Dear RC3,

Thank you for reviewing our manuscript and for your constructive and comprehensive feedback. Please find our responses below. To help distinguish between comments and our responses, comments are shown in black and our responses in Blue.

Summary

The authors present a novel data set of Digital Elevation Models and Orthomosaics generated from historical aerial imagery collected over the Antarctic Peninsula using a workflow adapted from the Historical Structure from Motion Package (Knuth et al. 2023). The DEMs are calibrated and independently validated against modern high-resolution satellite optical stereo (REMA) and laser altimetry (ICESat-2) observations over non-glacierized areas. The historical dataset will be very valuable in extending the historical mass balance observation record over this data scarce region. I have some comments which the authors can consider to clarify some issues in the manuscript, and I am looking forward to published manuscript post revisions.

Major Comments

The manuscript organization can be streamlined with clear division between the methods and results and discussion. There are some places where new experiments are introduced in the results section. I have detailed those at specific occurrences in the general comments below. Reorganizing the sections properly will go a long way in improving the readability of the manuscript.

Thank you for this comment about the readability of the manuscript. We originally chose to describe some of the methods in results and discussion to maintain a concise presentation, highlighting only DEM generation in the methods section. However, we agree that the manuscript may benefit from a clearer separation between the description of the methods and the presentation and interpretation of the results. We have therefore reorganized the manuscript by separating the Methods section while retaining a combined Results and Discussion section to improve readability and avoid unnecessary repetition.

General Comments:

• In the first figure, would it be possible to enlarge the Antarctica-wide subfigure, and add star markers to locations for which previous studies have provided historical mass balance, as described in the introduction literature review? It will provide context to a wider set of audience, and clearly establish the knowledge gap being filled by the new data set and manuscript.

Thanks for this great suggestion. We updated the figure accordingly.

• In auxiliary data section, confirm if the REMA strips are first co-registered and then mosaicking is, or the mosaic tiles are co-registered to CryoSat/ICESat. Currently it is mentioned that the mosaic strips are co-registered to altimetry observations.

Thanks for this comment. We used the REMA v2 mosaic downloaded from the OpenTopography website (Howat et al. 2022). According to the documentation provided by the PGC, REMA mosaic v2 is created to form 50 km x 50 km tiles from repeat strip DEMs with a median value of available elevation values at each pixel, with filtering applied to remove outliers. The tiles are then aligned to ICESat-2 and Tandem-X 90m PolarDEM data (not Cryospat/ICESat). Additional details on the mosaicking procedure are described in Howat et al. (2019).

We have updated L75-81 to clarify this as follows.

"We used the Reference Elevation Model of Antarctica (REMA, version 2) mosaic as a reference DEM to extract stable (or static) ground ……… derived from the aerial imagery (Howat et al. 2022). The REMA mosaic was downloaded from the OpenTopography portal (https://portal.opentopography.org/). It is compiled from multiple REMA strips that are generated using very high resolution (0.32 to 0.5 m) WorldView-1,2,3 and GeoEye-1 satellite imagery through Surface Extraction from TIN-based Searchspace Minimization (SETSM) software (Howat et al., 2019). The mosaic is created to provide a more consistent and complete DEM product with blending and feathering of strip DEMs to avoid edge artefacts. REMA mosaic tiles are co-registered to ICESat-2 and Tandem-X 90m PolarDEM."

• Page 6, Line 140: I am not sure if we can compute base to height ratio/convergence angles from pixel reprojection errors. If the camera parameters are estimated well, the reprojection error can be very well 0. Maybe this was a typo in the sentence, please check and consider modifying.

Thank you for spotting this, it was a mistake. The threshold of 10 refers to Metashape's "Reconstruction Uncertainty", not reprojection error (which is indeed unrelated to base-to-height ratio and can be driven to near zero with good camera calibration and optimization). In Metashape, reconstruction uncertainty is defined based on the geometric relationship of the cameras used to generate tie points, and high reconstruction uncertainty values indicate points reconstructed from images with a small baseline (i.e., small base-to-height ratio or small parallax angle) (Metashape manual, version 2.2, see references). We have rewritten the sentence to correct "reprojection accuracy" with "reconstruction uncertainty".

• First sentence in the paragraph on line 145 is unclear: we should define better what is meant by poorly localized reprojection error, and why were the sizes of these tie points larger than the others which caused this issue. Please clarify and revise.

Thank you for highlighting this. We agree. “Poorly localized projections” are the tie points whose image positions cannot be determined by Metashape precisely. This may happen when the feature used as a tie point appears large, blurry and less distinct than other tie points. This makes the average scale that was used for measuring coordinates of the projections of the tie-point in all overlapping images larger, degrading the “Projection Accuracy“ measure in Metashape. Therefore, we removed all the tie points with a projection accuracy value of more than 5.

We revised the text to make this aspect clear,

“We filtered out tie points with low projection accuracy caused by their poor localization. Tie points are poorly localized when the features they represent are large or less distinct, making it harder to locate their exact position in the images. To remove these points, we applied a projection accuracy threshold of 5, which was measured as the average image scale of the feature across overlapping images.”

