Replies to reviewer comments by Romain Hugonnet

We thank the reviewer very much for their appraisal of our manuscript and DEM product, which we truly appreciate.

We also appreciate the very constructive suggestions for improvements, and with these in mind we have clarified and adjusted the procedure as described below. 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 initially submitted, several of the input datasets have been updated, and we have now included these:

- 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 the paper.

Statement of expertise

I have expertise in the aspects of this paper associated with:

- Remote sensing of surface elevation data for glaciology,
- Laser altimetry with ICESat-2,
- Geostatistical methods used for data fusion, in particular variography and kriging,
- Uncertainty analysis in a spatiotemporal context for the predicted elevations.

Would be nice to complement with another reviewer being specifically expert in:

- Radar altimetry with CryoSat-2 (though I do have some).

General comment

In their very well-written manuscript, the authors present a rigorously derived dataset of annual (2019-ongoing) gridded surface elevation at 500 m resolution for the Greenland Ice Sheet, predicted from spatiotemporally sparse ICESat-2 and CryoSat-2 elevations.

The study is a truly a great piece of work, and I commend the authors on their attention to detail at all steps of their analysis to produce a robust product for the community. My core expertise is mathematics and DEMs, in particular spatial statistics and uncertainty analysis, and this study is a rare sight in our field in that aspect, and I think the way forward so that other scientists and users can know how reliable a data product really is.

I still have several comments for the authors, which I think could improve some parts of their analysis substantially. Some might be a bit of work, but are important and should be OK given the author’s knowledge in geostatistics, and others are optional.

Sorry if I cite my own work quite a bit in this review, there are not that many other studies looking at these specific problems in detail (geostats + uncertainty quantification + DEMs), and I hope it will be helpful to the authors!

Main comment 1: Refine the modelling of heteroscedasticity of the predicted elevation

Or "variability in error". This is something the authors have addressed quite a bit in the text and their estimates, in particular in Section 4.1 by constraining the variability in elevation error (CS2 - CS2) with surface roughness (for CS2; Figure 2b), and using the variability of 20 m IS2 points in the 250 m area.

However, Figure 8 shows that this modelling of the variability in error (including subsequent propagation with kriging) is not completely satisfactory:

- Too large predicted errors for the center of distribution,
- Yet does not capture the full range of variability (tails are quite significantly larger than a normal distribution).

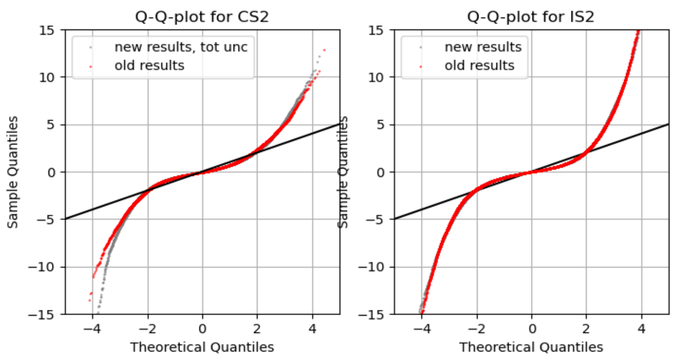
Here, I recommend the authors several pists that should improve their reported error.

As the reviewer correctly points out, the uncertainty of the resulting elevation estimates showed large spatial variability (high heteroscedasticity) as expected due to varying data density and spatially-changing terrain. Despite our efforts to account for this, the results of our validation (Fig. 8) did indeed show that our kriging uncertainties did not fully capture the true variability of the deviations: In most areas, the uncertainties were slightly conservatively estimated, but a significantly higher number of outliers than expected existed.

In line with the changes suggested by the reviewer to improve this (see below), we have therefore meticulously adjusted the uncertainty calculation of the CS2 and IS2 measurements. We have further changed the allowed variogram nugget (also suggested by the reviewer; more about this later), which also influences the spatial structure of derived uncertainties.

We also now normalize the prediction errors relative to the total uncertainty of the prediction and the estimated observation uncertainty for the validation data. This causes the normalized error distribution to slightly improve – in particularly for the CS2 data, which are associated with larger observation uncertainties.

