0.21 m).

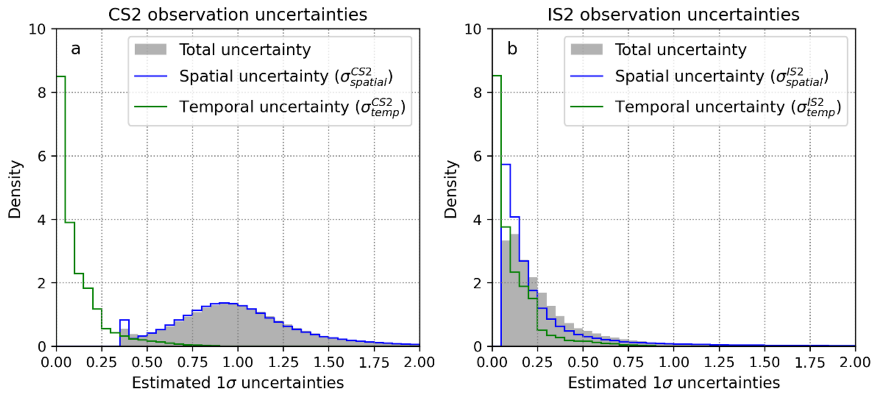


Figure 3: Distribution of derived total uncertainties (grey), along with the individual spatial (blue) and temporal (green) uncertainty components for the CS2 (a) and IS2 (b) altimetry data, respectively. Data from summers 2019 through 2022.

Another aspect brought up by the reviewer is the different resolution of the two datasets, with the IS2 data having much higher resolution than the CS2 data. In order to work with consistent datasets, we down-sampled the IS2 data along track to 250m resolution, which is of similar order as the resolution of the CS2 data (L90). We have added the following (in red) to this sentence:

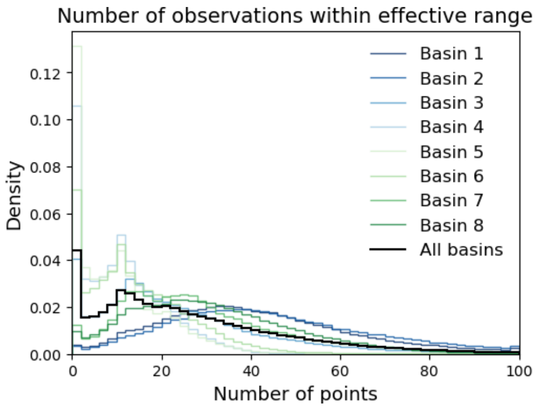
For consistency with the resolution of CS2 observations, and given that the topographic variations to be resolved in the 500m resolution PRODEM are much smoother than 20m, we further down-sample the ATL06 data by computing median values over 250m along-track segments of each beam.

The last aspect of the differences between the two measurement types mentioned by the reviewer is the difference in their spatial coverage: While valid CS2 observations do not exist for the rough terrain often located along the outer margin of the ice sheet, the IS2 radar is able to obtain valid observations also for these areas. Additionally, the IS2 data are acquired along straight lines, whereas the CS2 observations are spatially more scattered. However, these differences in spatial coverage will not affect the interpolation of the resulting DEM, apart from resulting in generally higher uncertainty of the resulting product in areas with less data.

- Regarding data coverage, the authors applied distinct interpolation radii in different regions. To enhance reader understanding, it would be beneficial to include information on data coverage, such as footprint numbers for CS2 and IS2 in each grid. This insight can shed light on the grids effectively covered by the altimeters, offering rationale for the choice of a 500 m resolution. It's noteworthy that the ICESat-2 DEM by Fan et al. (2022), constrained to 10 footprints at 500m resolution, covers only approximately 30% of the entire ice sheet, and a discussion on this limitation could be insightful for the readers.

It is actually not very relevant how many observations are located within each grid cell. Consider, for instance, an area of very flat terrain: It is still possible to interpolate a reasonable elevation value here, even for a grid cell in which no observations exist. The important factor is the number of observations within the correlation length of the varying terrain (i.e. effective length scale of the variogram). If no observations exist within this interpolation radius, the interpolated value will be set equal to the local median field.

We have added a figure to the paper showing the number of observations within the spatially varying effective length scale from each grid cell (see below), along with a discussion. We find that 13% of the grid cells suffer from sparsity in neighboring data, which we take to be less than 5 observations within a radius corresponding to the local effective length scale. And for 7% of the grid cells, no observations exist within the interpolation radius. The issue is most pronounced in the southern regions; for the southernmost drainage basin (basin 5.0) the percentage with sparse data coverage has increased to 37% (21% without any nearby observations). In general, however, the interpolated field is based on 22 nearby observations (median of the distribution). In regions without any nearby observations, the ArcticDEM elevations are adjusted with a constant depending on the local median anomaly field, as calculated from the 200 closest altimetry observations, and the resulting elevation field is thus highly reliant on ArcticDEM.



An additional parameter that one may use to evaluate the coverage of data points is the distance from grid centers to the closest data point. Histograms with these distributions are found in Figure 1a. When discussing this figure, we now have added a mention of the total percentage of grid cells having the closest data point more than 1 and 2 km away (L129):

For 19% of the grid cells, the closest observation is more than 1 km away, but for only 4% it is more than 2 km away.

