Dear RC1,

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.

Thota et al., present newly processed DEMs and orthomosaics from 1989 over the western Antarctic Peninsula; a climatically sensitive region in need of high-accuracy datasets to estimate glacier mass change. They use established methods for processing and validating the data, and deliver the dataset in an easily readable format. My assessment of the manuscript is that the work is of high scientific rigour, and my comments are mostly on the presentation of the data. I recommend a round of minor revisions to account for them, and congratulate the authors on a well-designed study.

General comments

I downloaded and visually inspected the DEMs and orthomosaics. The DEM quality seems to vary between being visually excellent to rougher looking in other spots, for example in the north-eastern edge of Adelaide Island (near Hansen Island). I think it would be a great addition to add a layer that could be used to filter these issues out for future uses of the data. For example, a per-pixel point count, standard deviation, or the “confidence” score from Metashape should all probably reveal where bad pixels are, which could be used for filtering by the user. More simply, perhaps just publishing the dense point clouds could be an option too (pre- or post-co-registration).

Thank you for your observation. As you rightly noticed, the quality of the DEMs is generally good, but the effect of stable areas for the coregistration dampened the quality in a few locations, apart from image quality, scale, and overlap. The north-eastern edge of Adelaide Island is related to the issue of insufficient stable areas for co-registrations (can be seen in Figure 3B in the manuscript). We agree that the availability of other data would present the user with a choice to judge or fix the quality issues in the future. Therefore, we decided to add pre-coregistration dense point clouds to the Zenodo data repository.

The use of ICESat-2 ATL06 elevation data as an independent validation method strengthens the case for the accuracy of the newly produced DEMs. I wonder, however, if issues with these data might be even worse than accounted for in the current version of the manuscript, leading to an overestimation of uncertainty in the new data when comparing the two. In other words, I think there is a chance that your data might be better than reported. The MSc thesis by Liu (2023)* details the use of ICESat-2 ATL08 data for snow depth retrieval, and find concerning issues on the accuracy of the product in both high slope and high curvature areas. This is briefly discussed in the published version of his thesis (Liu et al., 2025), but much information was unfortunately lost in the publication process. I am not sure where the slope/curvature errors are introduced; perhaps they are not present in the ATL06 product at all, but I nevertheless

recommend the assessment of not just slope but also planform and profile curvature. For example, binning the elevation difference by planform or profile curvature may reveal strong correlations that may be blamed on ICESat-2, not the newly produced DEMs. I recognize that ICESat-2 validation is not a pivotal part of the manuscript, so I leave the exact handling of my comment in the hands of the authors. I simply want to highlight that there may be a way to argue that your data may be better than reported. A minimum treatment could be to bring up in the text that ICESat-2 struggles in high slope/curvature regions, so the difference spread might get lower with further filtering.

We agree that ICESat-2 may exhibit degraded performance in regions of high slope and curvature, as highlighted in Liu et al. (2025) and Shen, X et al. (2022). As suggested, we computed profile curvature and binned the IfAG – ICESat–2 elevation offsets by curvature magnitude (see Figure R1). The results confirm a clear trend: median offsets increase systematically with positive profile curvature, supporting the reviewer's concern that ICESat-2 errors may contribute to the observed uncertainty estimate of IfAG DEMs in the areas of steep, higher curvature. However, with this information, it is difficult to conclude which of the sources is the major reason for the observed bias.

Here, we would like to highlight that our aggressive filtering helps to eliminate some of these outliers for validation against ICESat-2. As shown in Figure R2, a broader filter (applying a simple elevation difference filter of ±200 m between IfAG DEM - ICESat-2) increased some of the extreme offsets in high-curvature bins (higher median, NMAD), compared to our original multi-step filtering (±50 m, 2–98th percentile, 3×NMAD) (Figure R1).