Despite all these changes, our cross-validation shows that the resulting error distributions from the PRODEM validation are very similar to previously. A figure showing a comparison of the associated Q-Q plots for the LOOCV differences is provided below, along with an updated version of Figure 8 (now Fig 9). The figure comparing Q-Q plots for the previous and revised PRODEM versions is only for illustrating the differences to the reviewer, and it is not included in the revised paper.



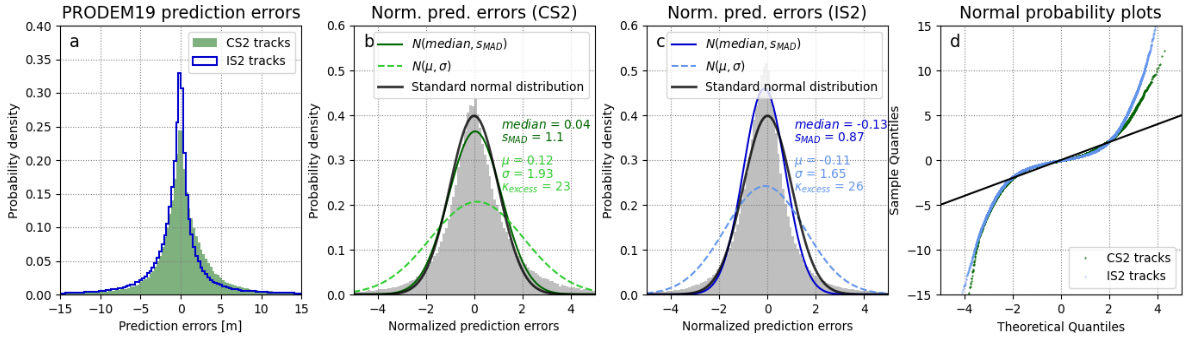


Figure 9: Distribution of PRODEM19 prediction errors of the elevation anomaly after removal of either one track of CS2 data, or one track (consisting of three beams) of IS2 data. a) Distribution of prediction errors prior to normalization. b, c) Distribution of prediction errors when normalized relative to the estimated 1σ uncertainty of the interpolated field and the observation uncertainty of the validation data. For both distributions are shown the associated normal distributions based on calculating the median and normalized MAD of the prediction errors (solid lines) as well as using the directly calculated mean and standard deviation (dashed lines). The standard normal distribution (solid black line) is shown for comparison. d) Normal probability plots comparing the quantiles of the prediction error distributions with those of the standard normal distribution. Deviations from a straight line (black) indicate departures from normality.

In other words, despite the enhancements in uncertainty estimation, some limitations persist within the reported error estimates. In the manuscript, L526-531 has been rewritten, so that this is clearly stated to ensure that the reader is aware of this issue. The section now reads as follows:

The analysis, however, also identifies some limitations with the reported uncertainties: All distributions display higher central peaks than that of a standard normal distribution and significantly heavier tails (higher kurtosis). The central clustering of prediction errors implies that most elevation uncertainties tend to be slightly over-estimated. The heavier tails, on the other hand, indicates that the reported uncertainties do not capture the full range of variability, and, consequently, that errors in elevation estimate substantially larger than the prediction uncertainty are much more likely than theoretically expected.

First, some small but important changes to avoid over-estimating the center distribution:

1. Consider using a Q-Q plot in Figure 8 (or add one), which will represent the tail differences in a clearer fashion,
   Q-Q plots have been added to Figure 8 (now Figure 9); see the updated figure above.