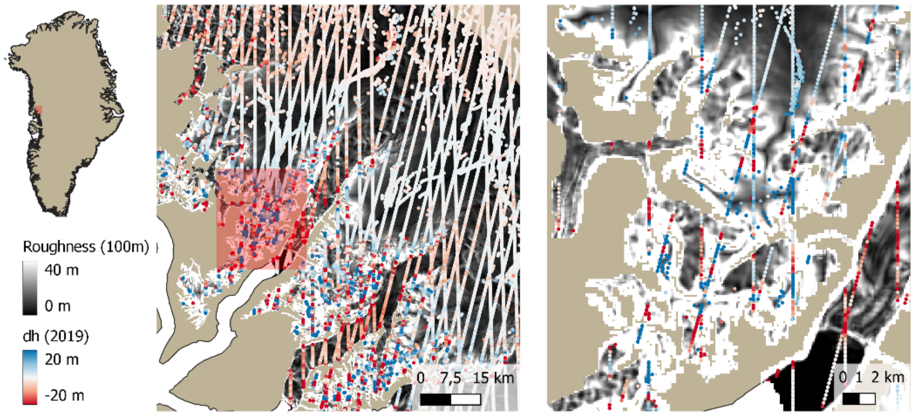
None of these measures can, however, directly be translated into an appropriate grid resolution, as this depends on the spatial representability of the data. Our rationale for selecting a grid resolution of 500 m is described later in this document.

- The authors should elucidate the rationale behind selecting the Kriging interpolation method. Given the steeper terrain at the ice sheet's edge, it is essential to address whether Kriging interpolation can effectively fulfill the task of estimating elevation in such rugged terrains. Clarification on the suitability of the chosen interpolation method, especially in challenging topographical conditions, would contribute to a more comprehensive understanding for the readers.

Our decision to employ Kriging interpolation was based on several factors, including its ability to account for variable uncertainty of observations, its robustness in capturing spatial variability and its ability to provide smooth and continuous surfaces. We have added the following to L316 (in red):

Kriging can account for variable uncertainty of individual data points during the interpolation, and due to its robustness in capturing spatial variability it is a widely used method to interpolate geophysical fields (e.g. (Bamber et al., 2001; MacGregor et al., 2015; Bales et al., 2001)).

In this respect is it very important to note that we do not apply kriging to the original elevation measurements, but to elevation anomalies relative to ArcticDEM. These elevation anomalies display a much smoother field, also around the marginal areas (see figure below), and to large extent this resolves the issue raised by the reviewer of whether kriging can also be used in these areas. Even for areas with very rough surface, such as those often existing at the very outer part of the ice sheet edges, there is consistency in the observed elevation anomalies across length scales of several kms. We have added this argumentation to the paper, along with the figure.



As long as the spatial correlation of the anomaly field can be correctly captured by the variogram, and the field (largely) fulfills the statistical requirement for kriging to be valid, there is no reason that kriging, which offers an optimal unbiased linear interpolation method, should not be a good interpolation method.

- There are reservations regarding the chosen 500m resolution. Satellite altimeters typically yield fewer effective observations at the ice sheet's edge (especially for CS2), and opting for higher resolution may result in a reduced number of usable observation grids. This reduction can potentially impact the interpolation performance. The question arises: Can a DEM heavily reliant on interpolation accurately portray the actual elevations, particularly in the steeper terrains at the ice sheet margins? Additionally, in regions with lower latitudes, where the gaps between satellite orbits are more substantial, concerns linger about ensuring effective spatial interpolation. Addressing these considerations would fortify the paper's discussion on resolution choice and its implications.

As the reviewer correctly points out, there are fewer observations around the very edges of the ice sheet with rough terrain, not least since these are not covered by CS2 data. Again, we are partially saved by using ArcticDEM as a reference DEM, which has high spatial resolution also in these regions. With the anomaly field here often displaying small correlation lengths, the resulting PRODEM tends to heavily rely on the ArcticDEM topography, the latter being slightly adjusted by the median value of the local anomaly field. But, agreed, the PRODEMs do have much larger uncertainty in these areas of very difficult terrain.

As previously mentioned, the spatial density of observations cannot directly be translated into an appropriate grid resolution, which is heavily dependent on the spatial representability of the data.

In terms of the chosen 500 m resolution: We have chosen to use Point Kriging, instead of Block Kriging, to obtain the elevation field, which means that the interpolated field is evaluated at the center of each grid cell. To improve the estimation of the mean value of the grid cell, we subsequently smooth the derived elevation field using a 3x3 mean filter. When taking the Point Kriging approach, it is important that the input data used for the interpolation is representative for the elevation field at the scale of the grid cell. Consider, for instance, the example of using highly resolved data (e.g. using only IS2 data without down sampling) with one observation located close to the center of the grid cell: The center value will then largely reflect the elevation of the nearest data point, regardless of whether this value is representative for the entire grid cell. In smooth areas, this may not be an issue, but in areas of rough terrain, this may not be very accurate. However, if the observation close to the grid cell center is representative of an area of approximately the same size as the grid cell, the interpolated value will correctly represent the mean grid cell value.

As the resolution of the CS2 is on the order of ~500 meters, we choose to down sample the resolution of the IS2 data to similar resolution (we decided to down sample to only 250 m along track to increase the data density), and, consequently, the appropriate resolution of the associated anomaly field must be chosen to be on the same order, namely 500 m.

We have added the following paragraph to the paper to clarify the chosen resolution of the product:

The PRODEMs are constructed using Point Kriging, in which approach the interpolated elevation anomaly field is evaluated at the grid cell centres. For this approach to accurately represent the mean field value within a grid cell, the grid resolution must align with the scale of variability of the observation data. After down-sampling of the IS2 data, both CS2 and IS2 data sets are representative for an area of a few hundred meters, and an appropriate resolution for the anomaly field of the interpolated PRODEMs is therefore on the order of 500 m.

