Title: **Observing glacier elevation changes from spaceborne optical and radar sensors – an inter-comparison experiment using ASTER and TanDEM-X data**

Author(s): Livia Piermattei et al.
MS No.: egusphere-2023-2309
MS type: Research article

Handling editor: Vishnu Nandan

Author's response related to the comments received to the manuscript: https://doi.org/10.5194/egusphere-2023-2309

We thank the two reviewers and the editor for their feedback on our manuscript. We revised our manuscript under consideration of their comments and suggestions, which helped to improve the clarity of our paper.

The main changes we made to the manuscript are as follows:

- We improved the method and discussion sections by clarifying the challenges for optical and radar data and the solutions implemented by the participants within their processing chain.
- Throughout the manuscript, we improved the description and clarity of the challenges of radar data, especially in mountainous areas, and the impact of radar penetration on the assessment of elevation changes.
- We adjusted Figure 4 by adding a list of all participants for each group.

We are convinced that these revisions - as detailed in the point-by-point reply below - addressed all the comments raised by reviewer 1 and the editor (based on the feedback from reviewer 2) and made our manuscript more understandable.

Point-by-point reply to reviewer comments: editor and reviewer comments are in black, and the authors' response in blue, with citations from the revised manuscript in green.

…………………………………………………………………………….

RC1: 'Comment on egusphere-2023-2309', Silvan Leinss, 16 Jan 2024

Review of "Observing glacier elevation changes from spaceborne optical and radar sensors – an inter-comparison experiment using ASTER and TanDEM-X data" by Livia Piermattei et al. EGUsphere-2023-2309.

General comment:

The paper is well written and almost completely clear to understand. It is well structured and contains all relevant information. The intercomparison of glacier DEM differences identifies the most critical and important processing steps and also indicates which processing steps are less relevant to the results. The work is very valuable to judge on the reliability of optical or radar DEM differences for glacier height changes. I have only a few minor specific comments, mainly addressing radar-specific topics for DEM generation over glaciers.

Response: Thank you for taking the time to review our manuscript and provide us with very valuable feedback. We greatly appreciate your constructive comments and interest in our work. We have addressed the minor comments you provided, and we included the implementation in the updated version of our manuscript as reported here below in green.

The authors note the lack of time-series based approaches to estimate glacier elevation change in their study; below I provide several relevant insights that can be learned from analyzing radar DEM time series over glaciers based on an IEEE publication from 2021, including an openly available dataset with 140 TanDEM-X DEMs from Aletsch Glacier (2011-2019).

Response: Thank you for pointing out this interesting study on the TanDEM-X time series on the Aletsch Glacier. While we cannot integrate these results into our experiment, we will cite them in the manuscript and discussion sections and compare the findings with our results on the Aletsch Glacier.

Specific comments:

250: "as reference for surface elevation for TanDEM-X production": Do you mean "for production of the TanDEM-X DEMs (as described in section 2.5.2)"? - To avoid confusion with the TanDEM-X DEM from DLR/Airbus or the TanDEM-X mission.

Response: We agree this needs clarification. For the production of our TanDEM-X DEMs (as described in section 2.5.2), we apply differential interferometry using the SRTM DEM as a reference. The processing of our raw DEMs follows the approach by Sommer et al. (2020) without the final co-registration or mosaicking, as this was the task for the intercomparison exercise. We have clarified this in the text.
Line 252: "The SRTM DEM is used as reference surface elevation for TanDEM-X DEM production based on differential interferometry following the approach by Sommer et al. (2020) (as described in section 2.5.2). Additionally, the SRTM DEM is directly used to estimate the elevation change rates of the Baltoro and Northern Patagonian Icefield for the period 2000−2010s."

Sommer, C., Malz, P., Seehaus, T.C. et al. Rapid glacier retreat and downwasting throughout the European Alps in the early 21st century. Nat Commun 11, 3209 (2020). https://doi.org/10.1038/s41467-020-16818-0

336: "... a mean or median ... is calculated, starting with the DEM with the smallest date difference to the validation": I don't understand this sentence. a mean or median is calculated over a set of data; why would a mean/median start with some specific DEM pair?