2. Avoid over-estimating the CS2 measurement error by dividing by square-root of 2 as the error source is combined twice in the cross-validation exercise (e.g., Equation 8, Hugonnet et al. (2022)).
   This has been corrected, and the following sentence has been added to the manuscript (L192):
   As the distribution of elevation differences integrates the error sources from both elevation measurements, we derive an approximate measure of the 1σ uncertainty of the CS2 observations as:
   $$\sigma_{spatial}^{CS2} = \frac{s_{MAD}}{\sqrt{2}} = \frac{1.4826}{\sqrt{2}} * MAD$$

3. Avoid over-estimating the IS2 measurement by taking a simple spread: separate segments will have largely uncorrelated errors, so you need to compute a standard error, i.e. divide by sqrt(N) with N the non-overlapping ICESat-2 segments.
   We have revised the approach used for computing the uncertainties associated with the IS2 measurements, so that it is now aligned with the method utilized for the CS2 data (more details below). Notably, we no longer rely on the spread of IS2 measurement along the 250m segments as an estimate for the terrain roughness, but instead use the terrain roughness calculated based

on ArcticDEM. In this respect, it is worth mentioning that the two roughness measures are highly correlated (correlation coefficient=0.8).

Then, to improve their modelled variability in error:

4. Have a consistent spatial variability dependency with terrain roughness for both IS2 and CS2 as it should be universal (e.g., Hugonnet et al. (2022) for slope + curv), you could apply the same scheme to both CS2 and IS2 (it sounds you did something similar for IS2 on L260-261? These statements are a bit unclear, consider explaining in more details, or adding a Figure if it tells something about the error!).

Inspired by the comments by the reviewer (above and below), we have made the following changes to the uncertainty calculations:

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 satellite tracks 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. We take this to be an approximate value for σ_{temp}.
7) We use cross-overs from all years (previously only data from 2019 was used)

Regarding 4), it is worth mentioning that the relationship between terrain roughness and spatial variability is found to differ significantly between the CS2 and IS2 data sets, and we can therefore refute the proposition that a universal relationship should exist between the two.

For IS2, we find a linear relationship between spatial variability and terrain roughness, whereas for CS2 the uncertainty is linearly correlated with the logarithm to terrain roughness (see the updated Figure 2 below, which now includes the IS2 uncertainty model). So, while we now apply a similar scheme to derive the CS2 and IS2 spatial uncertainties, the resulting two uncertainty models differ significantly, with the associated uncertainties consistently being largest for CS2.

We interpret the leveling off of the CS2 uncertainties at high roughness values as being due to the preferential sampling of CS2 at high points in the landscape (POCA), which will lead to a general under-estimation of the observation uncertainty due to spatial variability, particularly in highly irregular terrain. The IS2 instrument, on the other hand, does not display preferential sampling. We observe that the CS2 measurement variability is consistently larger for CS2 than for IS2 at all roughness values.

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. We note that with the new uncertainty estimates, the total uncertainties are generally slightly lower than before.

Sections 4.1.1-4.1.2 have been thoroughly revised accordingly. Figure 2 is expanded to also include the IS2 spatial uncertainty model, and a new Figure 3 is included showing the distribution of the individual uncertainty components for the input data. See the updated figures below.