- Notably, the entire DEM is subjected to Kriging interpolation, implying a potentially limited actual coverage area by the altimeter. Therefore, it is imperative for the authors to furnish the proportion of space covered by observational data. Elevation derived from Kriging interpolation might deviate from the altimeter-observed elevation, necessitating consideration of this difference, as the potential errors induced by interpolation are a critical aspect. Particularly, given the uneven spatial distribution of CS2, which may introduce additional uncertainty to Kriging interpolation, it becomes crucial to address how the authors accounted for the impact of observation gaps on the results. Discussing these considerations would enhance the paper's transparency regarding the intricacies of the interpolation process and associated uncertainties.

As previously mentioned, the important factor is the number of observations within the correlation length of the varying elevation anomalies, and we now provide a figure showing this distribution. In areas poorly covered by the observations, the resulting surface is highly reliant on the ArcticDEM surface topography used as reference, and the uncertainty of the interpolated surface is increased.

With proper observation uncertainty estimation and spatial variability model (aspects that we thoroughly investigated) kriging is a way to reconcile the surrounding observations according to their uncertainties and spatial covariance. For it to provide good measurements of the mean field value at grid cells, it is, as mentioned previously, furthermore imperative that the grid cell size is aligned with the spatial representability of the observations. As long as these (and other) conditions for kriging to perform well are fulfilled, the interpolated surface should only deviate from observations within their associated uncertainties.

In terms of the mentioned un-even spatial distribution of CS2, we remark that actually the more “random” distribution of CS2 is more appropriate for interpolation by kriging than the straight lines of IS2 observations (which are far from randomly distributed).

We have added the following paragraph (L333) to clarify the prerequisites and assumptions behind using kriging interpolation:

The accuracy of the interpolated field will be limited if the number of sampled observations (within the effective interpolation radius) is small or not adequately distributed relative to the properties of the field. The accuracy of the PRODEM elevation anomaly field will therefore be reduced in areas of high variability and few observations, such as conditions prevalent at the outer edges of the ice sheet.

- Addressing DEM uncertainty requires a more detailed exposition. While the spatial and temporal uncertainty is elaborated upon, clarification is needed on how the authors define the instrument and geolocation uncertainties mentioned in equation (1). It would be beneficial to delineate the specific contributions of each factor to the overall uncertainty and provide a proportionate breakdown.

We have thoroughly revised the procedure of addressing the CS2 and IS2 uncertainty in the following ways:

1) We now separate the uncertainty component into its spatial and temporal components
2) Elevation differences are now computed as the difference of elevation anomalies (instead of elevations); this approach accounts for a potential slope in the area.
3) We reduce the allowed spatial distance between measurements to 50m (instead of 100m)
4) The spatial component is computed only based on elevation differences at cross-overs acquired less than 15 days apart, in order to limit the impact of potential temporal changes.
5) The spatial uncertainty component for IS2 is calculated in a similar manner to that for CS2: We investigate the relationship between the terrain roughness and spread of elevation anomaly distributions. Based on this relationship, we construct a linear model between roughness and spatial uncertainty.
6) The temporal component is estimated based on a data set of average summer trends of surface elevation (Slater et al, 2021), multiplied with the difference between time of data acquisition and mid-summer.
7) We use cross-overs from all years (previously only data from 2019 was used)

Definitions of the various uncertainty components are provided in L156-163. Due to the way that we derive the spatial uncertainty component, this component includes the geolocation uncertainty, as well as the measurement uncertainty, and this is now clarified in the manuscript:

In the following, we develop separate uncertainty models for the two data sets to estimate representative values for their spatial uncertainty (including effects from geolocation and instrument measurement uncertainty), as well as a common model for the temporal uncertainty.

At small roughness values, we can evaluate the instrument measurement uncertainty from the combined spatial uncertainty as 40cm (CS2) and 8cm (IS2), respectively. These values are now mentioned in the text. We cannot reliably disentangle the geolocation uncertainty from the uncertainty due to spatial representability, as these are intricately intertwined.

We further elaborate on the relative importance of the two terms (spatial vs. temporal):

In general, for both instruments, the temporal uncertainty is notably less important than the spatial uncertainty component, particularly evident in the case of CS2. However, for individual observations the temporal uncertainty component may be non-negligible, with 21% of the IS2 observations having larger temporal than spatial uncertainty. For CS2, this is only the case for 0.4% of the observations.

Moreover, in discussing the uncertainty of CS2/IS2 elevations, the term 'crossovers' might be prone to misunderstanding.

We have gone through the manuscript and changed the word “cross-overs”. We now refer to it as “observations near intersecting satellite tracks”.

Consider utilizing track-based crossover analysis for data validation, ensuring a clearer understanding of whether the differences approach zero. This adjustment would enhance precision in evaluating the uncertainty associated with CS2/IS2 elevation comparisons.

We have followed the reviewer’s suggestion, and tested the resulting uncertainties at intersecting satellite tracks to check their consistency: Z-scores were computed from the observed elevation differences divided by the total uncertainty of the elevation differences. Despite heavy tails in the resulting distribution of z-scores (indicating the existence of outliers not fully captured by the estimated uncertainty) the distributions were found to adequately approximate a standard normal distribution.