Response: Thank you, and we agree that the sentence needs to be clarified.
Line 344: "A similar approach named "DEM mosaic" groups the DEMs in a time window of approximately two years around each target date. Then, for each DEM group, the mean or median of all overlapping cells is computed to estimate the elevation value and corresponding acquisition date. Another DEM mosaic solution uses the DEM closest to the target period first and then the remaining voids are filled with the other DEMs selected within the time window based on their acquisition date. Consequently, the elevation value and its respective acquisition date extracted from the DEM groups may vary from cell to cell."

343: "vertical deformation resulting from sensor behaviour...": I hope, the used sensors do not cause any deformation of the observed glaciers :) I guess, you mean offsets or spatially varying biases in the produced DEMs as written in the next sentence. I think you can simply drop "vertical deformation" here.

Response: We agree that it was not correctly formulated. We rephrased the sentence as follows.

Line 355: “Sensor behaviour, data acquisition, and processing can systematically bias relatively large fractions of the DEMs, resulting in incorrectly measured spatial trends and vertical deformation….”

346: "ramps/constant offsets/phase jumps": I doubt that ramps are caused by phase-jumps, at least not by locally constrained phase-jumps that appear as a well visible constant offset over a bounded area. Furthermore, ramps indicate linear (or higher order) trends rather than "constant offsets". I guess, ramps could also be caused by small (cm-scale) errors in the given satellite state vectors. The TanDEM-X orbits have at least centimeter-accuracy, but an error of 1.5 cm line-of-slight would already correspond to a full phase cycle in the interferogram and a driftig error would cause a phase ramp. Therefore, I think that phase ramps could originate from orbit inaccuracies. See also (Krieger 2007, page 3331-3332). Of course, phase jumps would cause other kinds of errors and could also lead to ramps when fitting a plane over an area with locally bounded phase-unwrapping errors.
Reference: G. Krieger et al., “TanDEM-X: A satellite formation for high-resolution SAR interferometry,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 11, pp. 3317–3341, Nov. 2007, doi: https://doi.org/10.1109/TGRS.2007.900693.

Response: Thank you for pointing out this wrong formulation. We rephrased the sentence, clarifying that ramps are not caused by phase jumps.
Line 359: “TanDEM-X formation baseline errors cause tilts in range. Additionally, radar scenes can be affected by spurious elevation jumps originating from phase unwrapping errors (“phase jump”), which have a magnitude of multiples of the height of ambiguity (Rizzoli et al., 2017). Furthermore, phase ramps might also partially originate from the SRTM reference DEM due to along-track undulations with a frequency of several kilometres (Farr et al., 2007).”

Rizzoli et al. Generation and performance assessment of the global TanDEM-X digital elevation model, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 132, 2017, Pages 119-139, ISSN 0924-2716, https://doi.org/10.1016/j.isprsjprs.2017.08.008.

348: It might be irrelevant here, but the SRTM also shows along-track undulations with a frequency several kilometers that are caused by oscillations of the boom where the second radar antenna was attached. If the SRTM was involved as any auxiliary data for DEM generation, mentioning these oscillations here might be relevant.
T. G. Farr et al., “The Shuttle Radar Topography Mission,” Rev. Geophys., vol. 45, pp. 1–33, 2007, doi: 10.1029/2005RG000183. (section 5.3) and G. Franceschetti, A. Iodice, S. Maddaluno, and D. Riccio, “Effect of antenna mast motion on X-SAR/SRTM performance,” IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 5, pp. 2361–2372, 2000, doi: 10.1109/36.868892.

Response: Thank you for your comment. We included the potential phase ramp error introduced by SRTM in the text by citing the suggested reference. Please see the full sentence in our reply to the previous comment. (Line 362).

375: "bias...ignored if they were smaller than the pixel resolution": Could you specify this a bit more? I guess, the coregistration approach contains already a resampling of the DEM even when the shifts were well below a pixel resolution. Otherwise, an offset of 1 pixel (assuming 30 m pixel spacing) could cause a height error of 30m on 45° inclined terrain. What exactly is meant with "remaining biases"? I guess, all remaining biases are in the z-direction because the figure-of-merit function is usually the height-error between two DEMs? Or are there any other remaining biases in the horizontal direction?