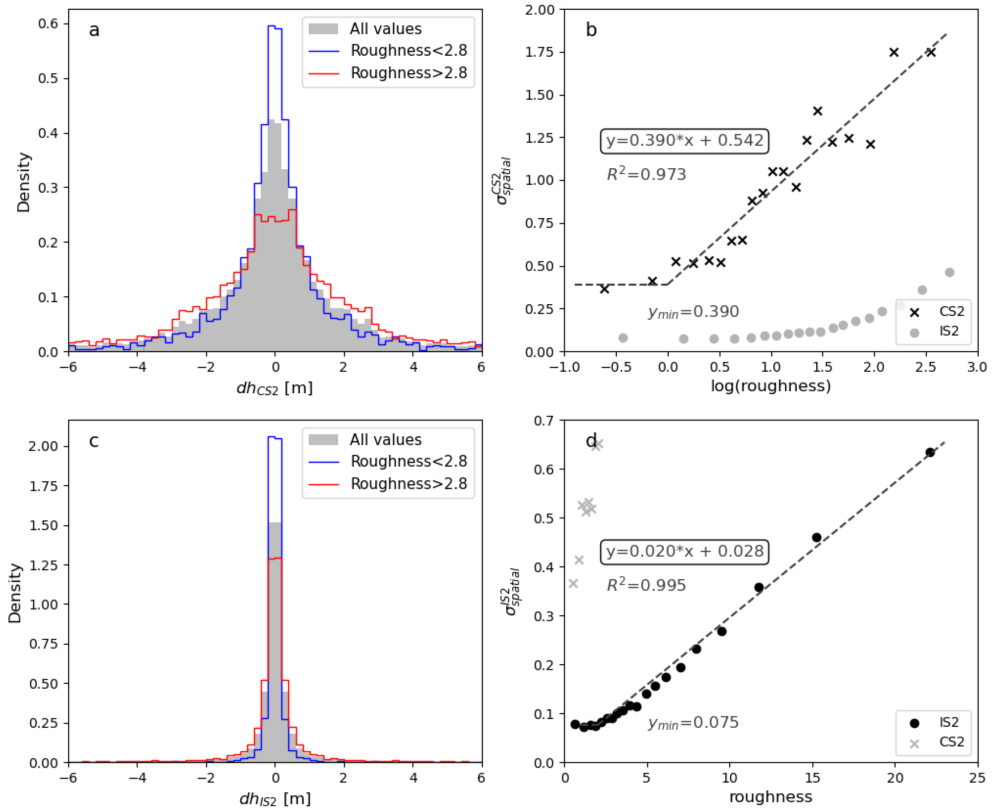


Figure 2: a, c) Distribution of measured elevation differences (dh) at all CS2 (a) and IS2 (c) satellite track near-cross-over locations within the study area (grey), and cross overs in areas with roughness higher (red) or lower (blue) than the median roughness value of 2.8, respectively. b, d) For CS2 (b), a linear relationship (dashed grey line) exists between the logarithm to the local 100m-scale roughness value and the estimated spatial uncertainty based on the dispersion of associated elevation differences (Eq. 1). For IS2 (d), on the other hand, we observe a linear correlation directly between the roughness and estimated spatial uncertainty. Data from summers 2019 through 2022.

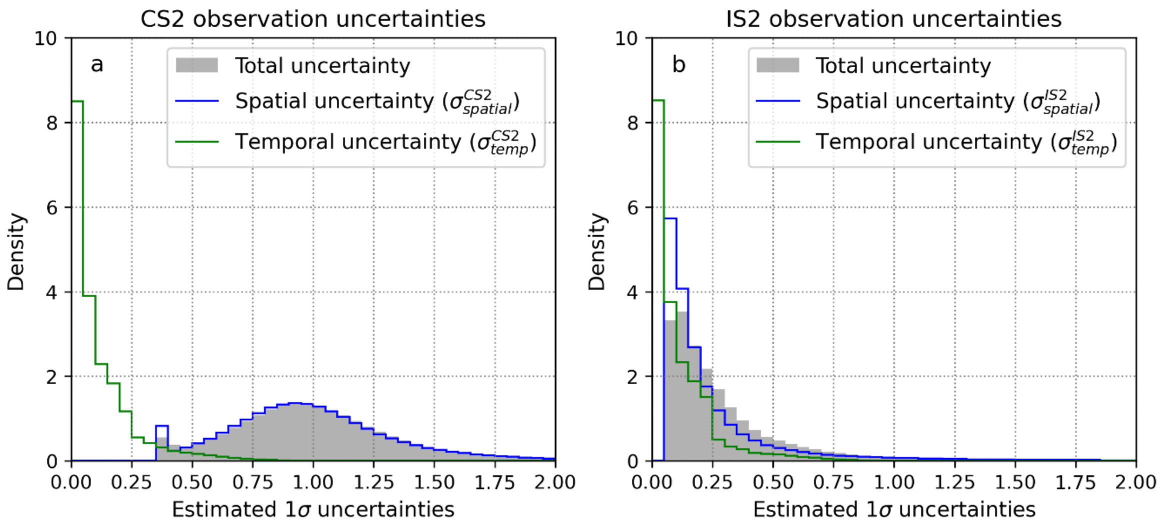


Figure 3: Distribution of derived total uncertainties (grey), along with the individual spatial (blue) and temporal (green) uncertainty components for the CS2 (a) and IS2 (b) altimetry data, respectively. Data from summers 2019 through 2022.