For some marginal areas, the temporal component was not well defined due to lack of data, and for these areas, we used the average value of trend of summer elevation across summer to estimate the temporal representability. The z-score distributions best approximate a normal distribution for cross-overs located in areas where the temporal component was best defined. This issue is particularly important for the IS2 observations, which to higher degree are located near to the ice sheet margin, where the temporal component is poorly constrained.

We have added the following paragraph to the paper:

The derived observation uncertainties were subsequently tested for consistency with the observed elevation differences near intersecting satellite tracks. Z-scores were computed from the observed elevation differences divided by their total uncertainty. Despite heavy tails in the resulting distribution of z-scores, indicating the existence of outliers not fully captured by the estimated uncertainty, the distributions were found to adequately approximate a standard normal distribution.

- The evaluation of the proposed DEM solely through cross-validation may be deemed insufficient. It is recommended that the authors endeavor to gather actual measurements, such as airborne datasets, for Greenland elevation. This would facilitate an independent evaluation of the generated DEM, offering accuracy metrics under various terrain conditions. Such an approach would provide more robust and precise evaluation information, enhancing the credibility of the study's findings.

We agree that it would be great to also validate the PRODEMs against independent data. However, airborne data sets that fulfill the requirements of being acquired during the summer months of 2019-2022 are exceedingly sparse. To our knowledge, only one such campaign exists, namely the CryoVEx Summer 2019 campaign, and it was directed towards obtaining data over sea ice rather than land ice. One track does, however, cover a small section of the very margin of the north-eastern part of the ice sheet. We will investigate whether this track contains sufficient land ice data for independent validation (which in any case will be very limited), and if so, we will report the results in the manuscript.

As this data set in any case is too limited to offer accurate evaluation metrics under diverse terrain conditions, we instead stratify the prediction errors from the cross-validation exercise, and consider the error distributions as function of increasing latitude, elevation, roughness, aspect, ArcticDEM uncertainty and summer elevation trends. We have added a section on this to the manuscript, along with the figure provided below.

In short, the analysis allows us to identify troublesome areas, where one should be careful regarding the reported uncertainties in the PRODEMs, these being characterized as having s_{MAD} values of the normalized prediction uncertainties significantly above one, which we here take to be larger than 1.2. In order of importance, these areas are characterized by having: Roughness above 13 m, ArcticDEM MAD values above 0.4 m, and/or elevations below 300 m. These conditions span respectively 25, 22, and 4% of the area, and largely cover the same areas along the very margin of the ice sheet.

