Reponse to Reviewer 1

1 Summary

The authors provide a comprehensive assessment of the impact of void-filling routines on the calculation of glacier elevation and volume changes. This is an important work that has relevance for a wide variety of both local and regional scale glacier change studies utilising geodetic datasets. This is a timely study and a topic I've been interested in for some time. The manuscript is of high-quality, is very well written, largely free from errors, and suitable for publication in The Cryosphere. I would recommend acceptance following minor revisions, providing that the authors address the following minor comments. I'd like to congratulate the authors on an interesting study and an important addition to the growing body of knowledge on regional-scale glacier volume change estimation. This paper will be an excellent companion to the equally good Nuth & Kaab TC study of 2010.

We would like to thank the referee for their careful and constructive comments that have helped to improve both the clarity and focus of the manuscript. Our responses to the comments below are in blue, with the original comments in black.

2 Minor comments

- Title: There is an inconsistency between the use in the title of the term geodetic mass balance' and what is referred to elsewhere in the manuscript (and what is actually calculated) which is volume change. I know why you have it up front in the title, as this is motivation for the study, but as you calculate only relative estimates of volume change' (4,23-24), the title is in fact incorrect. You do not assess the sensitivity of geodetic glacier mass balance in this work. The title therefore needs to be revised to volume change'. However, keep the geodetic mass balance mentions in the abstract and elsewhere, as they're used correctly there, and provide the important context to this work.

  We agree, and have changed the title to better match the text.

- page 1, line 18: can provide

  Changed.

- 1,21: has been calculated

  Changed.

- 1,24-25: this isn't quite right, though may just be a quirk of language. The geodetic method does not have to require extrapolation of sparse measurements, but it still can if measurements are sparse. Centreline elevation changes extrapolated to full width and differenced are still the geodetic method' (see, for example, Arendt references in your list). A couple of other studies, including one of mine, have directly compared mass balances calculated from full coverage DEMs and extrapolated centreline elevations (Barrand et al., 2010, J. Glaciol., 56, 199, doi:10.3189/002214310794457362).

  Added a parenthetical statement to make clear that we aren't excluding, for example, laser altimetry/ICESat studies from the 'geodetic' label.

- 2,4: not sure glacier water resources' is quite the phrase you're looking for as that gets into ice thickness / total water equivalent volume territory. Perhaps something like the scale of glacier change'?

  Changed to 'scale of glacier change.'

- 2,35: I know you detail from where the DEMs are from later, but this sentence is fragmentary and would benefit from a very brief description of the source of the data.

  See response to next comment.

- 3,1-2: this sentence is strange. So, you're measuring volume changes but we should not interpret these as mass balance estimates? Why would we, given the additional density correction step that is necessary to calculate mass change? Why not calculate volume changes only (and present these) and avoid any mention of mass balance entirely? Then you solve the problem of seasonal timing. This looks to be what you've done (from the following sentence). If the estimates presented here . . .should not be interpreted as mass balance estimates..', then you need to change the title of the paper and the content of the abstract, to reflect this.

  We have moved this sentence to the end of the paragraph, and included information about where the DEMs come from (C-band vs. X-band). We have also made it clear that additional corrections (density, seasonal timing) must be made before these values are interpreted as mass balances. Additionally, we have made it more clear that we are looking at the effects on volume changes, which can then be used to estimate mass balances, in response to your previous comment.

- 17,1: it's not clear to me why the elevation data in this figure should be presented in a categorised colour scale. I think it would be clearer to view and interpret if the background hillshade was slightly opaque, and the DEM data were presented in a continuous colour scale. The dark grey outlines are presumably the ice-covered land, though this is not specified in the figure itself or the caption. With a more opaque hillshade, the ice cover would then be more discernable.

  The color scale is continuous, but QGIS displays the legend as a non-continuous scale. We have added a continuous color scale to the legend, specified what the dark outlines are, and increased the transparency of the background hillshade.

- 3,9-14: I don't think there is, but is there any reason to believe that findings from a single DEMs scene from this region would differ from elsewhere in the world (perhaps regional differences between SRTM tiles?). Can you justify here why this study uses just a single difference DEM from this location, rather than multiple difference DEMs from elsewhere?

  We don't believe that there would be a significant difference in the results from this region vs. another region, in part because of the diversity of glacier types, sizes, etc. that are found in this region. The reason to choose a single DEM difference is that then the results are not dependent upon variations in the changes through time. In this respect, the effects of void interpolation is most easily extracted and understood using a single DEM difference. By using a large collection of varying glaciers in one region, we also simulate something similar to multiple DEM differences over one glacier.

- 3.20: qualify here that SRTM is commonly used at regional-scales and over medium to long time periods as it is not exceptionally accurate and likely wouldn't be as much use for e.g. 2000-2001 mass balances.

  Added a clause, "though typically over longer time periods ($>$ 10 year separation between DEMs)." to this sentence.

- 3,24-30: due to these problems, would it not have been better to select a region for which two high-quality regional-scale DEM products exist? Say, Iceland?

  Perhaps, but finding high-quality, regional scale DEM products with known dates is not an easy task. For example, the Iceland National DEM has significant errors/interpolation artefacts, as many areas are interpolated from old topographic contours. While the glacier surfaces may be quite good, these artefacts and errors make estimating the uncertainty in the calculated volume changes much more difficult. In response to another reviewer, we have provided additional information about the size of the area impacted by these voids.

- 4,11-12: what's the justification for this omission now that we know that these very small glaciers are quite important? (Bahr & Radic, 2012, Cryosphere, doi:10.5194/tc-6-763-2012).