8) Better constrain the variability of CS2/IS2 error (which might explain Fig. 8 entirely).

We have now improved the uncertainty model of the CS2 and IS2 data points, as described above. However, despite the enhancements in uncertainty estimation, limitations persist within the reported error estimates, as visible from Figure 8 (now Figure 9; see the updated figure in the previous comment).

We suspect that the main reason for this discrepancy lies within our estimation of the spatial uncertainty component: When constructing the spatial uncertainty component from temporally close cross-overs, the resulting elevation difference distributions (which we assume describe the observation uncertainty) are not gaussian distributed, but have much heavier tails. In other words, outliers are much more likely to exist that we expect from a gaussian distribution. Nevertheless, as kriging requires gaussian data uncertainties, we approximate these distributions as gaussian. Consequently, we are not properly assigning these extreme data observation uncertainties.

I was happy seeing Equation 1, and then disappointed that the sources were not actually individually refined. I do understand that separating the sources is not feasible with the available data, but it is still possible to separate the variability in error mixed within your single error proxy (here, your cross-over differences).

The authors have 40,000 CS2 cross-over and 60,000 IS2 ones at disposal, a good sample to understand different variabilities. For the temporal variability: the authors could bin these differences with the temporal lag to your reconciled summer date represented by your predicted elevation (middle of your June 1st to September 30th period? This "middle reported date" absolutely needs to be reported in the paper). This should allow to constrain an average variability with the time lag (e.g., Hugonnet et al. (2021), Extended Data Fig. 5ab, which also includies spatial correlations). However, I suspect this variability will likely depends on thinning at the surface at each location, not only on the time lag. You would have to use an estimate of long-term dhdt at the surface (doesn't need to be super resolved, e.g., Smith et al., 2020, or your own product for multiple years) to combine this into a 2-D (time lag + dhdt) binning. This might explain most of your temporal (and general) error variability.

As suggested by the reviewer, we now separate the uncertainty component into its spatial and temporal components.

We investigated several approaches for estimating the uncertainty caused by the temporal variability, but it is hard to disentangle from the spatial variability. The issue is complicated by the following factors:

- The temporal variability tends to be largest at the margin, where also the spatial variability is largest. The two are therefore not independent, and the spatial component is generally the most important factor. Only considering areas with low spatial variability, i.e. primarily central areas, will not accurately capture the temporal variability outside this area.
- The surface elevation displays a trend during summer, the general trend of which has been mapped by Slater et al. (2021). This seasonal trend is very different from the long-term dh/dt trends as in e.g. Smith et al, 2020.
- At any location the surface is not likely to display a constant linear trend throughout the summer, with periods of melting being interrupted by periods of increasing surface elevations due to e.g., new snowfall events.

We attempted the suggested approach of binning with regard to time lag in a similar manner as in Hugonnet et al. (2021), but even when performing a 2D binning by also dividing into regions of similar dh/dt (we tested both using summer trends or long-term trends) the results were inconclusive: The variability was generally constant with time lag, instead of increasing with time lag as expected. We suspect that this is caused by our binning of time lags from different time periods during the summer, and that the surface does not display a steady evolution during this period.

We therefore refrain from estimating an average variability with time lag directly from the data. Instead, we take advantage of the Slater, 2021 map of average dh/dt summer trends, and use these to estimate a temporal uncertainty for each observation as follows:

$$\sigma_{temp}(x, y, t) = \frac{dh}{dt}_{Slater}(x, y) * (t - t_0)$$

With t_0 being the reconciled mid-summer date. This date (August 1st) is now reported in the paper.

We further tested the resulting uncertainties at cross-overs 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. The distributions best approximate a normal distribution for cross-overs located in areas where the Slater et al (2021) map of summer dh/dt were defined, and worse for areas where the temporal component was estimated based on the mean summer trend for the entire ablation zone. The difference was particularly important for the IS2 observations, which to higher degree are located along the ice sheet margin, and therefore outside the Slater data region.