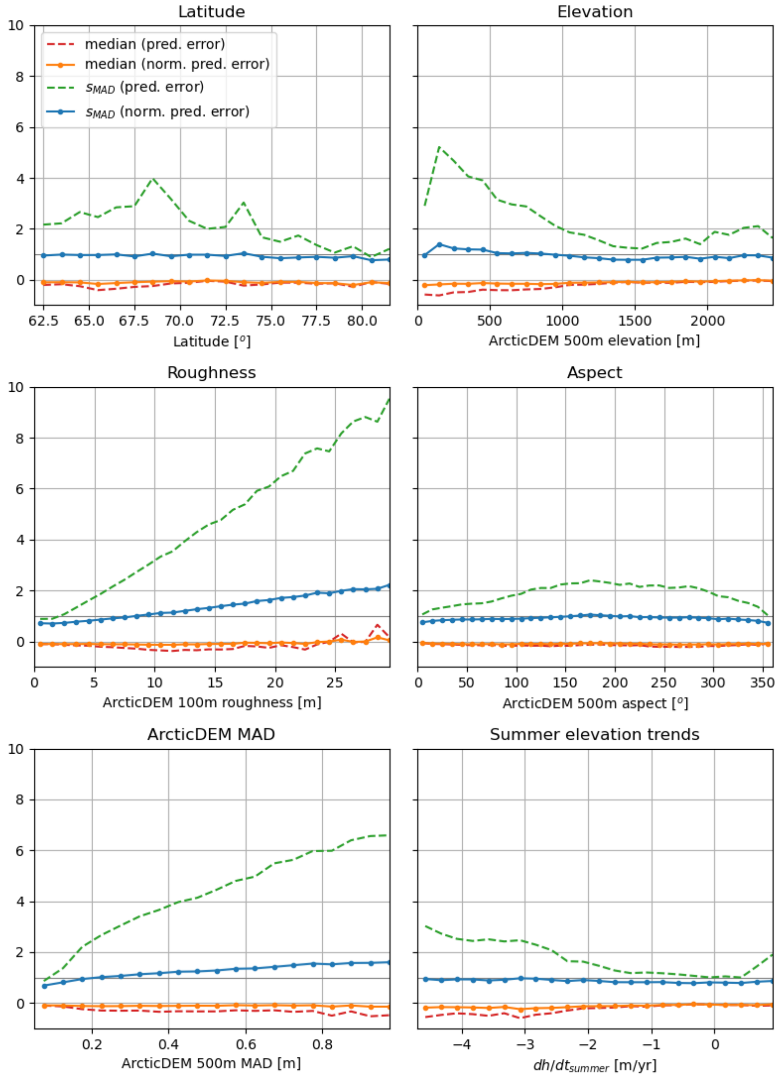
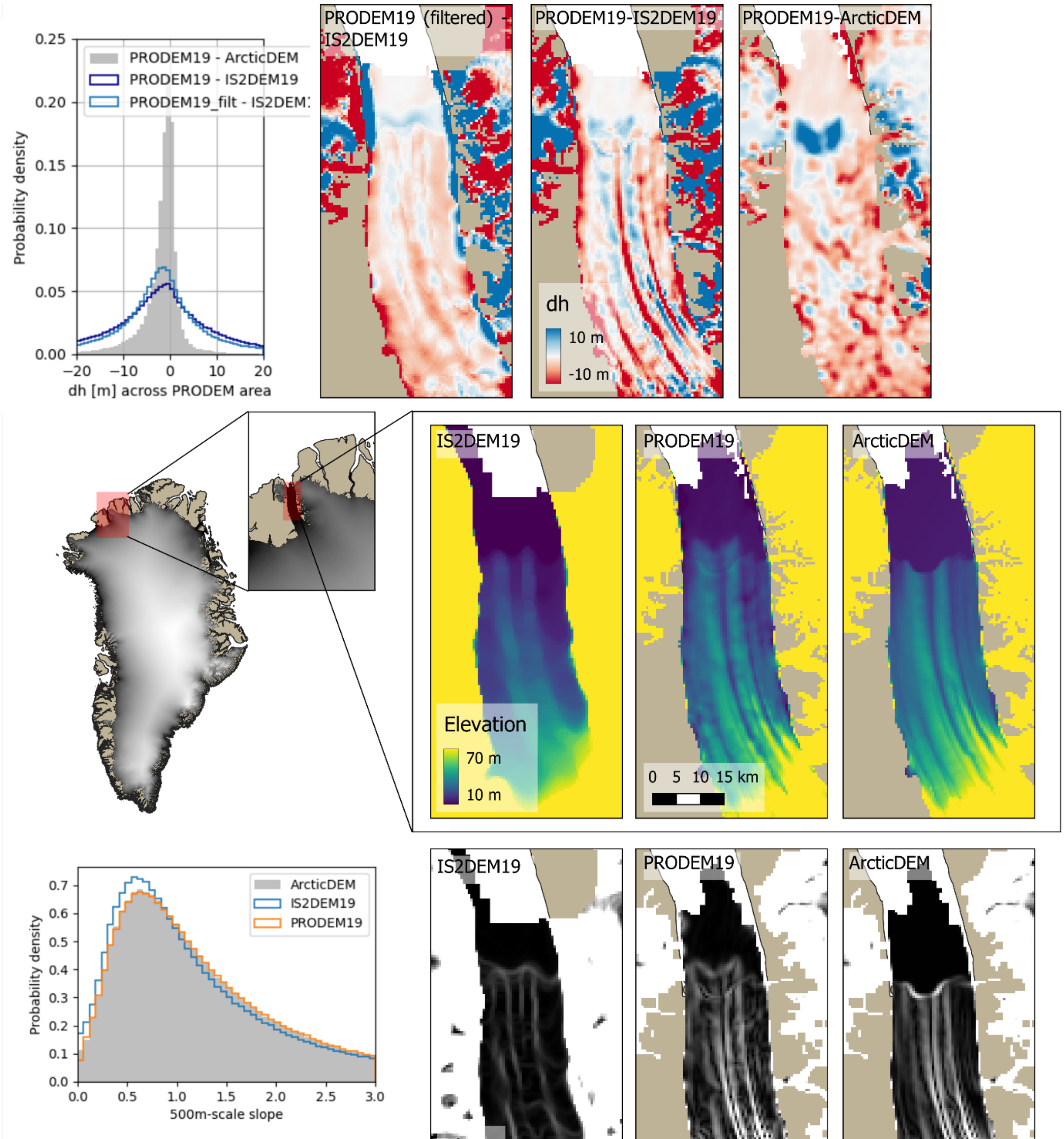


Figure 11: Descriptive measures (median and s_{MAD}) of the distribution of prediction errors, both raw (dashed lines) and normalized errors (solid lines), when stratified relative to various location and terrain properties that may impact the PRODEM product performance. Bins are created to cover the 5 to 95 percentiles of each data set. The ArcticDEM MAD value is computed as the median absolute deviation of the DEMs contributing to ArcticDEM at a given location, and therefore provides an uncertainty estimate of ArcticDEM. Data based on prediction errors from PRODEM19.

- In comparing with other DEMs, presenting a distribution map alone may lack depth. Additional comprehensive comparisons, such as showcasing the spatial distribution or numerical values of elevation differences, and establishing relationships between these differences and slope and roughness, are warranted. To enhance the display of the DEM, the authors might consider incorporating a slope map or shaded relief map. These additions would provide a more nuanced evaluation of the DEM data and contribute to a thorough understanding of its effectiveness.

We have added a slope map to Figure 9 (now Figure 10), along with elevation difference maps relative to IS2DEM19 and ArcticDEM, and histograms of the associated distributions, see the updated figure below. Since the PRODEMs have much more structure (as evident, for instance, from the slope map), we also compare it to a smoothed version of IS2DEM19.

As the three DEMs are not representing the same period of time, and the IS2DEM19 is of lower actual resolution than the others, we cannot use the DEM comparisons for validation of PRODEM19. We therefore do not consider it relevant to dive further into the investigation of relationships between elevation differences and e.g. slope and roughness.



Slope in PRODEM area	Mean	Median
IS2DEM19	1.65	1.05
PRODEM19	1.78	1.16
ArcticDEM	1.78	1.15

Figure 10: A comparison of the three elevation fields: IS2DEM19, PRODEM19, and ArcticDEM across Petermann Gletsjer. Coverage of the three DEMs varies slightly across the region. The top panel shows elevation differences between the three DEMs, including a filtered version of PRODEM19, produced using a Gaussian filter with 5 pixels radius and relative standard deviation of 50%. The middle panel shows the elevation fields across Petermann Gletsjer. The lower panel compares the slope distribution for the three elevation models. a) A histogram of the distribution of elevation differences across the PRODEM area, and b-d) maps showing the spatial distribution across Petermann Gletsjer. e-g) The elevation fields across Petermann Gletsjer. h) Distribution of slopes across the PRODEM area, i) mean and median values of the slope distribution, and j-l) the spatial slope distribution across Petermann Gletsjer.