• The description of the uncertainty estimation between 185 to 190 is a bit repetitive. Consider describing the Seehaus et al. 2019 approach first, which is unchanged for both cases. Then just mention you did a 50 m outlier filter on elevation residual additionally on the altimetry measurements before applying the general Seahaus et al. approach to get rid of outliers due to clouds.

We would like to clarify that our uncertainty estimation with respect to ICESat-2 differs from the procedure used for REMA (and therefore the approach of Seehaus et al. 2019). In addition to applying a 50 m outlier filter to remove clouds, we also applied 2-98 percentile data and a 3 times NMAD threshold on the whole dataset without slope binning. This results in a conservative removal of outliers tailored to the sparse distribution of ICESat-2 elevation measurements.

• The image quality section in the first result section describing the Shanon index is a bit out of place. It should be mentioned first in the methods. And the results & discussion sections should only describe the results and interpretation, not a new method.

We agree. Now, we chose to present the description of Shannon entropy definitions in methods section and the results (along with their discussion) in separate sections to improve clarity and structure.

• Why is Section 4.3.1 stated as “relative accuracy”? In general relative accuracy would refer to residual values between maybe the overlapping regions of the historical DEM strips. In comparing against external reference like REMA, should this not be absolute accuracy?

Thank you for this important point. We chose the term "relative" mainly because the reference DEM itself carries spatially variable uncertainties. The vertical accuracy of REMA is varying and depends on terrain characteristics, image geometry, and processing conditions (Howat et al., 2019). Since REMA is not an error-free "truth" surface, differences against it are best interpreted as accuracy relative to the reference DEM rather than true absolute accuracy.

In addition, the subsection title "**4.3 Evaluation of IfAG DEMs with REMA**" explicitly indicates that this assessment is in reference to REMA. We therefore believe it should be clear that the term relative accuracy refers to accuracy relative to the REMA, rather than to internal relative accuracy derived from overlapping historical strips.

To avoid any ambiguity and to explicitly caution readers, we have added the following sentence at L186,

"Note that these uncertainty estimates relative to REMA include potential errors present in REMA itself (Howat et al., 2019). Therefore, to obtain an independent estimate of vertical accuracy, IfAG DEMs were also compared to ICESat-2. The uncertainty of the IfAG DEMs is assessed using ICESat-2 data by….."

• The description of the variogram analysis could be improved. Like the previous comment, the experiment setup should be mentioned in the method section. In the description section, some interpretation of the the three ranges and sills should be provided. Some speculative points on why was a range upto a particular value on the three different scales, what do we understand from the different sill measurements would be useful. Right now there is only a comparison with a double nested spherical model and not much weight is given to the main 3-variogram result presented in the manuscript.

Thank you for this great comment. We agree that the description of the variogram analysis can be clarified and better integrated into the manuscript. In the revised version, we have moved the description of the experiment setup to the Methods section. In the Results and Discussion section, we now provide a more detailed interpretation of the three variogram ranges and sills, including what they imply about spatial correlation at different scales. Following is the updated text about its description in **Results and Discussion**

"We fitted a triple-range spherical variogram model to characterize the spatial autocorrelation of elevation error in our IfAG DEMs (Figure 10). Each nested component in the variogram fit represents a distinct contribution, with its sill indicating the proportion of total error variance associated with that spatial scale. The short-range correlation (range of 303.39 m, sill of 0.7969) accounts for the largest share of variance, suggesting that most elevation error arises from local sources such as sensor noise. The medium-range component (range of 2312.29 m, sill of 0.3970) contributes the next major fraction of the variance and likely reflects residual lens distortion that introduces correlated errors over several kilometres (Dehecq et al., 2020). A double-range model failed to capture this substantial medium-scale variance (see Figure A2), so

we adopted a triple-range spherical model to represent this physically interpretable structure, consistent with known error sources. The smallest proportion of error variance is associated with the long-range correlation (range of 72606.89 m, sill of 0.1219), which reflects broad regional variances caused by co-registration errors, such as misalignment across image subsets (Dehecq et al., 2020; Hugonnet et al., 2022).“

• Reviewer 1 raises a good point about the degraded IS-2 accuracy in areas with high slopes and surface roughness. It would be good to acknowledge this in the discussion section. Also refer to papers by Schenk et al., 2022; Csatho et al. 2024 and other members of IS-2 science team etc on this topic.

Thank you. We have now updated the L228,
“These biases reduced to 0.3 m (Mean offset for IfAG mainland mosaic for filtered, slopes less than 30 degrees), when only lower slopes are considered (Table 6). Furthermore, the accuracy of ICESat-2 is known to degrade at higher curvatures (Shen, X et al. 2022); the uncertainties of our DEMs may therefore be overestimated in such regions.”

References :

Howat, Ian, et al., 2022, 'The Reference Elevation Model of Antarctica - Mosaics, Version 2', https://doi.org/10.7910/DVN/EBW8UC, Harvard Dataverse. Accessed 2025-12-03
Howat, I. M., Porter, C., Smith, B. E., Noh, M.-J., and Morin, P.: The reference elevation model of Antarctica, The Cryosphere, 13, 665–674,395
2019.

https://www.agisoft.com/pdf/metashape-pro_2_2_en.pdf

https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.082023.3031.1