  We omit these smaller glaciers because errors/inaccuracies in glacier outlines are much larger for smaller glaciers. As our goal is to investigate the effects of void interpolation methods on estimated volume changes, it is best to have a larger sample of on-glacier pixels to work with; voids over small glaciers result in more limited data from which to extrapolate. Also, since our objectives are methods oriented, the question about small glaciers being important is not so relevant. For further comparison of results over small glaciers, we would suggest that higher spatial resolution DEMs are required as opposed to the medium resolution DEMs used here. We have attempted to clarify this in the manuscript. We now clarifiy this in the text.

- 5,1-2: specify most spaceborne stereo optical sensors'. Sensors onboard airborne platforms or historical aerial photographs will not have identical spectral range or resolution, and therefore may not be comparable with processing of ASTER scenes.

  Done.

- 5,13: mean and median, or the mean or median? Which? See also 6,7-8.

  Mean or median; changed to clarify.

- 5,20: if this is to be replicable then some more detail is required. Which surrounding pixels? Just those immediately proximal to the void? If so, this could be problematic as there may be inaccurate elevations just beyond the low correlation areas cutoffs. If not the very next pixel, then how many back from the void space? Provide enough detail of this method for another to reproduce your procedure exactly. See also 5,25

  The interpolation is carried out using `scipy.interpolate.griddata`, which triangulates the input data and performs linear barycentric interpolation (`https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.griddata.html`). We will make the scripts used to fill the voided DEMs, as well as a csv file of resulting volume changes, available through a github repository upon the acceptance of this paper.

- 6,26: why 10%? What's your justification?

  The 10% assumption is based on a conservative estimate of the error reported by the RGI (e.g., Pfeffer and others, 2014), found elsewhere in the literature (e.g., Brun and others (2017); Kääb and others (2012)). We have added these references to the text.

- 6,27: over what scales does spatial autocorrelation occur? I see this on the next page. But, why is it assumed to be 500 m (and why only 500 m given that it can occur on a range of scales simultaneously)?

  We have chosen 500 m based on the value used in other studies, including Brun and others (2017); Fischer and others (2015); Rolstad and others (2009); Magnússon and others (2016). We are aware that it could be smaller, but feel that 500 m is a good, conservative estimate based on this previous work.

- 18, Figure 2: Can you differentiate between the colour of the glacier outline and the ASTER correlation score mask? The middle panel all looks the same colour to me (except the red), even though I think its supposed to be dark grey outline and black mask.

  Done.

- 19, Figure 3: Shaded grey around elevation changes refers to uncertainties? If so, please state in the caption.

  Mean ± one standard deviation, now included in text.

- 7,9-10: why would you find the most voids occurring in the middle of the elevation range when from an optical image feature matching perspective (where the ASTER DEM gets its correlation score) you would expect fewer features and poorer correlation the higher up you go?

  For most of these glaciers, the higher elevations are on much steeper slopes with significantly higher contrast. The middle elevation ranges tend to be the flatter, more featureless parts of the accumulation area.

- 20, Figure 4: Background Landsat scene is a bit awkward to see as its so dark. Can you adjust the contrast, or similar to a previous comment, turn up the opacity to de-emphasise the background and emphasise the elevations changes? Looks like a graded colour scale, yet legend shows categories. Shouldn't the legend by a graded colour bar too? Likewise other figures.

  Regarding the color scale, see comments for Figure 1. We have changed the background to be a pan-sharpened Landsat scene with more contrast.

- 7,18: by acquisition area, do you mean accumulation area? If you're going to list individual glacier names in the main text, these need to be listed or shown in the figure somehow.

  In this case, we are referring to the 2012 and 2013 acquisition years for the IfSAR DEM, not the glacier accumulation areas. Glacier names are shown in Fig. 1, which we now refer to here.

- 7,24: I would say patterns' isn't quite the right word here. Some of the variability' perhaps?

  Changed.

- 7,25-26: is it therefore worthwhile to consider repeating this exercise at the local glacier (rather than regional) scale? And for simple vs complex perimeter glaciers?

  It could be interesting to consider the local glacier as well, but this will be heavily dependent upon each individuals glacier change with the amount and location of voids. In this study, the local scale is covered by about half of the extrapolation methods applied (See Figs. 5, 7, and 8). Furthermore, it is not necessarily the perimeter of the glaciers that's important here, it's the variability in elevation changes, where you have some surging/advancing glaciers, many heavily retreating glaciers that reach low elevations, and other glaciers that are also retreating, but don't necessarily have the same loss vs. elevation as larger glaciers due to dynamics.

- 23, Figure 7: Great figure, but for readability perhaps the RGI60.01.' part can be removed from each individual glacier on the y axis and be included in a single y axis label? Can you also indicate in the figure caption how the individual glaciers are sorted along the y axis? It doesn't appear to be by RGI ID number, or by volume change. Is it north-south, or by glacier area, or something else?

  Thank you. We have adopted your suggestions, and added "sorted by glacier area in descending order" to the figure caption.

- 24, Figure 8: It would be interesting to see this analysis extended to smaller glaciers, or the entire sample, but I understand if this is too time-consuming and therefore not possible.

  In general, the pattern is similar for the smaller glacier classes, just with more outliers.

- 9,1-20: some very small paragraphs here (comprising just one sentence sometimes). Is this necessary?

  We have combined the last two paragraphs, and added to the paragraph beginning at line 6.

- 9,18-20: can you add some value judgments between these best three, perhaps quantifying precisely how each do and therefore which performs best? Actually, nevermind that, I see it in the next paragraph.

  Never minded.

- 11,8: please quantify rather than just stating performed well'.

  Added "producing estimates within the uncertainty of the original estimates"

- 11,20-25: please replace do well', does well' etc, with perform(s) well'.