- Consideration should be given to the potential benefits of updating the DEM with the new ArcticDEM mosaic v4, as it has the potential to significantly enhance DEM accuracy and reduce differences in dh (as observed in Figure 10). Such an update could serve as a viable solution to the issues identified and contribute to an overall improvement in the quality of the DEM.

The PRODEMs have been updated used the new ArcticDEM v4.1 mosaic. We note that we observe a checkerboard pattern in dh also when using the new version of the ArcticDEM as reference (although less pronounced than previously, and appearing more significant in other areas). The updated Figure 10 (now Figure 11), from a different location, is provided below.

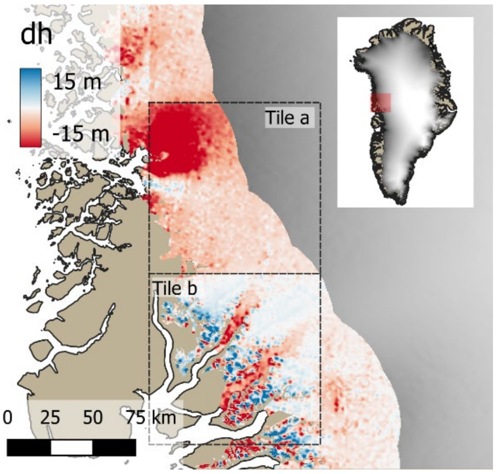


Figure 11: Elevation anomalies (here for PRODEM19) relative to ArcticDEM display a checkerboard pattern due to the way ArcticDEM was constructed. Two ArcticDEM tiles of 100x100 km are indicated.

Specific comments:

(p: page, l: line)

Section Introduction: The current paragraph lacks substantial content. To enhance its depth, it is recommended to incorporate the significance of ice sheet elevation monitoring. Elaborating on the importance and implications of monitoring ice sheet elevation would provide readers with a clearer understanding of the broader context and relevance of the study.

We have revised this section so that it to higher degree focuses on the importance of ice sheet elevation monitoring.

p2l44, Supplementing the existing DEM datasets and providing commentary on them is advised. This additional information will not only enhance the comprehensiveness of the study but also offer valuable insights and context for readers evaluating the significance and limitations of the proposed DEM datasets.

In the introduction to the revised manuscript, we now also describe existing DEM datasets, and describe how the PRODEMs will offer new insights into ice sheet monitoring efforts.

p2l53, CS2 is equipped with a Ku-band radar altimeter, which is highly sensitive to moisture changes. Notably, surface melting can introduce significant interference to the echo signal, posing challenges in obtaining accurate elevations during such periods. While the author's emphasis is on capturing surface elevations during summer, it is worth considering the possibility of melting snow, even in areas designated as snow-free. Clarification on the impact of any residual melting snow on elevation data would enhance the understanding of potential influencing factors.

The effect of surface melting on the radar signal properties is primarily relevant in the firn zone (i.e. outside the PRODEM area), where surface melting will affect the structure of the near-surface firn, this causing significant biases in the elevation estimates.

Indeed, the ice sheet summer surface in the marginal areas is inhomogeneous (experiencing surface melting surface, the existence of supraglacial lakes, may at times be partly snow-covered etc.), which may impact the altimetric measurements of the ice sheet surface, as it may make the surface less well-defined and thereby more difficult to track in the radar waveform. However, we consider the potential impact of this on the elevation estimates to be well within the uncertainties of the CS2 observations.

To eliminate bad CS2 measurements (due to this factor as well as e.g., incorrect geolocation), we filter the altimetry observations prior to interpolation.

P2l69, what distinguishes the Baseline-D and E datasets, and is there potential for this difference to impact the consistency of the CS2 data? Clarifying the distinctions between these datasets and addressing their potential influence on CS2 data consistency would contribute to a more comprehensive understanding of the data sources employed in the study.

Since submission of the paper, the baseline-E data set has become available for the entire period, and we have therefore produced a new version of the PRODEMs using the baseline-E for the entire period, so the CS2 data set is now fully consistent.

P2l74, in contrast to IS2, CS2 exhibits a relatively limited dataset with irregular coverage, further reduced through quality control measures. It is imperative to elucidate the specific spatial distribution of CS2 data after quality control. Any potential uneven distribution in the data may raise concerns about its impact on subsequent analyses and differences. Addressing these considerations would provide a clearer understanding of the reliability and potential biases associated with the quality-controlled CS2 dataset.

The CS2 datasets constitute ~25% of the full dataset applied to construct the PRODEMs (L93).

For best performance of the kriging interpolation, the observations should be distributed in an adequately balanced way with respect to the properties of the field to be interpolated. In this respect, the CS2 data is actually much better distributed than the IS2 data, which are acquired with a far from random distribution. Indeed, straight-line arrangements may give rise to biases and inaccuracies in the kriging predictions, as it may lead to overfitting in densely sampled regions. The only way to deal with this, however, would be to entirely remove parts of the IS2 data, which is not a desired solution.