Main comment 2: Report spatial correlation of errors for your data product

As pointed out by the authors in the short discussion (L617-620), even though your estimates have a reasonable error, it might not be random at all because of the temporal differences in acquisition.

While it would be quite a task to correct these temporal errors in detail (feasible by extending your kriging framework with a dimension of time or seasonality, for instance, or with a more simple models of your multi-annual changes during the summer season), the authors can still model the spatial structure of these errors. And this information would be quite important for downstream applications.

As noted by the reviewer, the uncertainties of the PRODEM product are not random, but display systematic deviations due to the spatial correlations of acquisition time of measurements, with one satellite track being measured at the same time. We agree that it could be interesting to extend the kriging framework with a time dimension to fully account for these seasonal changes. However, as also pointed out by the reviewer, this would be quite a task, which we consider to be out of scope for the current paper.

Instead, we follow the reviewer's suggestion to model the spatial structure of errors for the PRODEMs, as described below.

At the local scale, those are systematic errors due to common temporal sampling. But, as long as there is no bias in the time of data acquisition, those will be centered on 0 on average (which the authors validate with their comparison to IS2 in a LOOCV, Table 1). Which means that, at the larger scale of Greenland, they can be interpreted as structured errors with spatial correlations due to the temporal sampling discrepancies (Hugonnet et al. (2021)).

Here, the authors should at least report an "average" variogram that represents the spatial correlation in errors their products due to temporal sampling (this has be done on the LOOCV differences you compute in Section 7.1).

As part of the validation of the PRODEMs, we have now added a section describing the spatial structure of correlated errors based on the normalized prediction errors from the LOOCV. Since the error structure is likely to be aligned with the satellite tracks, in particular the IS2 observations forming straight lines, we construct directional variograms along and across track directions. In the along-track direction, the error structure has an effective correlation length of a few kilometers, whereas errors in the across-track direction are essentially uncorrelated.

We further provide an analysis of the prediction errors (raw and normalized) stratified relative to surface roughness, elevation, latitude, ArcticDEM uncertainty etc., allowing us to identify troublesome areas for the predicted elevation and reported uncertainties.

Even better would be to have a varying variogram depending on a map of "time lag to closest elevation used" or similar, that would reliably represent the errors in both space and time!

You could then validate your variogram representing your errors using Block LOOCV, and studying that the spread matches that predicted by your variogram in space (following classical error propagation with a spatial correlation term, e.g. Hugonnet et al. (2022), Equation 18).

However, the "swath" nature of your input data, and hence the temporal sampling, is less than ideal to capture errors in temporal sampling with omnidirectional variograms. It's still an important added value, and I don't see an easy solution to this directional problem, except trying to come up with a better scheme to correct these temporal errors seasonally to your middle summer date, using your own multi-annual data (as stated at first).

As previously mentioned, we now provide a better estimate of the input data uncertainty of representing the mid-summer elevation surface due to their different acquisition time.

While we agree with the reviewer that it would be ideal to construct the PRODEMS using a 3D kriging framework to fully account for the seasonal changes, we consider this to be out of scope for the current paper – but we appreciate the comments, and would like to work on improving this aspect for future editions of the PRODEMs.

Main comment 3: (Optional) Regionally-varying kriging: could also decompose and standardize your elevation field variance sources to explain the variability at the source

I must have been quite the effort to setup this regionally-variable scheme with GSTools!

First, while the Cressie estimator is reasonably robust to outliers, it is still fairly sensitive for what we get in elevation data (ArcticDEM outliers). You could consider using Dowd's estimator (Dowd, 1984) that actually matches your NMAD in terms of how it is derived. This estimator is implemented in SciKit-GStat (Mälicke et al., 2022) which has functions to export models to GSTools.

We now produce the experimental variograms using Dowd's estimator. Consequently, we are now using SciKit GStat for experimental variogram estimations, instead of GSTools where this estimator is not implemented. We only use the tool to produce the experimental variograms.