  Done.

References

Brun, F., E. Berthier, Patrick Wagnon, A. Kääb and Désirée Treichler, 2017. A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, *Nature Geoscience*, **10**(9), 668–673.

Fischer, M., M. Huss and M. Hoelzle, 2015. Surface elevation and mass changes of all Swiss glaciers 1980-2010, *The Cryosphere*, **9**(2), 525–540.

Kääb, A., E. Berthier, C. Nuth, J. Gardelle and Y. Arnaud, 2012. Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, *Nature*, **488**(7412), 495–498.

Magnússon, E., J. Muñoz-Cobo Belart, F. Pálsson, H. Ágústsson and P. Crochet, 2016. Geodetic mass balance record with rigorous uncertainty estimates deduced from aerial photographs and lidar data Case study from Drangajökull ice cap, NW Iceland, *The Cryosphere*, **10**(1), 159–177.

Pfeffer, W. T., A. A. Arendt, A. Bliss, T. Bolch, J. G. Cogley, A. S. Gardner, J. O. Hagen, R. Hock, G. Kaser, C. Kienholz, E. S. Miles, G. Moholdt, N. Mölg, F. Paul, V. Radić, P. Rastner, B. H. Raup, J. L. Rich, M. J. Sharp, L. M. Andreassen, S. Bajracharya, N. E. Barrand, M. J. Beedle, E. Berthier, R. Bhambri, I. Brown, D. O. Burgess, E. W. Burgess, F. Cawkwell, T. Chinn, L. Copland, N. J. Cullen, B. J. Davies, H. De Angelis, A. G. Fountain, H. Frey, B. A. Giffen, N. F. Glasser, S. D. Gurney, W. Hagg, D. K. Hall, U. K. Haritashya, G. Hartmann, S. J. Herreid, I. M. Howat, H. Jiskoot, T. E. Khromova, A. Klein, J. Kohler, M. König, D. Kriegel, S. Kutuzov, I. Lavrentiev, R. Le Bris, X. Li, W. F. Manley, C. Mayer, B. Menounos, A. Mercer, P. Mool, A. Negrete, G. Nosenko, C. Nuth, A. Osmonov, R. Pettersson, A. E. Racoviteanu, R. Ranzi, M. A. Sarikaya, C. Schneider, O. Sigursson, P. Sirguey, C. R. Stokes, R. Wheate, G. J. Wolken, L. Z. Wu and F. R. Wyatt, 2014. The randolph glacier inventory: A globally complete inventory of glaciers, *Journal of Glaciology*, **60**(221), 537–552.

Rolstad, C., T. Haug and B. Denby, 2009. Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western Svartisen ice cap, Norway, *Journal of Glaciology*, **55**(192), 666–680.

Reponse to Reviewer 2

McNabb et al. compare different strategies of filling data gaps or interpolating sparse measurements of glacier elevation change in order to obtain the best estimate of total volume change (and ultimately glacier-wide mass balance). They assess the relative performance of the different gap-filling methods by comparing their results to the true volume change from the complete map of elevation change, an assessment both at the scale of individual glaciers and at the regional scale.

This is a certainly welcome study and I foresee that it is going to be widely cited. Indeed, almost all studies performing geodetic mass balance estimates need to handle data gaps. The procedure to assess the influence of different gap-filling method (i.e. taking a complete map of elevation change dh and creating realistic data voids in it) is adequate. That said, I was somewhat disappointed by the paper. It is not always clear and the writing could be improved. More importantly, I ended up with some questions that, I think, could have been, at least partly, answered. More work is needed to fully exploit this nice dataset and to transform this good study in a benchmark paper for the community.

We would like to thank the referee for their careful and constructive comments that have helped to improve both the clarity and focus of the manuscript, and we hope our revisions will be satisfactory. Our responses to the comments below are in blue, with the original comments in black.

1 General comments

1. Choice of unit to report the results. The authors have chosen to report their total volume change (and their departure from the true value) in km 3 . I do not find this unit really useful, as it is so much dependent on the glacier area. This is why, most studies use the very convenient unit of m w.e. yr -1 (or kg m -2 yr -1) to report mass balances. With the latter unit, it is easy to compare different glaciers within a region or glacier mass balance from different regions. I fully understand (and support the fact) that the authors do not want to provide mass balances here because many additional corrections would be required to obtain a meaningful value. Thus, I suggest that they use glacier-wide or region-wide elevation change (thus in meters), together with % of error (as already done).

   In accordance with this comment, as well as the comments from another reviewer, we have re-presented the analysis using $\mathrm{m\,a^{-1}}$, that is, the area-averaged rate of volume change.

2. How to handle data gaps in the error estimate. A missing section/discussion is how to take into account the data gaps in the formal error estimate. Right now, authors performed a sound sensitivity analysis and conclude on the best strategies, which is already useful. However, a remaining question is how to include the uncertainties dues to data gaps in the formal error estimate. I do not think this is done well in the literature so far and I was hoping to find an

answer here. Authors would increase the impact of their work if they could provide, at least, some suggestions. I know this is not straightforward but really hope they can tackle this issue.

We agree, both that this is not done particularly well in the literature, and that this would be a very good addition to the literature. Unfortunately, we are unable to include a correct, formal handling of the uncertainty introduced by data gaps, but rather provide insight into the magnitude of the potential error given a certain void percentage. We hope that the improved section 4.5 and Figs. 10-12 in the updated manuscript provide some useful insight on this topic. The formal handling of this error must somehow include also the ability to guess at the missing changes, and will depend strongly upon each individual dataset. For example, a dataset with a few random measurements over a geometrically simple accumulation area may be less uncertain than over a more complex accumulation area with multiple basins where both accumulation and dynamics will be spatially varying. Therefore, also at this point in time, we are not sure how to best formulate an extrapolation error equation for data gaps.