P2l75, Significant disparities exist between LRM data and SARln data. The authors should expound on their considerations regarding the consistency between these two datasets. Providing insights into the strategies employed to address or account for these differences would contribute to a more thorough understanding of the data integration process and potential impacts on the study's outcomes.

We only include LRM data for a very tiny region. The number of datapoints included consists of 0.1% of the total data set, and does not have any major impact on the study. We decided to include it in order to provide CS2 data for the entire 50km of the margin.

P3l83, A higher sampling density often correlates with improved elevation simulation results. In the context mentioned, it would be beneficial to clarify the purpose of avoiding oversampling and elaborate on the potential consequences associated with it. Understanding the reasoning behind this choice and its potential impacts would enhance transparency regarding the sampling strategy employed in the study.

Ideally, data points should exhibit spatial variability representative of the underlying field, capturing both local fluctuations and broader trends. Straight-line arrangements can introduce biases or inaccuracies into the kriging estimates, as it may lead to overfitting in densely sampled regions, where local variations are excessively emphasized, while neglecting broader trends. A balanced distribution of data, capturing both local and global variability is therefore essential for robust kriging interpolation.

The straight-line arrangement of the IS2 data cannot be changed. However, to obtain a more balanced distribution we remove data from the weak IS2 beams, which are located right next to the strong beams. Retaining these data would provide no new information, but lead to excessive oversampling.

P3l88, However, such downscaling might diminish the inherent advantages of IS2. Additionally, in the case of CS2 data, it is essential to elucidate how the authors addressed the potential impact of observation resolution. Clarifying the strategies employed to mitigate the effects of observation resolution on CS2 data would contribute to a more comprehensive understanding of the data processing methodology.

Agreed, to some extent the downscaling diminishes the inherent advantages of IS2. It was, however, necessary to do this downscaling in order to assimilate the data set with the CS2 data, see also our answer to one of the previous comments. We do not see any need to mitigate the lower resolution of the CS2 data, as we do not attempt to create the elevation anomaly field in higher resolution that the approximate resolution of the input data.

P3l93, it is evident that IS2 data plays a more central role in DEM generation. Consequently, it raises the question: What specific role does CS2 serve in this context? Considering the non-uniform distribution of CS2, there arises a concern about the potential introduction of uncertainty into elevation estimations. Addressing the distinct contributions and potential uncertainties associated with CS2 data would enhance clarity regarding its significance in the overall DEM generation process.

The CS2 data provides 25% of the total elevation data set, which in our opinion is not a negligible portion.

We have in the new version of the manuscript improved the uncertainty estimation of both data sets.

P4l118, the impact of CS2 signal penetration depth in snow is not explicitly addressed in this section. Providing insights into how the authors considered and accounted for this factor would enhance the completeness of the discussion.

This is not relevant since the PRODEMs are based on data from the ice sheet margin during summer, and therefore largely covers only the snow-free zone.

P4l124, the reason for choosing a 500m resolution is not explicitly stated. Including the rationale behind this resolution choice would provide clarity and context for readers seeking to understand the decision-making process.

This is now more clearly stated:

The PRODEMs are constructed using Point Kriging, in which approach the interpolated elevation anomaly field is evaluated at the grid cell centres. For this approach to accurately represent the mean field value within a grid cell, the grid resolution must align with the scale of variability of the observation data. After downsampling of the IS2 data, both CS2 and IS2 data sets are representative for an area of a few hundred meters, and an appropriate resolution for the anomaly field of the interpolated PRODEMs is therefore on the order of 500 m.

P5l146, the uncertainty associated with Kriging interpolation is not specified. Offering details on the uncertainties inherent in the Kriging interpolation method would contribute to a more thorough understanding of the potential limitations in the DEM generation process.

The uncertainty associated with Kriging interpolation is calculated based on the equation provided in L436.

P7l236, the stated measurement precision applies specifically to the interior areas of Antarctica. It's crucial to note that for steep terrains, the accuracy is expected to be lower. Emphasizing this distinction would provide a more nuanced understanding of the precision associated with different topographical features.

As correctly pointed out, the stated measurement precision specifically applies to the flat interior areas of Antarctica, and we have added this to the sentence (the addition in red):

The measurement precision of the IS2 ATL06 product has been documented to be as small as 9cm over the flat interior part of Antarctica.

Note, however, that we do not use this value in the subsequent analysis, but instead use elevation differences from crossing satellite tracks to provide an estimated uncertainty for the individual observations depending (primarily) on terrain roughness. In areas of low roughness, we estimate an instrument measurement uncertainty of 8cm, which is in accordance with the results from the interior of Antarctica.

Section 4.1, The calculation method for DEM uncertainty, particularly how the authors integrated data from the two altimeter datasets, is not clearly presented in this section. Including details on the calculation method would enhance transparency and facilitate a more comprehensive assessment of the uncertainty analysis.

We have revised the uncertainty model for the individual observations, so that the spatial (including geolocation and instrument measurement uncertainty) and temporal components now are separately defined. The total uncertainty on an observation is found by adding the two uncertainty components in quadrature. The text in this section has been updated accordingly.

The DEM uncertainty from the kriging interpolation is calculated based on equation 5 (L436), which allows to account for the different uncertainty of the surrounding observations.

P8l279, Notably, there appears to be a relatively large elevation difference between CS2 and IS2. It is pertinent to inquire whether the authors have considered any correction methods to address or minimize this discrepancy. A discussion on potential strategies or considerations in correcting the observed elevation differences would contribute to a more thorough evaluation of the data integration process.