Response: We agree that this can be better clarified. The resampling depends on the co-registration algorithm; here, we are referring to the remaining vertical bias as, ideally, there is no remaining horizontal bias after co-registration. To correct the remaining vertical bias, one could add or subtract the related offset. Solutions adopted by the participants included applying a correction for the across and along-track shifts in ASTER DEMs and removing a first-order spatial polynomial vertical bias. We clarified this in section 3.3.2 Spatial bias correction. So, we removed this sentence in the co-registration section (Line 393) to avoid confusion and repetitions.
Line 370: "... DEM generation as proposed by (Girod et al., 2017). After co-registration, additional vertical bias corrections were applied to the ASTER DEMs. UGA removed a first-order spatial polynomial vertical bias, while LEG corrected the across and along-track shifts based on Gardelle et al. (2013)."

381ff: Could you also mention artifacts due to wet snow or terrain with very low backscatter? When the backscatter drops below the noise-equivalent-zero, the coherence is completely lost. I guess, the impact is similar to radar shadow.

Response: Thank you. We agree, and the mentioned effects are added to the manuscript as follows.
Line 400: "...Artefacts in the TanDEM-X DEM are produced by phase unwrapping errors and low or lost coherence in areas with low backscatter or even below the noise-equivalent-zero. This problem especially occurs in mountainous regions, where wet snow areas have very low backscatter values. Additionally, steep slopes and vertical cliffs cause artefacts and voids in side-looking InSAR acquisition geometry due to shadowing and layover."

402: For penetration biases, you could add a reference to Leinss (2021), Section VI-D: S. Leinss and P. Bernhard, "TanDEM-X: Deriving InSAR height changes and velocity dynamics of Great Aletsch Glacier," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 1–18, 2021, doi: 10.1109/JSTARS.2021.3078084.

Response: Thank you for the suggestion. We added the reference in the manuscript as follows.
Line 423: "... but average penetration depths of up to several metres (> 5 m) have been reported previously on Aletsch Glacier (Leinss et al., 2021; Bannwart et al., 2023) and other mountain glaciers in the Alps and High mountain Asia (Dehecq et al., 2016; Li et al., 2021) as well as the ice sheets (Abdullahi et al., 2018; Rott et al., 2021), while for the longer wavelength C-band of SRTM, …"

612-613: I suggest to provide numbers objectively describing the deviations from the validation dataset rather than the subjective words "perfectly fit", "underestimate", "overestimate" because order of magnitude of "perfectly fit" and "under/overestimate" in Fig. 11 is almost the same. - although within the standard error (also here: provide a number). Could you mention if the means of ASTER and TanDEM-X include or exclude the low-confidence results?

Response: Thank you for this comment. We improved the text by adding the reference to the figure legend in the text, which was missing, and we also reported that the ensemble means are calculated excluding low-confidence results. This information was only stated in the figure caption. However, we prefer not to add the values here as the corresponding mean values and standard error are in the figure legend.
Line 645: "...The ensemble mean of the full experiment sample (excluding low-confidence results) perfectly fits the validation result, while the ensemble means of the ASTER and TanDEM-X underestimate and overestimate, respectively, the validation data, although within

the standard error of the corresponding sample. Values are reported in the legend of Figure 11."

624: "large spread between the experiments": Could you provide a number how larger the spread is (between results with reasonable confidence)?
Response: We agree. We reported the range of the spread in m per year, excluding low confidence results, for the three study sites and referred to the supplement table for all dh results.
Line 657: "...that the large spread between the experiment results (ranging from 0.7 to 1.0 meter per year, as shown in Tables S16-S18 under "dh_T2_final") represents a major challenge…"

633: "none of the groups used a time series approach with the TanDEM-X data, partly, to avoid issues with radar penetration, which is expected to have a larger impact in cold winter than in wet summer snow conditions":
Certainly, radar penetration is a major uncertainty for radar interferometry-based DEMs, however, Leinss and Bernhard (2021) used a time series of ~140 TanDEM-X acquisitions over Aletsch glacier. In Section VI-C (p.4808) the authors observe that their height time series (Fig. 11) show the smallest variability between December and January, indicating relatively constant penetration during this period. In consequence they used only the winter(!) acquisitions for temporal regression of their time series, because during summer and especially during spring, the observed height changes vary significantly stronger due to variable penetration (dry/wet snow) and ice melt in summer. For the period 2011-2019 they observed a height change rate of -1.54 m/yr for the observed area (59 km^2). The observed area was reduced compared to the 78 km^2 given in Fig 1 (this paper) due to radar layover and shadow affecting mainly the accumulation area where height changes were negative, but relatively small (layover affected to a minor fraction also the glacier tongue with significant height changes of -3m/yr). A back-of-the envelope calculation, referencing to the 78 km^2 (Fig. 1, this paper), results in a height change rate of (-1.54 m/yr * 59 km^2 + 0.3 m/y * (78 km^2 - 59 km^2))/78 km^2 = 1.24 m/yr which is close to the validation value of -1.14 ± 0.15 m/yr (Table 1, this paper). The TanDEM-X DEM time series are available online: https://doi.org/10.3929/ethz-b-000482456