3. % of data gaps. The gap creating method makes sense. However, I had the feeling that the % of data gaps was not very high and the data voids not large. Are these percentages of data gaps in line with published values? A more aggressive gap creating threshold is discussed, but too briefly. How much data gaps are created in this case? I think many readers would be curious to know if the conclusions hold when 50% (or more) of data gaps are present.

   In the updated manuscript, section 4.5, we have gone into much greater detail, using multiple correlation thresholds to analyze the differences for individual glaciers. In particular, we examine the performance vs. percent void for the best-performing methods chosen based on the 50% threshold case.

4. Variability of dh in the study region. I miss a more thorough description of this variability. This is important here because in an end-member case (hypothetic) where there would be no spatial variability of elevation change, then most gap-filling methods would work well. How does variability vary with elevation? I expect less dh variability at high elevations where data gaps tend to be concentrated, which may explain why the local hypsometric approach works well. To quantify variability, individual glacier mean elevation change (not glacier-wide mass balance) could be calculated and the spread shown. How does this spread compare to earlier studies? It would also help to discuss whether the study region is representative.

   This is a welcome suggestion, and we have included this discussion in its own section, 4.2, which looks at both the variability within the region and compared other published values. This part of Alaska seems to be somewhere in the middle of values for regions including other parts of Alaska, the Arctic, and High Mountain Asia. Fig. 3 in the text shows the variability within (normalized) elevation bands, which confirms your expectations here - less variability at higher elevations, higher variability at lower elevations.

5. Global hypsometric approach, normalized elevations or not? To take into account the diversity of the altitude range of glaciers in a region, some earlier studies have normalized the elevation in order to extrapolate to un-surveyed areas. This is also what the authors do here to plot dh in their Figure 3. I was wondering if the normalization helped or not for the extrapolation. This procedure seems to make sense and it would be good to test its added value.

   Other studies, such as Arendt and others (2006), have looked into this with more detail. We discuss this point in more detail at 9.31-10.2, indicating estimates may be improved for individual glaciers by using the normalized elevations; however, we obtained good results using absolute elevation changes with the global mean hypsometric method, which may indicate that it is not needed.

2 Specific comments

- 1.1 mass balance does not imply sea level. Glacier mass gain/loss does.

  Changed.

- 1.2 Mentioning glaciological measurements in the abstract is not really useful. Not the core of the paper.

  Removed this sentence, replaced with "Recently, glacier mass balance has been estimated on individual glacier and regional scales using repeat, full-coverage digital elevation models (DEMs)."

- 1.5. Is "based" the best word here?

  Changed to read "the properties of which depend on..."

- 1.18. One further and strong important limitation of the glaciological mass balances is that they seem to be performed on glaciers where the mass balances tend to be more negative than the regional average (Gardner et al., Science, 2013).

  Indeed, and we have added a sentence here to reflect this.

- 1.20 They must be a reference for the WGMS data and also for the number of glaciers on Earth

  RGI reference added for the number, and WGMS reference added.

- 1.22 A reference to a review? Possibilities I see are:

  - Bamber, J. L. and Rivera, A.: A review of remote sensing methods for glacier mass balance determination, Global and Planetary Change, 59(14), 138148, doi:10.1016/j.gloplacha.2006.11.031, 2007.
  - Bamber, J. L., Westaway, R. M., Marzeion, B. and Wouters, B.: The land ice contribution to sea level during the satellite era, Environmental Research Letters, 13(6), 063008, 2018.
  - Marzeion, B., Champollion, N., Haeberli, W., Langley, K., Leclercq, P. and Paul, F.: Observation-Based Estimates of Global Glacier Mass Change and Its Contribution to Sea-Level Change, Surveys in Geophysics, 38(1), 105130, doi:10.1007/s10712-016-9394-y, 2017.

  We have included a reference to Bamber and Rivera (2007) here.

- 1.25. Do the authors exclude from the geodetic method (and thus from the study) all ICESat-based estimates of glacier volume change? ICEsat provides sparse measurements that need extrapolation. To be clarified.

  Added a parenthetical statement to make clear that we aren't excluding laser alimtetry/ICESat studies from the 'geodetic' label.

- 2.6. Acronym DEMs to be used here, as defined already. Do the authors understate that they exclude estimate based on ICESat or sparse GPS surveys?

  It was not our intent to exclude ICESat surveys from this consideration, and we have updated the text accordingly.

- 2.9 Maybe a short statement that this is certainly true for old imagery (8-bits) but that this issue is strongly reduced using state-of-the-art 11- or 12-bits stereo data? In the end, I also note that the data gaps are not so concentrated in the accumulation area.

  added “though this problem has been reduced with improved radiometric resolution of more modern sensors.” to the end of this sentence.

- 2.25 I think the interpolation methods should be described only once but not "briefly". They are the heart of the study.

  The methods are described in more detail in the actual body of the paper, but we feel that a brief summary is appropriate here.

- 2.28. Does it make a difference that the elevation with altitude is used to fill unsurveyed values vs. just multiplied by the area of the altitude band? For the glacier-wide mass balance (or the glacier-wide dh) I think it is the same. Maybe state it to avoid confusion for some readers.

  Changed to read “and estimating elevation change as a function of elevation, integrating this curve with the glacier hypsometry”, as the practical difference between the two approaches is at most very small.

- 2.30 I very strongly suggest using regional instead of global. I found global confusing (I immediately thought about the whole Earth). Or did I miss a difficulty linked to the use of "regional"?