Second, while the regionally-varying kriging is well-implement and definitely valid, it might be over-fitting your data locally by using that regionalized variogram subsample. This would result in an over/under-smoothing of the final interpolated field compared to reality, because of the way the scheme is implemented. Another way to avoid the non-stationarity would be to describe the nature of your spatial covariance and its variability (summed variance components with a certain correlation range, and their individual variability). The variability in error with long correlation range would for instance likely linked to time lag and dhdt (Main comments 1 and 2). Then, using standardization + decomposition (as in done a bit in Hugonnet et al. (2022) with instrument resolution and noise, but was less important there), you could transform your data to reach second-order stationarity, and apply kriging.

I am mostly presenting this option for the authors to realize that there are ways to avoid a regionally-varying variogram implementation. But I think that, given the complexity of this scenario (varying time sampling that is uncorrected), the current approach is completely valid and it would be hard to implement a decomposed + standardized approach.

We appreciate the suggestion. We did at first try to implement a standardization scheme, but were not successful in making it work properly, and hence decided to go to regionally varying kriging approach. One likely issue for implementing the standardization scheme was that ArcticDEM used as reference DEM had different properties in different regions; with the new version of ArcticDEM this may (or may not) be less important.

Main comment 4: Justify the choice of 500 m for the final product

I don't think it is explained anywhere in the manuscript why the authors chose 500 m for their study, it would be good to add a paragraph on this!

We have added the following paragraph on why we use 500m resolution for the final 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.

Additionally: ArcticDEM mosaic v4 is released since last month, with significant improvements. If it is not too much work, the authors could update their pipeline with this latest version.

This has been updated, and the text and figures throughout the paper have been changed accordingly.

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, from a different location, is provided below.

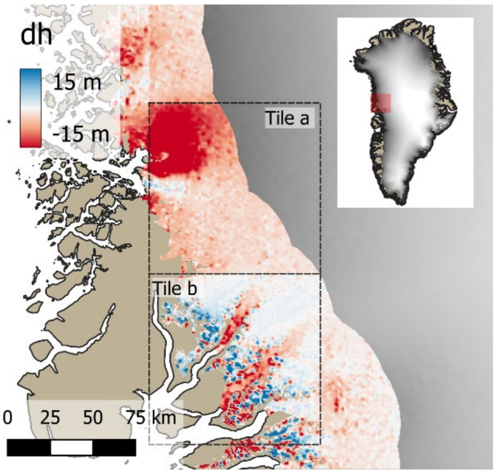


Figure 10: Elevation anomalies (here for PRODEM19) relative to ArcticDEM display a checkerboard pattern due to the way ArcticDEM was constructed. Two ArcticDEM tiles of 100x100 km are indicated.

Main comment 5: Drop the nugget for the kriging

Using a nugget in a variogram represents a very specific type of discontinuous spatial pattern at the boundary of the measured location. This is really only relevant to point observations that actually belong to a small spatial ensemble (such as gold nuggets), and is not relevant to inherently continuous elevation data on a grid.

You can actually see on Fig 4. the misfit of the black and red variograms at short spatial lags: they should not have a nugget at all, and it would show if more short lags were sampled.

We do not agree with the reviewer that the nugget should be dropped completely - in our case, the nugget should reflect the observation uncertainties. However, as the reviewer points out, during modelling of the variogram, the nugget value is often overestimated. We therefore now prescribe the nugget value as the median of the uncertainties of the surrounding observations when constructing the variograms.

However, this change gave rise to several other changes in the variogram estimation procedure. With a prescribed nugget value, it was often not possible to construct well-fitting variograms out to a distance lag of 15km. Since we only use the closest 200 points for interpolation, we only need to know the variogram structure at relatively close distances. We therefore changed the variogram binning, decreasing both the bin size and the maximum bin size. With this adjustment, the misfit of the variograms often improved at small spatial lags.

Constraining the nugget value also meant that the R2-score no longer was an appropriate measure to describe how well the variogram model fitted the experimental variogram: In some locations, the calculated semi-variance at the first lag (and first lag only) was substantially larger than the prescribed nugget value (as well as the semi-variance at the subsequent lags). For some of these, the R2-value became negative despite a visually good fit. We therefore now refrain from using the R2-score to filter out outliers in the variogram parameter maps.