Reference: Leinss and P. Bernhard, "TanDEM-X: Deriving InSAR height changes and velocity dynamics of Great Aletsch Glacier," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 1–18, 2021, doi: 10.1109/JSTARS.2021.3078084.

Response: Thank you for clearly describing the results obtained using a time series approach with TanDEM-X and providing the radar penetration findings. We also appreciate the comparison of this study with our inter-comparison results. We added this reference and your key insights regarding radar penetration (see green text below). However, the observation periods and areas were different and, hence, not entirely comparable, and hence, we decided not to report these dh/dt values in the discussion.
Line 666: "We note that none of the groups used a time series approach with TanDEM-X data. This could be due to the experimental setup, with much fewer available DEMs than those used, for example, by Leinss and Bernhard (2021), who extracted the elevation change of Aletsch glacier using a time series of approximately 140 TanDEM-X acquisitions. A dense DEM time series is required to evaluate periods of stable signal penetration throughout the year. Leinss and Bernhard (2021) observed that their elevation change time series shows the smallest variability between December and January, indicating relatively constant penetration during this period. As a result, they used only these winter acquisitions for temporal regression of their time series because, during summer and especially spring, the observed elevation changes vary significantly due to variable penetration caused by dry/wet snow and ice melt in summer."

713-735: Section "Differences in Survey Period": I think this section is mainly valid for optical data and not well suited for radar data because height changes due to penetration can be stronger than height changes due to accumulation. As discussed in Section VI-C in Leinss (2021), reference above, the apparent (radar-)height decreases in early winter (Oct-Dec) because refreezing of firn causes more penetration bias that the additional accumulation bias due to fresh snow. Furthermore (adressing line 731), radar data ist least suitable for comparison of spring DEMs because penetration varies most strongly during spring.

Response: Thank you for this comment. We agree that for TanDEM-X, the temporal correction based on the glaciological method is not really suited. This is why TanDEM-X results were submitted already temporally scaled by most groups. We clarified this point in the manuscript as follows, and we added the finding about early winter penetration into the radar penetration subsection.

Line 758: "…no correction as the DEM selection was already close to target dates, corrections filling temporal gaps with annual glaciological observations, and corrections based on long-term annual trends derived from the selected DEMs. While the first two strategies mainly applied to optical DEMs, they were not well-suited for radar data. Most groups used winter radar DEMs, and the few tests with TanDEM-X DEM from the end of the ablation season were considered low confidence due to unrealistically low values of elevation change over ten years (e.g., Hintereis DLR-1 and UIO-5, Fig. 7 and Table S16). Therefore, TanDEM-X results were submitted already temporally scaled by most groups. Long-term trends were scaled by…."

Line 802: "This bias, which can be as high as 8 meters, aligns with the findings of recent studies by Leinss and Bernhard (2021) and Bannwart et al. (2023) conducted in the same location, while time series analysis for other regions even revealed higher seasonal signals (Vijay and Braun, 2017). According to Leinss and Bernhard (2021), the height changes due to penetration changes can be stronger than height changes due to accumulation, and the apparent (radar-)height decreases in early winter (Oct-Dec) because refreezing of firn causes more penetration bias than the additional accumulation bias due to fresh snow."

745: "radar penetration of a few decimeter per year": Here, an absolute dh value should be given, rather than a change rate. Penetration matters less for larger time intervals between DEMs.

Response: We agree, thank you for this point. We added the absolute values and clarified the sentence.