  We have chosen to use “global” and “local” terminology based on the same terms used in “global” and “local” methods of interpolation. In the former, a single function is used to interpolate over the domain of a dataset, whereas in the latter, the function chosen changes based on the properties of a subset of the data. We have now stated this in the text.

- 2.31 "basin" needed after "glacier"?

  Removed, and elsewhere.

- 2.35 The sentence "In this paper, we use two high-quality, radar-derived DEMs." does not appear to be complete and break the flow of the introduction.

  We have moved this sentence to the end of the paragraph, and added more information about how we perform the study.

- 3.14. I think the key point for this study is that the authors have a large intra-glacier and inter-glacier variability of elevation change (a consequence of the variety of glacier type). Make it clear and quantify better (see general comments). The authors may note that some previous workers have separated different glacier types while extrapolating.

  Sentence changed to read “As such, it is an ideal region to estimate the effects of using spatially incomplete DEMs to estimate glacier volume changes, as it provides a diverse sample of glacier types, sizes, and altitude ranges, with a high variability of intra- and inter-glacier elevation changes.”

- 3.24. % of data gaps in SRTM for this study area?

  Fewer than 2.5% for the glacierized area; we have added this information here.

- 4.26 Could also have been done on the SRTM. Maybe state that this is an arbitrary choice.

  True, and stated.

- 5.3 How did the authors handle clouds in ASTER?

  ASTER scenes were chosen based on being mostly cloud-free over glaciers. If clouds were present in the images over glaciers, this is reflected in low correlation scores.

- 5.9 As said before, description of each interpolation method is central to the study. So we do not want to have a "brief summary" only. In fact the description is detailed enough.

  Removed "brief" from this sentence.

- 5.14 Here and elsewhere I found the use of "glacier basin" instead of "glacier" a bit problematic. For me a glacier basin includes the glacier + the off-glacier terrain included in this basin. Why not using "glacier" simply ? (everywhere)

  To reduce confusion, we have removed references to 'glacier basins' throughout the paper, and replaced with either 'glacier outline' or simply 'glacier'.

- 5.22 "linear interpolation". Should not it be "bilinear"?

  Yes, it should be.

- 5.22 "because the voids are relatively small" is not a very precise statement. It lack quantification (void size?) and one also would like this study to address the case of large data voids.

  We have removed this statement.

- 6.8 is 'original elevation' clear enough?

  Changed to "elevation in the earliest DEM"

- 6.27. IMPORTANT. I see no reason why the systematic error in elevation difference (epsilon_bias) obtained using triangulation between the DEMs should be divided by the square root of the number of effectively independent pixels. Either justify or correct.

  Thank you for catching this typ-o. We have fixed this equation and the following one.

- 7.8 I would have expected a higher percentage of voids in the accumulation area. This is not the case. This should be discussed.

  It is actually the case, though. The bulk of the area-altitude distribution corresponds to the relatively flat, mostly featureless portions of the accumulation areas, while higher elevations tend to be on much steeper slopes where there is more contrast in the ASTER scenes (which leads to higher correlations).

- 7.14 title of section 4.2 is not really meaningful. Improve section and sub-section titles if possible.

  Changed to "Impacts on Individual Glacier Estimates"; changed 4.3 to "Impacts on Regional Total"

- 7.16. An elevation change can be negative, not a pattern.

  Changed to read "elevation changes in the region are negative, especially at lower elevations."

- 7.21 " The pattern of elevation change shown on Rendu Glacier in the elevation difference maps". Authors need to improve the text.

  replaced "the elevation difference maps" with "Fig. 4"

- 7.26 to 8.2. These sentences are not really well written and the reasoning is hard to follow. In fact, I do not see the rational for using volume change in km^3 (and quoting an average volume change). This unit is so much dependent on the size of the glaciers whereas the global hypso method consist (if I understood correctly) in using mean/median dh per elevation band. So if a glacier (whatever his size) as a dh vs. altitude pattern like the rest of the region then the method should work.

  We have updated the text to use units of m a^{-1}, and indicated the variability in glacier elevation changes as suggested previously.

- 8.4 the fact that the authors do the conversion here to average elevation change (in meter), nicely illustrates the limit of the total volume approach (in km3).

  We have removed this paragraph.

- 8.7. IMPORTANT. The fact that the authors interpolate "over much smaller areas" (and the authors are aware of that) is quite problematic. It suggests that the authors are in a configuration (with sparse data voids) where local gap filling methods will all perform reasonably well. A much more aggressive gap creating strategy should be considered in an alternative scenario.

  Kääb (2008) used contour lines, and differences at contour lines, to interpolate a DEM, hence our use of "much smaller areas". We have removed this paragraph in the updated text, and greatly expanded the discussion and analysis in section 4.5 to show much more aggressive cases.

- 8.8-10. I do not follow the reasoning. Contour lines are maybe (certainly) biased at high elevation but a DEM created from them does not have data gaps. So the fact that contour line is floating is a different problem (like radar penetration) and does not influence the errors due to gap filling. Or better explain if I missed something.

  As above, we have removed this paragraph.

- 8.30 Showing the dh with altitude for each of these 20 glaciers and the regional mean value would nicely illustrate the text.

  Thank you for the suggestion. With the number of figures we have, we feel that it would not be an effective use of space to show this here as well.

- 9.11 Did authors used the term "global fits" before. I do not think so. If they want the readers to follow them, then they need to stick to a terminology.

  Changed to use "methods", in line with the previous terminology.

- 9.22 "Differences" of what?

  Differences to true values. Text updated.

- 9.26 authors need to clarify "relative". Is it normalized? If yes, I think they should quantify the added value of the normalization for the same global mean hypsometri method.