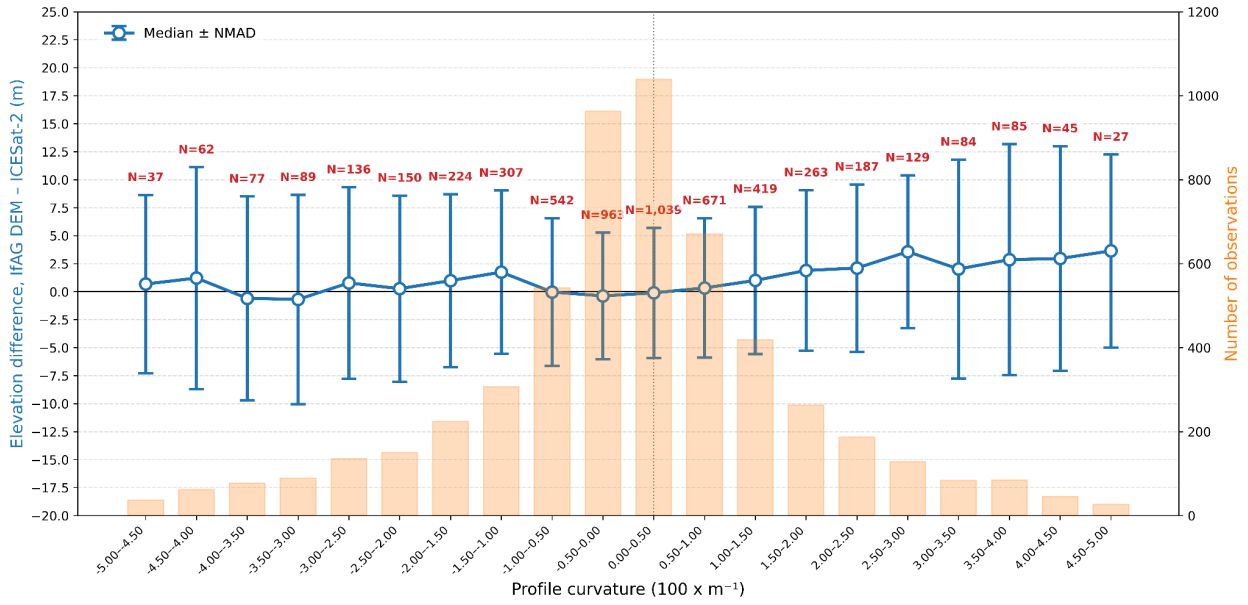

Figure R1. Elevation difference between IfAG DEM and ICESat-2 on ice-free areas with respect to Profile curvature, with our multistep filtering (described above) applied on elevation differences

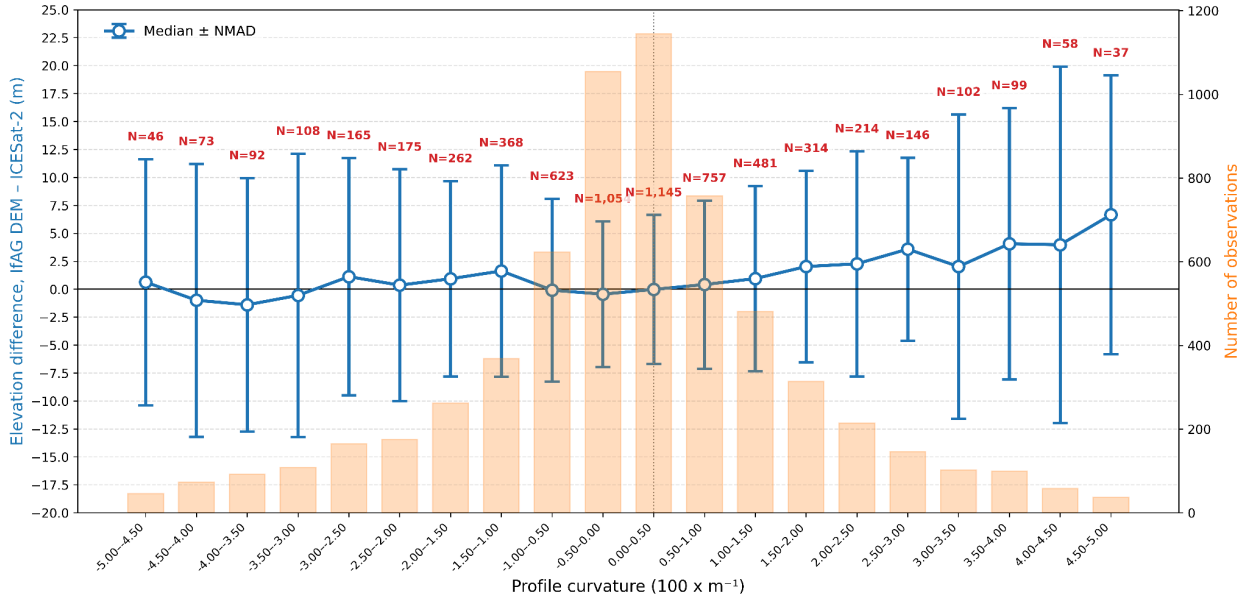

Figure R2. Elevation difference between IfAG DEM and ICESat-2 on ice-free areas with respect to Profile curvature, with a simple filtering (described above) applied on elevation differences

Given that ICESat-2 validation is not central to our study, and considering the known limitations of ICESat-2 in high-relief areas, we have chosen to retain our original filtering and reported statistics. However, we added a brief discussion in the revised manuscript (Section 4.4) acknowledging that ICESat-2 performance degrades with increasing curvature, and that the true accuracy of our DEMs may therefore be underestimated in such regions.