Line 791: "In our experiments, only two groups – UST and LFR – addressed this issue by applying radar penetration corrections for Baltoro, while others included this in their uncertainty estimate. The corrections made by UST and LFR were a few decimetres per year, or 1.2 m and 2.1 m, respectively, over 2000-2012 when using SRTM and almost negligible in the case of two TanDEM-X DEMs (LFR group), with a correction of 0.07 m per year over the period between 2012 and 2019."

Caption, Figure 14: A summer winter comparison is not only biased by penetration but equally by ice melt especially at lower elevations. The text mentions this, but it needs also to be mentioned in the caption of the figure.

Response: We agree, thank you for the suggestion. We updated the figure caption as follows.

Line 811: "Figure 14: The potential impact of radar penetration at high elevations and the ice melt at lower elevations on the Aletsch glacier when two TanDEM-X DEMs from different seasons are subtracted. The elevation change of summer (2013.08.11) and winter (2013.03.21)

DEMs are shown with the glacier hypsometry (grey bars, second y-axis). For elevation bands with limited areas (less than 0.1 km^2), the elevation change is shown as an empty circle."

770: "in the glacier forefield": add: "and at the glacier margins", because less ice thickness usually comes along with a reduced ice extend at all glacier margins.

Response: We agree, thank you. We added it to the manuscript.
Line 808: "...or no ice loss in the glacier forefield and at the glacier margins (within RGI outlines)."

678, 719, 806 vs. 630: The selection of the best acqusition time seems to differ between optical and radar data. While for optical data, the end of the ablation season seems to be the best option, for radar data DEM pairs from the mid-winter seems to be best suited as the radar penetration ist most similar in deep winter (line 630, and Leinss 2021). In the accumulation area (and only there), radar height changes appear still significant at the end of the ablation season, probably due to refreezing of firn (Fig. 11(a) in Leinss 2021).

Response: Thank you for this comment. We agree that a distinction between optical and radar scene selection strategies is appropriate. However, if for Aletsch and similar regions, the same mid-winter-to-winter penetration has been observed, we believe this cannot be generalised, as this strongly depends on the regional setting (e.g., how much surface melt is there, is it across the entire glacier, and so on ...) and might be completely different for other regions. We rephrased this section accordingly.
Line 859: "Optimal selection of DEMs based on their quality and timing is essential for accurately detecting long-term trends while minimising noise and seasonal variability. However, the optimal acquisition time for optical and radar data seems to differ. For optical data, the best option is to select DEM pairs of good quality at the annual mass-balance minimum, which occurs at the end of the ablation season. This helps to avoid biases due to snow accumulation and related temporal corrections. Conversely, for radar data, the mid-winter period appears to be the most suitable time for DEM pair acquisition since radar penetration is most similar in deep winter (Leinss and Bernhard, 2021). In the accumulation area, radar height changes still appear significant at the end of the ablation season, probably due to the refreezing within the firn (Fig. 11a, Leinss and Bernhard, 2021), but this is dependent on the regional setting (e.g., distribution and location of surface melt). Mosaic or time series approaches can help to reduce the effects of random error associated with any individual DEM (optical or radar) and extract additional information from the DEM stacks."

851: "steep terrain during phase unwrapping": I would split that in two parts: Radar missions have to cope with steep terrain (causing terrain masking by radar shadow and layover) and with phase unwrapping (most difficult in steep terrain due to masking and a high fringe-frequency).

Response: Thank you for the suggestion. Also, we would like to add that with more up-to-date and accurate reference DEMs, the phase unwrapping process can be supported, and fewer phase unwrapping errors can be expected. Shadow and layover can't be avoided except with multiple acquisition geometries.
Although the suggestions are quite detailed for a conclusion section, we agree with this separation and implement the suggestions in the updated manuscript.
Line 910: "The experiments also revealed clear challenges with each data source. Optical imagery has some limitations in notoriously cloudy or textureless regions (the latter holding especially for sensors with limited dynamic range), which can result in gaps and noise in the DEM. The quality of DEMs from radar missions can deteriorate in steep terrain due to the masking effect caused by radar shadow and layover. Due to masking and a high fringe

frequency, phase unwrapping is also challenging in such areas. Additionally, radar data can be affected by wet snow, which has very low backscatter and can cause correlation loss, as well as seasonal and annual variable radar penetration."