  Updated text to use 'normalized' in place of 'relative'.

- 9.30 "one explanation for the value". Do the authors want to discuss a high/low value? Clarify. Did they expect this method to perform better? Avoid such understatements.

  Changed to read "the overall worse performance of the elevation interpolation method versus linear interpolation of elevation change", as this is the comparison we intended to make.

- 10.9 do the authors suggest using the median rather than the mean as a metric of centrality for an elevation bin? I think it could be dangerous because the dh distribution could also be quite skewed with an elevation bin (when it comes to large glaciers for example). At least this needs to be discussed.

  We suggest that using the median for an elevation is less bad than using the median for an entire glacier; that said, it clearly does not do as well based on our results. We have added a sentence at the end of this paragraph to clarify this point.

- 10.11 Authors need to provide the corresponding % of data gaps? Does this more aggressive threshold really lead to a strong increase in data gaps? Where on the glacier?

  Please see the updated text, which discusses these issues in much greater detail.

- 10.22 Authors should detail how the ASTER DEMs were derived. Depending on the methods (and correlation threshold) the percentage of data gaps will change quite a lot. The following question is thus raised: Is it better to keep only the most reliable values in the DEM and increase data gaps (and filled them afterward) or alternatively try to get the DEM processing parameters resulting in the most complete DEM. If the authors could also contribute to this research question they would increase the impact of their study.

  We have added more details to this paragraph, and attempted to answer the bigger question at the end of this section.

- 10.25 was dDEM defined already? (not sure)

  Yes, at 5.24.

- 10.29 Is this value of 0 km^3 the volume change estimate, suggesting surprisingly no volume change? Or the difference to the true IfSAR/SRTM value?

  This is the volume change estimate, not the comparison to the "true" value. We have added 'total' to this sentence to help clarify this point.

- 10.31 "3.6" positive value of volume change? OK?

  That's the value that we get, yes.

- 11.1 surprising statement that the two methods perform as well when authors just illustrated the danger of the linear interpolation method...

  Added 'in the idealized case presented here' to clarify.

- 11.25 the dependence on the size of the voids has unfortunately not been examined sufficiently.

  We hope that the updated section 4.5 is sufficient.

- 11.30 I do not think this issue of proximity has been really addressed so that such a conclusion can be made. Or I misunderstood the statement? Do the authors suggest using a modified global method using only the glaciers in the vicinity of the one for which volume change needs to be calculated?

  This statement covers the local hypsometric methods, as well as the spatial interpolation methods, which use values either from a given glacier, or from areas within a small area around the glacier in the case of glacier complexes. We have clarified this in the text.

- 11.33 "suffice" well anyway there is no other choice right? If only a few "anomalous" glaciers are sampled than the regional total could be strongly biased

  Added a sentence 'Additionally, the regional bias for such a case may be strongly biased, as discussed in previous studies.' to further clarify this point.

- Table 1. Can the authors tell if these are simple (as I guess) or area-weighted statistics? Maybe remind in the legend the number of individual glaciers on which these statistics are obtained.

  These are not area-weighted statistics, but 'simple' statistics as you have guessed. We have included the total number of glaciers in the legend, as suggested.

- Figure 1. I could not find name of glaciers on this figure.

  We have increased the font size to make this easier to read.

- Figure 3. What is the envelop around the mean/median dh? 1-sigma of data?

  Yes. This is now indicated in the figure caption.

- Figure 2. An extra panel showing the distribution of data gaps for the more aggressive correlation threshold would be welcome. Also provide on each panel the % of data gaps for Taku Glacier.

  We have provide the percentage of data gaps for Taku, as suggested, but have not added a panel to the figure.

- Figure 5. The authors use Actual volume change here but true volume change in the text. Homogenize. Are all the acronyms used to name the different methods in the figure defined (in the text or the legend)?

  Label change to use "true" instead of "actual." Acronyms are now introduced in section 3.2, and kept consistent throughout the figures and tables.

- Figure 6. Rather than showing the dh maps for all methods (with some maps that are very similar), it would probably be best to show only the ones with strong difference. Also it would be good to show the map with data voids. So that the reader as a good sense of where the voids where. Authors could also consider moving this figure (or the suggested revised version of it) to the supplement. Showing instead the pattern of change with altitude for Taku derived from these maps could likely better illustrate some of the subtle differences mentioned in the text.

  By including all maps, even those that look very similar, we illustrate that in some cases, there aren't many differences between the methods investigated. We have included the data voids in panel a, but have otherwise left the figure unchanged.

References

Arendt, A. A., K. A. Echelmeyer, W. D. Harrison, C. S. Lingle, S. L. Zirnheld, V. B. Valentine, J. B. Ritchie and M. Druckenmiller, 2006. Updated estimates of glacier volume changes in the western Chugach Mountains, Alaska, and a comparison of regional extrapolation methods, *Journal of Geophysical Research*, **111**(F3).

Bamber, J. L. and A. Rivera, 2007. A review of remote sensing methods for glacier mass balance determination, *Global and Planetary Change*, **59**(1–4), 138148.

Kääb, A., 2008. Glacier Volume Changes Using ASTER Satellite Stereo and ICESat GLAS Laser Altimetry. A Test Study on Edgeøya, Eastern Svalbard, *IEEE Transactions on Geoscience and Remote Sensing*, **46**(10), 2823–2830.