To improve your spatial lag sampling in the variograms, consider using a non-uniform binning. Those can be long to compute and, for pairwise sampling efficiency, in SciKit-GStat, there are different types of metric spaces just for this, one based on the sampling scheme presented in Hugonnet et al. (2022), Fig. S13. Those are also wrapped more conveniently in xDEM for raster data (xDEM contributors, 2022).

We agree that this is something to consider for future PRODEM versions. For now, however, we have decided to go a slightly different way. When modelling the experimental variogram, we now weight the value at a given lag with the number of pairs of observations within the lag. In this way, a well-defined semi-variance value is weighted higher when fitting the model. While incorporating this, we adjusted the weighting to decrease over 1km instead of 5km (L374), as this (due to the overall spatial distribution of observations) resulted in an approximately similar weighting scheme as previously.

Line-by-line comments:

Title: Having “present” in the title for the period is probably not a good idea long-term.

We have changed the title to “(2019 through 2022)” for improved clarity.

L40-42: This statement unlikely holds true. Typical random errors in modeled ice thickness are +/- 100 meters near flux gates with important spatial correlations as well as systematic errors (Kochtitzky et al., 2022; 2023). While predicted surface elevation errors are typically less than 10 meters (Hugonnet et al., 2021). Even measured (not just modelled) ice thicknesses are usually less precise and with larger correlations in errors than measured surface elevation. The authors could simply make the case that we need more annual data to represent the surface elevation of the ice sheet.

This has been removed, and we have changed the introduction so that it now focuses more on the importance of ice sheet elevation monitoring, and how the PRODEMs will offer new insight into ice sheet monitoring efforts.

L48-52: Same remark.

See answer above.

L81 and onwards: Need to add spaces between numerical values and unit.

Done.

L190: Could directly introduce the Normalized MAD or “NMAD” directly, which has been around for while in elevation data analysis (e.g., Höhle and Höhle, 2009).

Agreed. We now write as follows:

While not strictly adhering to a Gaussian distribution, an approximate Gaussian standard deviation describing the spread of the distribution can be obtained using normalized MAD: $s_{MAD} = 1.4826 * MAD$*. This approach has the advantage of being more robust to outliers than a direct calculation of the standard deviation.*

L319: Need to specify: "second-order" stationary!

Done

References:

Mirko Mälicke, Romain Hugonnet, Helge David Schneider, Sebastian Müller, Egil Möller, & Johan Van de Wauw. (2022). mmaelicke/scikit-gstat: Version 1.0 (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5970098

Hugonnet, R., Brun, F., Berthier, E., Dehecq, A., Mannerfelt, E. S., Eckert, N., & Farinotti, D. (2022). Uncertainty Analysis of Digital Elevation Models by Spatial Inference From Stable Terrain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 6456–6472.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., & Kääb, A. (2021). Accelerated global glacier mass loss in the early twenty-first century. *Nature*, *592*(7856), 726–731.

Kochtitzky, W., Copland, L., King, M., Hugonnet, R., Jiskoot, H., Morlighem, M., Millan, R., Khan, S. A., & Noël, B. (2023). Closing Greenland’s mass balance: Frontal ablation of every Greenlandic glacier from 2000 to 2020. *Geophysical Research Letters*, *50*(17). https://doi.org/10.1029/2023gl104095

Kochtitzky, W., Copland, L., Van Wychen, W., Hugonnet, R., Hock, R., Dowdeswell, J. A., Benham, T., Strozzi, T., Glazovsky, A., Lavrentiev, I., Rounce, D. R., Millan, R., Cook, A., Dalton, A., Jiskoot, H., Cooley, J., Jania, J., & Navarro, F. (2022). The unquantified mass loss of Northern Hemisphere marine-terminating glaciers from 2000–2020. *Nature Communications*, *13*(1), 1–10.

Höhle, J., & Höhle, M. (2009). Accuracy assessment of digital elevation models by means of robust statistical methods. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing , 64(4), 398–406.

xdem contributors. (2021). xdem (v0.0.2). Zenodo. https://doi.org/10.5281/zenodo.4809698