We added the following text to L288 of the manuscript,

"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."

Specific comments

L40: What did they find in the "follow-up analysis" of Fieber et al., (2018)? The sentence stops quite abruptly. For example, adding an average geodetic mass balance, such as in the sentence below, would complete it.

We added the findings of the paper to make the sentence meaningful.

"In a follow-up analysis, Fieber et al. (2018) analyzed the surface elevation changes of 16 individual glaciers, grouped at 4 locations on the AP and surrounding islands, between 1956 and 2014. They reported that 81% of the glaciers exhibited significant thinning, with an average

annual mass loss rate of 0.24 ± 0.08 m w.e. Most notably, this was observed at Stadium Glacier, where losses reached up to 62 m w.e. and the glacier front retreated by more than 2.2 km."

L50: "[…] against external elevation data such as […]"; I recommend being more specific than "such as"; you validate against REMA, ICESat-2 and other published DEMs.

We updated the sentence to "*The DEMs have been co-registered and evaluated against external elevation data such as Reference Elevation Model of Antarctica (REMA), ICESat-2 and other historical DEMs (Howat et al., 2019).*"

Figure 1. Great overview, but the locality labels are very small to the point of being unreadable at 100% A4 zoom. Please also add Adelaide Island and Pourquoi Pas Island, as they are repeatedly mentioned in the text.

We updated Figure 1 according to the review's suggestions.

L80: According to their webpage, it seems that only early REMA releases used ICESat-1 co-registration. As far as I understand, the mosaic is not using these data. However, REMA version 2 seems to be using ICESat-2 and TandemX data for co-registration. I can also not find any mention, apart from the REMA front-page, that discusses Cryosat-2, which is strange on their part. Please adjust the text according to the dataset version that was used. https://www.pgc.umn.edu/guides/stereo-derived-elevation-models/pgc-dem-products-arcticdem-rema-and-earthdem/

Thanks for bringing this out. We used version 2 of REMA in our work, and they seemed to have used ICESat-2 and Tandem-X 90m PolarDEM for the co-registrations as per the cited documentation. We added the following text to reflect this update.

"We used the Reference Elevation Model of Antarctica (REMA, version 2) mosaic as a reference DEM to extract stable (or static) ground elevation for co-registering our historical DEMs derived from the aerial imagery…….strip DEMs to avoid edge artefacts. REMA mosaic tiles are co-registered to ICESat-2 and Tandem-X 90m PolarDEM."

L99: Which version of Metashape did you use? I see that this information is provided on L131, but it should be the mentioned the first place where Metashape is mentioned.

Updated.

Sect. 3.1/3.2/Figure 2. I was mildly confused by starting to read about the extrinsic parameter estimation, then reading the workflow of Figure 2 which starts with intrinsic parameter estimation (specifically fiducial estimation). I think it would be easier to read if the camera intrinsic estimation section (3.2) came before the extrinsic (3.1) to stay consistent with the figure.

Thanks for your suggestion. We have updated the sequence of sections 3.1, 3.2.

Figure 2: Generally a great figure! It took me some passes, however, to understand that the lines without arrow signs were detailed explanations of whatever they were connected to. I especially got lost in the multi-stage co-registration as my eyes were flying back and forth between all the arrows when and I tried to logically arrive to the ICP box. I have no great suggestion for how to fix the clarity of the detail boxes (perhaps like chat boxes in comics?), but I suggest a small revision to the styling. I see that the boxes are rhombohedral to separate "parameter" boxes, but two of them are not ("OpenCV matchTemplate" and "Key point identification […]"). Another alternative could be to color their background differently. Finally, "Nuth and Kaab" should be "Nuth and Kääb".

Thanks for your perspective. We updated the figure in the revised manuscript accordingly.

L110: What happened to the images with one or zero principal points? Were they discarded? I suggest adding a short sentence on that here.

We assume that you meant one or zero 'fiducial' points, as clarified in L113-114; we discarded all the images with fewer than two fiducial points detected from further analysis.