851: Possibly relevant is also wet snow with very low backscatter causing correlation loss.
Response: Thank you, this is true, and it is an additional argument for selecting DEM scenes mid-winter instead of the end of ablation in this region. We added it to the manuscript, see the sentence above.

Technical corrections/typos:

306: remove by before reference or use format "by Hugonnet et al. (2021)"
Response: Thank you, implemented.

344: remove "at a scale at a horizontal scale" -> "at a horizontal scale
Response: Thank you for spotting, removed.

……………………………………………………………………………………………………..

RC2: Editor, 21 Feb 2024

We note that the anonymous reviewer 2 provided an extensive report, which lists more than 90 generically formulated points, each with a list of sub-points. Unfortunately, these points are too generic to be addressed effectively, and many of these points have already been implemented in the manuscript or are clearly beyond the scope of the study. In fact, this report reads to us as if it was produced using Artificial Intelligence assistance, which would be against EGU recommendations: https://www.egu.eu/news/1031/statement-on-the-use-of-ai-based-tools-for-the-presentation-and-publication-of-research-results-in-earth-planetary-and-space-science/

Consequently, in agreement with the Editor, we have limited the response to the points below, which are selected and further clarified by the Editor based on the original comments from reviewer 2.

Specific comments:

Line No. 318: While the authors have indicated that details of corrections applied by each group are provided in tables S4 to S15, it would be beneficial to include a note on the lack of incorporation of seasonality corrections by many groups. Additionally, specify the buffer area chosen by each group to consider the stable area?

Response: We agree that most participants did not apply seasonal correction except for two groups (ETH and LMI). However, this depends also on the type of approaches used (e.g., pair or time series) and the selected DEMs. For example, in the case of a pair approach, where the two DEMs are very close in date to the target period but with different years, an annual correction was required rather than a seasonal correction. This was the case for the ASTER pair approach in the Hintereisferner case study. In the discussion section (lines 714-734), we explained the type and impact of the temporal correction according to the survey date.
However, in subsection 3.3.6 about temporal correction, we added a sentence summarising the main corrections used by the participants.
Line 452**:** "It is important to note that most groups did not apply seasonal corrections. However, the type of corrections applied depended on the approach used to estimate the elevation change

(i.e., single pair or time series and related regression approaches) as well as the closeness of the selected DEMs to the target period when using the pair approach."

Thank you for asking about the buffer. Indeed, we mistakenly reported a buffer of the glacier outline in the DEM co-registration section, while participants only applied the buffer for the area uncertainty calculation in section 3.4. We removed the sentence about the RGI buffer (line 374).

Line no 399-401, 403, 409-410: Indicate theoretical penetration depths that has been observed from these studies, with respect to C- and X-band wavelengths. The authors do provide a good literature reference but could be improved if you could provide a quantified threshold to the depths observed on both cold and warm glacier ice and through 'snow' (tricky right?).

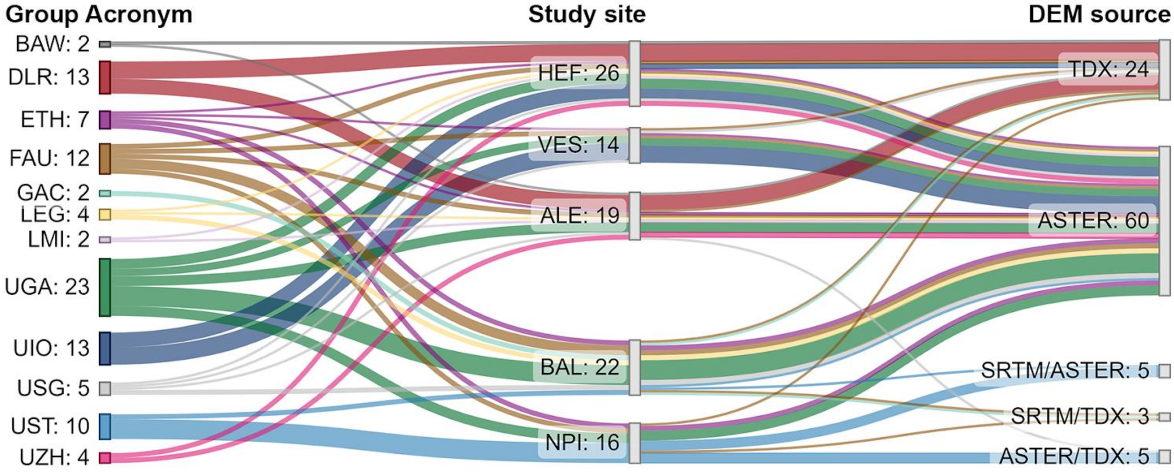
Response: As mentioned in the discussion section (Line 754), radar penetration is challenging and requires further research to determine its impact in different areas of interest and for different wavelengths. Setting a threshold for cold and temperate ice is difficult, as this will strongly depend on snow depth and underlying firn density/structure - and, therefore, on the climate and the relative elevation of the glacier.