Reponse to Reviewer 3

In this paper, McNabb et al. investigate the effect of missing data (called voids in the article) on the glacier volume change that can be obtained from digital elevation models (DEMs) differencing. The methodology is rather straightforward, they differentiate two DEMs acquired over Southeast Alaska that (almost) cover the entire glacierized area. These data are used as reference data. Then, they artificially generate voids in the data and evaluate the impact of different void-filling/interpolation methods on the on the regional glacier volume change estimate, but also for each individual glacier. They investigate 11 different void-filling/interpolation methods that are often used in the literature, providing a unique and comprehensive assessment. They conclude that most interpolation methods introduce very little bias (<1%) on the regional glacier volume change. However, individual glacier volume change estimates can be severely affected by the choice of the interpolation strategy.

This paper is rather narrow-focused, but its scope fits very well within The Cryosphere, where it will certainly reach an adequate audience. The topic is timely and very relevant, as the geodetic method is more and more widespread in glaciology. To my opinion, this paper has the potential to become a classic paper in the field of geodetic mass balance. However, and while I appreciate the concision of the paper, I have the feeling that the authors could discuss some aspects more in depth. Moreover, I sometimes had a hard time following the paper and found that it lacks clarity in its current form. These are my two major comments.

We would like to thank the referee for their careful and constructive comments that have helped to improve both the clarity and focus of the manuscript, and we hope our revisions will be satisfactory. Our responses to the comments below are in blue, with the original comments in black.

1 Major comments

1. **Volume change vs. geodetic mass balance**
   The title of the paper mention the sensitivity of geodetic glacier mass balance, but actually discuss only glacier volume changes. This decision is somehow understandable, because it is the quantity that is directly affected by the void-filling strategy. However, the impact of void-filling strategies on the individual glacier volume change expressed in km^3 is not very intuitive, and not as informative as it could be. First of all, the IfSAR DEM was acquired over two years and it would be better to present the annual mean instead of the totals, in order to get rid of this temporal inconsistency. Second, the results are largely dependent on the glacier area considered, larger glaciers being more sensitive to the interpolation (P8L12-13), mostly because for a similar elevation change they have larger volume change, due to their larger area. For example, for figures 7 and 8 (and 5?), I suggest to present the results in kg m^{-2} a^{-1} or in m a^{-1} (if the authors do not want to make any density assumption).

   In accordance with this comment, as well as the comments from another reviewer, we have rewritten the analysis using units of $m\,a^{-1}$, that is, the area-averaged rate of volume change.

   If the authors want their study to be reproduced and the conclusions of this article to be applied elsewhere, they need to analyze the influence of the void-filling strategy more in depth. I feel like the paper misses some basic, yet interesting analysis. For instance, what is the influence of the percentage of voids for individual glaciers? Of the glacier-wide mass balance/mean rate of elevation change? Of the glacier area?

The authors probably analyzed these influences already and found that they were limited/not interesting, but I think it is probably worth mentioning them, in order to apply their conclusions to a different setting.

We have significantly expanded the discussion on the impact of voids for individual glaciers and methods in section 4.5, in particular adding Figures 10, 11, and 12, and focusing on the percentage of voids where the 'best' methods start to give unreliable results. We hope that this helps to increase the impact and depth of the analysis.

2. **Some clarifications needed**
The objective of the study is quite straightforward, but a number of confusions and unclear statements prevent from an easy understanding of the paper. I had to go back and forth a number of time reading the paper, and I have the feeling that the clarity of the paper could be much improved if the authors address the three comments below. First, the author mimic the voids of a standard DEM difference, based on ASTER correlation map patterns. Consequently, I expected that they would investigate the influence of the void-filling strategy for this purpose. However, they also investigate such methods as the global ones, which are generally used for regionalization of Lidar surveys. They should make a clear distinction between these two applications when relevant. In other words, I do not think that the global methods are relevant for DEM differences void filling at the scale of individual glaciers. Correct me if I'm wrong, but I do not know any paper which studied individual glacier mass balances obtained with such global methods to fill in the holes of a DEM difference.

You are correct in saying that the global methods have not really been used to estimate individual glacier mass balances in prior studies. Our goal with including these methods was to investigate whether useful results could be obtained for an individual glacier using global methods, as well as to compare the regional estimates obtained using a variety of methods. In the updated text, we have tried to make this distinction more clear, in particular at P6L20-24.

Second, the different methods described are relatively basic, however their description should be clearer. For instance, adding equations to the description of each method would be beneficial. Alternatively, you could share the code you wrote, which would also support your conclusion in which you encourage others to test different methods when dealing with voided data.

We have added a statement in both the Methods section, and the Code Availability about where the scripts used to generate and interpolate the voids will be available.

Third, I found the example about Taku Glacier extremely confusing. If I understand correctly figure 2, the tongue of Taku Glacier is mostly free of voids. However, the global methods (panels e to g) totally change the pattern of areas where data are available! Consequently, the methods should be described as Interpolation and not Void-filling (for instance the title of section 3.2 should be changed), because they also apply to areas without voids (and it is technically a non-exact interpolation method). If I did not understand correctly figure 2, you can ignore this comment, but you should consider changing figure 2.

You are correct that the tongue of Taku Glacier is mostly free of voids; you are also correct that these methods would be better described as 'void interpolation' rather than 'void filling', given that we aren't actually 'filling' the voids, per se. We have updated the text to reflect this.

2 Specific comments

- Fig. 2: confusion between the glacier and voids outlines. You should draw the glacier outlines in a different color/line thickness, such as the panel 3 is easier to understand.

  We have changed the color of the voids in Fig. 2 in order to avoid this confusion.

- Fig. 5: add a scale/grid on the inset.

  The updated figure no longer has an inset.

- Fig. 6: this figure is extremely confusing to me. First of all, I'm missing the location of the voids on panel a, and I had to go back on the previous figures to understand where the voids where. Then the figure shows the strong impact of the void-filling procedure on the tongue of Taku Glacier, but from what I understood of figure 2, there were no voids on the tongue. This comment is in line with my major comment 2.