L111: The text states that 29.50% of all images had three fiducial markers detected, then that the principal point is extracted from the centroid of the detected fiducials. I hope you mean the centroid of two opposing fiducials in the case of three detections! Otherwise, the estimated principal point would be very far off from the real one. I suggest phrasing it to make sure that you've thought of the case of three fiducials.

Thank you for this careful observation. We take this opportunity to clarify. When three fiducials are detected, the principal point is computed as the midpoint of the two opposing fiducials that form a complete pair (e.g., left–right). The third detected marker is not used in this case.

We updated the text to clarify this.

"The principal point was estimated from fiducial markers that passed quality checks. It was computed as the average of axis-aligned pairwise midpoints (left–right and/or top–bottom). For instances where only two non-opposing markers were detected (e.g. left and bottom), the principal point was estimated using the X coordinate from either the top or bottom marker and the Y coordinate from either the left or right marker (HIPP, 2021)."

L120: Missing "a" in "[…] Pourquoi Pas Island (PPI, see Figure 3) as calibration site."; "a calibration site".

Corrected.

L121: Why use only one site for intrinsic calibration? I don't understand why they were not estimated over the entire survey instead. I'm sure there's a good reason, but I don't learn it by reading the paragraph. Please add a sentence why you did this.

Thanks for this great question. Intrinsic calibration is performed at only one site because only one camera is used for the entire survey. Moreover, we chose this particular site for optimising the intrinsic parameters of the camera, mainly for two reasons: 1. It has well-spread stable areas for optimising intrinsics, given that initial camera locations are not accurate, and 2. It has high image quality compared to other regions of the study area. We have updated L120-124 to the following

"To estimate the intrinsic parameters of the single camera used throughout the survey, we performed camera calibration at Pourquoi Pas Island (PPI, see Figure 3). This site was selected for two main reasons: (1) it contains well-distributed, stable terrain with significant terrain features representative of the broader Antarctic Peninsula, enabling robust parameter estimation despite initial positional inaccuracies (Cziferszky et al., 2010), and (2) it offers the highest image quality in the dataset, with cloud-free coverage and strong visible contrast."

Table 2: Very interesting relationships between coverage, resolution and uncertainty. Thank you for informing about that! Small technical correction: the quality flag is "Ultra High", not "Ultrahigh".

Thank you. We corrected the text.

L145: Pedantic comment from me: Technically, are the tie points not used to inform the intrinsic/extrinsic estimation, which in turn allows for a generation of a dense point cloud? Currently, it sounds like the tie points are used to generate the dense cloud, but that is not exactly true as far as I understand it.

Thank you for this helpful clarification. You are correct, the tie points are used to estimate the intrinsic and extrinsic camera parameters, which then enable the generation of the dense point cloud through the dense reconstruction algorithm of Metashape. We revised the text to,

"The filtered tie points were used to estimate the intrinsic and extrinsic camera parameters, which were then applied to generate a dense point cloud for each subset at medium quality, i.e., at a scale sixteen times lower than the original image scale."

L146: Here it says that "medium" quality represents 1/16 image scale, while Table 2 states that medium is equivalent to a scale of 1/4.

It is a mistake in the table; 'medium' quality represents 1/4 on each side of the image, therefore, 1/16 image scale. We corrected this information in Table 2

L150: I suggest adding “originally” or something to the resolution parenthesis “(~3.5 m)”. It took me a while to understand how that aligned to the 10 m REMA resolution, before realizing that I had misunderstood the sentence.

We updated the text accordingly.

L162: “refined the alignment” → “the alignment is refined”

Updated.

L168: I would add one or half a sentence about why sub-pixel co-registration is required after ICP. I know that it’s required because minimum distances will always converge to an exact pixel offset when co-registering regular grids, but the reader might not know that (no need to be elaborate, but I’m just pointing out that it’s not obvious). Also, which tool did you use for the Nuth and Kääb (2011) implementation? xDEM? If so, then please state that.

Thanks for raising this point. We have used the demcoreg tool for Nuth and Kääb (2011) implementation. We updated the text to,

“We used the Nuth and Kääb (2011) algorithm from the demcoreg package for subpixel coregistration over stable areas, which has a higher accuracy compared to ICP, as demonstrated by a reduction of NMAD after coregistration (Shean et al., 2021). This method estimates and corrects systematic offsets by relating elevation differences to terrain slope and aspect.”

L197: As far as I can tell, this is the first time “other historical DEMs” are mentioned. Please make sure to mention this earlier in the text (c.f. my comment of L50).

Updated.