According to our experiment, we observed a maximum penetration of up to 8 meters on the accumulation area of Aletsch glacier, a finding confirmed by other studies such as Leinss and Bernhard (2021) and Bannwart et al. (2023) over the same glacier. As written below, we reported the penetration depth estimates of C- and L-band from the studies by Dall et al. (2001) and Rignot et al. (2001).

Line 423: "... For TanDEM-X, the penetration depth is within a few metres, but average penetration depths of up to several metres (> 5 m) have been reported previously on Aletsch Glacier (Leinss et al., 2021; Bannwart et al., 2023) and other mountain glaciers in the Alps and High mountain Asia (Dehecq et al., 2016; Li et al., 2021) as well as the ice sheets (Abdullahi et al., 2018; Rott et al., 2021), while for the longer wavelength C-band of SRTM, even higher penetration depths can occur (Dall et al., 2001; Rignot et al., 2001). Dall et al. (2001) show that the height difference between lasers and InSAR in East Greenland changes from 0 m to a maximum of 13 m in the soaked and percolation zones, respectively. Similarly, Rignot et al. (2001) discovered that the radar penetration depth of C- and L-bands varies in different zones of Greenland (e.g., cold polar firn, exposed ice surface and marginal ice) and ranges from 1 to 15 m. For temperate glaciers in Alaska, the penetrations are between 4 and 12 m in C- and L-bands, respectively, with little dependence on snow/ice conditions. These studies highlight the challenge of establishing a radar penetration threshold for cold and temperate ice, as it depends on snow depth, underlying firn density/structure, climate, and the relative elevation of the glacier."

Figure 4 and section 3.x.x.x. – I wonder if there is a better way to show all 12 groups in the figure and in text. Readers have to refer to Supplement to know more about the groups and the members, which sometimes can be annoying. Can you provide a table in the manuscript referring to these members and groups. I know it will be cluttered but atleast useful not to check multiple documents at the same time. Think about it!

Response: We think that the Sankey diagram clearly shows how the study sites and DEM sources are connected to different groups. However, we agree that making the group members more visible is important. To address this, we included a table in Figure 4 that lists the names of the participants in each group with their main institution. Thank you for this suggestion.



Group Acronym – Main Institute	Participants	Group Acronym – Main Institute	Participants	Group Acronym – Main Institute	Participants
BAW – Bavarian Academy of Sciences and Humanities (GERMANY)	Anja Wendt	**GAC** – Gustavus Adolphus College (USA)	Laura Boehm Vock, Jeff D La Frenierre	**UIO** – University of Oslo (NORWAY)	Livia Piermattei, Désirée Treichler, Ruitang Yang, Luc Girod, Robert McNabb, Andreas Kääb
DLR – German Aerospace Center (GERMANY)	Lukas Krieger, Dana Floricioiu	**LEG** – LEGOS, Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (FRANCE)	Etienne Berthier	**USG** – United States Geological Survey (USA)	Christopher McNeil, Caitlyn Florentine, Louis Sass
ETH – Eidgenössische Technische Hochschule Zürich (SWITZERLAND)	Romain Hugonnet	**LMI** – National Land Survey of Iceland (ICELAND)	Joaquín M.C. Belart	**UST** – University of St. Andrews (UK)	Tobias Bolch, Gregoire Guillet, Atanu Bhattacharya, Daniel Falaschi, Owen King, Sajid Ghuffar
FAU – Friedrich-Alexander-Universität Erlangen-Nürnberg (GERMANY)	Christian Sommer, Thorsten Seehaus, Philipp Malz, Matthias Braun	**UGA** – Université Grenoble Alpes (FRANCE)	Amaury Dehecq, Friedrich Knuth, Shashank Bhushan, David Shean, Erik Mannerfelt, Romain Hugonnet	**UZH** – University of Zurich (SWITZERLAND)	Inés Dussaillant, Michael Zemp

"Figure 4: … Sankey produced with Sankeymatic (SankeyMatic, 2023). At the bottom is the list of the participants in each group with their main institution."