Indeed, in some places we observe quite some elevation differences between closely-spaced (50m) CS2 and IS2 observations. It is worth noting that the outliers in this distribution primarily arise from satellite track crossings in areas of high roughness, where the terrain may vary significantly within a short distance, and a substantial elevation difference therefore is to be expected. Investigating the spread of the distribution as function of terrain roughness (in a similar manner to the analysis performed for estimating the spatial uncertainty of CS2 and IS2), we find a linear increase in dispersion (MAD) of the elevation difference distribution with roughness (see the figure below). At small roughness values (roughness < 2), the dispersion of the distribution is described by a MAD value of 0.48m (average value), corresponding to a 1σ=0.50m. This number should be seen in relation to the spatial uncertainty of the CS2 and IS2 observations at small roughness values, which is 0.39m (CS2) and 0.08 (IS2), respectively. Combining the two in quadrature, we therefore expect an uncertainty of their associated elevation differences of 0.40m. The spread of the elevation difference distribution can therefore largely be attributed the individual uncertainty of the two observations.

In regard to estimating any bias, we must instead consider the mean or median of the distribution. Considering the latter to provide the best estimate for the bias, we notice that its value is lowest in areas of low roughness (-5 cm). This is despite these smooth areas often being located at high-elevation central areas, which may periodically be covered by snow during the designated summer period, which we expect to lead to some, minor, differences in the measured elevation from the two satellites. We therefore consider 5cm to be an upper value for the potential bias.

It is worth noting that this value shows significant spatial variability; magnitude as well as sign differs when conducting the analysis on drainage basin scale: The median value ranges from -0.26 to +0.40 depending on drainage basin, although it should be noted that the analysis for some basins is hampered by poor statistics.

We therefore limit ourselves to note that the maximum potential bias value of 5cm bias is well within the uncertainties of the CS2 datapoints. In accordance with previous literature, we therefore cannot identify any significant bias between the two sets of satellite observations, and consequently we cannot justify to make any bias correction.

We have revised Section 4.2 to include more information on the bias estimation procedure, and a new figure (replacing Fig. 3a).

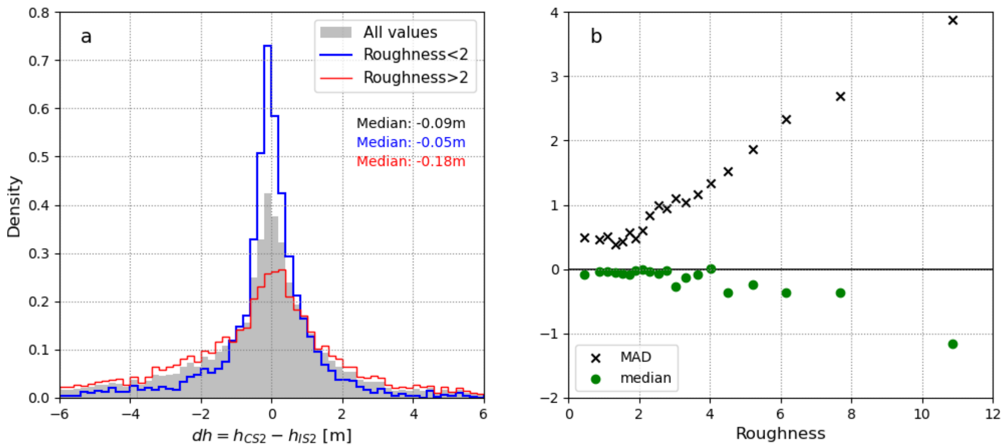


Figure 4: a) Distribution of elevation differences, computed as $dh = h_{cs2} - h_{is2}$, near intersecting CS2 and IS2 satellite tracks, when imposing strict limits in terms of maximum distance and time between acquisitions. b) Evolution of median and MAD of the resulting elevation difference distribution as function of terrain roughness. Data from the PRODEM area during summers 2019 through 2022.

P9l292, it would be beneficial to expound on the specific steps involved in DETREND. Additionally, clarification on the basis for employing linear interpolation and whether it introduces any uncertainties would be useful.

We decided to do the detrending in the simplest possible way using linear interpolation in the ArcticDEM grid. We note that linear interpolation tends to conserve the general trend of data between grid points, making it suitable for interpolating values within a regular grid, and we have no reason to believe that a more sophisticated method would perform any better.

P9l298, the term "ELEVATION ANOMALIES" requires clarification, and providing details on the proportion of outliers would offer a more precise evaluation of the data characteristics.

We have added to following (in red) to L293:

Prior to interpolation, the satellite altimetry data are detrended by subtracting a reference DEM (ArcticDEM) from each altimetry point measurement using linear interpolation, and the resulting data set is termed elevation anomalies.

Details on the proportion of outliers are provided in L300-301.

Section 5.3, Explicitly mentioning the software used for Kriging interpolation and listing relevant parameter settings would provide essential information for readers seeking to replicate or understand the interpolation methodology.

We only used the software GSTools to construct the experimental variograms, and all relevant parameter settings are provided in the manuscript (L344-347). For the revised version, we have replaced this software with Scikit-Stat, which allows us to use the Dowd estimator instead of the Cressie estimator (as suggested by reviewer 1).

Section7.1, I strongly recommend incorporating airborne elevation data into the DEM evaluation process. While cross-validation results are valuable, utilizing airborne data for evaluation would provide an additional and crucial perspective on accuracy. This approach ensures a more comprehensive assessment and strengthens the overall validity of the DEM.

We will investigate whether this can be done based on the very limited airborne elevation data available. See comment above.