Figure 4. Please consider looking over the caption of this figure. The second sentence starts with “One subset on […]” and it is unclear what this refers to. “with a background LIMA” (and no period) could be rephrased to be more clear. Also, the latitude/longitude labels on the right hand side of the figure overlap so they cannot be read.

Thank you, it is a mistake. We corrected it.

Figure 5. Please define “with a background” better . There is also a missing period after the description of panel B. See my comment on Figure 4 of “with a background LIMA”. I think it would be nice to add a brief comment if the REMA strip is the cause for gaps or the IfAG DEM. I presume the latter?

We updated the figure caption. You are correct, the gaps are from the IfAG DEM. We have now added it to the text.

Sect. 4.2 / 4.2.1 / 4.2.2 headers: I suggest changing the header name to something with the word “uncertainty” or “accuracy” in it to more properly reflect its contents.

We have updated the headers accordingly.

L229: It is unclear what you mean with the adjustments directly influencing the accuracy of the point clouds means. Please rephrase this sentence. Do you mean that it significantly improves or harms the accuracy?

Thank you for pointing this out. We have rephrased the sentence to clarify that the adjustments improve the accuracy of the point clouds and DEMs. Larger adjustments indicate areas with higher initial uncertainties, and applying these corrections reduces errors in those regions. We updated the text to,

“Applying these corrections improved the accuracy of the point clouds and resulting DEMs, with larger adjustments corresponding to regions that initially had higher uncertainties due to rugged terrain or limited data coverage.”

Table 4: Are you sure that the focal length unit is in pixels and not millimeters? Metashape usually translates to millimeters if you use their fiducial-aware features.

We did not use the “film camera with fiducial marks” option in Metashape because we were working with preprocessed images from HIPP (HIPP, 2021). Therefore, the Metashape takes focal length in pixels.

To clarify this, we updated L138 in the manuscript to

“Preprocessed images were imported into Metashape, and the tie points were generated at the native resolution of the imagery, facilitating precise and independent alignment of each subset.”

Figures 7 and 8: Please add that these are comparisons to REMA (I presume) in the captions.

Updated.

Table 5: I find it strange that no bias is zero since co-registration has been perfomed. It also does not look like that all <30° slope differences would even out to 0 since all medians and means are negative. Is the mask different from the co-registration mask, or does the Nuth and Kääb implementation not include bias-correction (the xDEM implementation does, which is why I ask)? Please help me and future readers to understand the discrepancy!

We appreciate this insightful comment. As rightly noted, the co-registration surfaces used are different from the mask used to report the uncertainties in Table 5. There are false positives in the existing ice-free area/rock outcrop masks that we filtered to perform the co-registration of our historical DEMs (as explained in Section 3.4 of the manuscript). But, we chose to report the uncertainty of the product over unfiltered, raw ice-free areas from Silva et al. (2020), to stay consistent with previous studies in this region.

We would like to clarify that our Nuth and Kääb (2011) implementation through demcoreg package does include bias correction. Nevertheless, the biases may not be zero after co-registration in places where there is limited availability and spatial distribution of stable areas (and may also be due to relatively lower quality historical DEMs). Therefore, the reported uncertainties should be interpreted in light of the spatially variable reliability of the underlying ice-free area masks.

Figure 9: I would rephrase the end note to be even clearer that it's just scaled for visualisation. For example, it could be "For clearer visualisation, ice free […]" or something similar. Another alternative could perhaps be to change the area unit to normalized area per category, as it may be even clearer then; I leave this up to the authors to decide.

We updated it as follows *"For better visualisation, ice free…"*

Figure 10 caption: The numbers for the variogram model fit are presented both in Sect. 4.3.2 and in the caption. Since they are quite difficult to read out, I wonder if one of them could be removed for clarity. If you want to keep them, then I suggest clarifying the reporting: e.g. what is the hyphen for in "length- 303.97 m"? Perhaps change it to "of": "length of 303.97 m".

Thanks for pointing it out, updated.

L274: Replace comma with a semicolon: "[…] camera lens distortion; a double-range model […]".

Updated.

L281: Some words are missing in the parenthesis. Perhaps change to "(see the total number of observations in Table 6)"

Updated.

References

Liu, Z., Filhol, S. and Treichler, D., 2025. Retrieving snow depth distribution by downscaling ERA5 Reanalysis with ICESat-2 laser altimetry. Cold Regions Science and Technology, p.104580.

Shen, X., Ke, C.Q., Fan, Y. and Drolma, L., 2022. A new digital elevation model (DEM) dataset of the entire Antarctic continent derived from ICESat-2. Earth System Science Data, 14(7), pp.3075-3089.