The explanation for UIO-4 and UIO-5 not showing stable terrain due to their use of a time series approach based on elevation bins (Table S12) is somewhat vague. Please provide a more detailed explanation of why the time series approach they used could provide stable area changes?

Response: Thank you. We provided more details in the figure caption about the lack of stable area in the dh/dt map.

Line 531: "Note that UIO-3 and UIO-4 do not show stable terrain as they used a time series approach based on elevation bins within the glacier area (cf. Table S12)."

Line No. 684-685: The authors mentioned, "Different tools, in combination with a priori or posteriori bias corrections, were used to correct shifts, rotations, and scale effects between DEMs." To enhance clarity, elaborate on the specific bias correction methods used instead of referring to them as a priori or posteriori? Additionally, instead of writing "a few meters," specify the value of the correction factor applied for the study?

Response: Thank you for the suggestion. We agree that the sentence regarding bias correction was unclear. We have now clarified it and explained what "corresponding corrections" means.

Instead of repeating the methods used by the participants, we prefer to refer to the method section regarding the co-registration and bias approaches. Additionally, we replaced the generic phrase "few metres" with a specific value, as reported below.
Line 713: "... Different co-registration tools (Sect. 3.3.3) were used to correct shifts, rotations, and scale effects between DEMs in combination with bias corrections applied before or after the co-registration to correct spatial trends and vertical deformation (Sect. 3.3.2). Figure 10 illustrates that co-registration corrections can be up to 2 metres per year and, hence, …."

Line 704: The authors stated, "Participants have applied various void-filling techniques to calculate glacier-wide elevation changes." To improve understanding, provide examples of the techniques applied for void filling?

Response: Thank you. We prefer not to report additional details in the discussion section. Our aim here is to summarise the main factors that impact elevation change estimates and highlight which factor has the most impact according to different approaches and data. However, we added the reference to the method section where we described the solutions implemented by the participants for the void filling (line 392). In the method section, we rephrased the sentence and provided more details about the used approaches.
Line 408: "Voids in the DEM difference maps were filled by almost all groups using spatial methods such as the local or global hypsometric approach (McNabb et al., 2019). This method divides the glacier into elevation bins ranging from 20 m to 100 m intervals and assigns the average elevation change of the corresponding elevation bin to the data voids. Additional solutions included the weighted version of the local hypsometric method (ETH, Hugonnet et al., 2021), and the UIO group also applied a simple Inverse Distance Weighting interpolation."

Line No. 765-774: Regarding glacier area, did the authors compare RGI 7 glacier data with other globally available glacier boundaries for Baltoro Glacier (Karakoram, Pakistan) and other data for High Mountain Asia, such as GAMDAM? Clarify

Response: For the inter-comparison experiment, participants were provided with the RGI6 outline of each study site and requested to provide elevation change estimates within that area. Therefore, we did not compare the results with other glacier inventories like GAMDAM Glacier Inventory as it was beyond the scope of the experiment, as reported in line 762: "In addition, there are some issues that were not covered in the present inter-comparison experiment, but that can have an important impact on geodetic glacier mass change assessments."

We used RGI6, as stated in line 148 because at the time of the experiment, RGI7 had not yet been released. We also received estimates from the Hugonnet et al. (2021) study that relied on RGI6 within the experiment.

Line No. 775: Density conversions: It was unclear what value of density was considered for the study 900-850? Clarify.

Response: As for the glacier area changes, the issue of density conversion is not part of this experiment. However, in the discussion section 5.2, we listed and described all the factors that can affect the elevation change estimates, including the problem associated with the density conversion (line 775). So, here, we do not refer to any specific values as this section is a broad description of the issue; instead, we reported the two most commonly used values in the community for the density conversion.
To clarify, the density conversion factor of 850 kg m^{-3} following Huss (2013) was used in this experiment for the temporal correction of the estimates provided by the participants, using the methods developed by Zemp and Welty (2023) as reported in section 2.6 and the discussion (line 728).