  See response to your major comment 2 above. We have shown the void locations in the updated Figure, panel a, to help make this figure more clear.

- Fig. 7: I think here the reader loses the information about the difference to truth relative to the total glacier volume change. I guess the larger the volume, the larger the error.

  The updated figure shows the differences averaged by glacier area, rather than the volume change in km^3, in accordance with other comments.

- P1L3 and P1L17: I do not fully agree with your definition of the glaciological method, which does not really monitor changes in surface height (that would actually be the geodetic method). The glaciological method directly measures the surface accumulation and melt.

  Changed to read, "Traditional estimates of glacier mass balance have involved *in-situ* seasonal or annual measurement of accumulation and ablation at select locations, and extrapolation of these sparse measurements to the entire glacier..."

- P1L8-9: add a word about the artificially generated voids

  Done.

- P1L11: define ASTER

  Done.

- P2L6: Digital Elevation Models-¿ DEMs

  Done.

- P2L33-P3L6: this paragraph is not really well structured in my opinion. You should describe more clearly the philosophy of your study. I suggest to move completely your warning statement about the radar DEM difference (P3L1-2) to the other place where it is mentioned (P4L19-24). You need to add something about the artificially generated voids and to better justify the choice of ASTER, among a large choice of (optical) sensors.

  We have moved the warning statement to the end of the paragraph, and added something about artificially generating the voids used. We feel that the warning is important to have in the introduction, as well as later on in the paper. We don't feel that the choice of ASTER needs justification here, as it is a widely-used sensor in glacier studies with a long (nearly 20-year) record.

- P3L9-14: what is the glacierized area?

  $\sim$5900 km^2, added to text.

- P3L19-20: provide references about studies that used SRTM to estimate geodetic glacier mass balance

  Done.

- P3L29: what proportion of the glacierized are if affected by these small voids?

  Fewer than 2.5% of the glacierized area; we have added this information here.

- P4L2-6: more references and details are needed in this paragraph. What is the precision of the IfSAR DEM? What is the proportion of voids? Has it been used in other glaciological studies?

  We have added more information here. The metadata report an accuracy of ∼1 m, while the USGS-reported accuracy is 3 m. Voids in the original acquisitions are small (<1% according to the metadata), and are filled in post-processing using proprietary algorithms.

- P4L5-6: is this sentence useful?

  Probably not, so we removed it.

- P4L12: can you justify the exclusion of glaciers smaller than 1 km^2 from your analysis?

  We omit these smaller glaciers because errors/inaccuracies in glacier outlines are much larger for smaller glaciers. As our goal is to investigate the effects of void interpolation methods on estimated volume changes, it is best to have a larger sample of on-glacier pixels to work with; voids over small glaciers result in more limited data from which to extrapolate. Also, since our objectives are methods oriented, the question about small glaciers being important is not so relevant. For further comparison of results over small glaciers, we would suggest that higher spatial resolution DEMs are required as opposed to the medium resolution DEMs used here. We now clarifiy this in the text.

- Method section: at some point you need to explain how you calculate the regional estimate. Is it the sum of the individual glaciers, or do you consider all the glaciers as a single body of ice? Do you include the glaciers smaller than 1 km^2 in this estimate? If not you might bias it.

  It is the sum of the individual glaciers. We have included this to help clarify this point.

- P5L9: see my major comment 2 -¿ you might want to rename this paragraph Interpolation instead of Void Filling

  Changed to "Void Interpolation"

- Void-filling section: how do you deal with the temporal inconsistency between your two IfSAR DEMs? I guess for most method you interpolate the rate of elevation change and not the elevation change? This should be written clearly. However, this is not possible for the method based on the interpolation of elevation.

  As stated in section 2.2.3, we remove any glacier outlines that fall 10% by area or more in both collection years, to avoid this problem, adjusting the area of the remaining glacier outlines that fall within both DEM acquisition years accordingly.

- P6L4-6: here the reader wonders why using ASTER voids for regionalization applications (i.e., Lidar based studies)? In order to test the influence of regionalization, one could extract elevation changes in your DEM difference along Lidar flight lines... but this would be another paper!

  Indeed, and it has been done by other papers, for example Berthier and others (2010). Here, we are attempting to show that while these methods may not be a useful way of estimating the mass balance of an individual glacier, it can still be useful to estimate the mass balance of a region using data from the glaciers where you do have measurements.

- P7L6: define normalized glacier elevation

  Added "(i.e., the elevation divided by the elevation range)"

- P7L14: consider switching the order between the sections 4.2 and 4.3.

  We prefer the section order as-is.

- P7L15-23: this paragraph is rather disconnected from the rest of the analysis.

  We disagree. This paragraph discusses the variability of elevation changes in the region, which helps to explain some of the differences seen between the methods.

- P7L16: the pattern of elevation change is negative -¿ the phrasing is not clear to me

  Changed to read "elevation changes in the region are negative, especially at lower elevations."

- P7L26-27: the average geometric volume change has probably little influence, contrary to the mean elevation difference.

  Per your and another reviewer's suggestions, we have changed the units from km^3 to $\mathrm{m\,a}^{-1}$.

- P8L3-10: can you say a word about the constant methods? And about the 1km neighborhood method?

  Added in the updated text in the following paragraphs at lines .

- P8L4: define RMS. This sentence is not completely clear to me.

  We have removed this paragraph. RMS values are now defined in the text at P8L16.

- P8L13-14: the larger glaciers are more sensitive than the others, because they have larger volume change for a similar elevation change, due to their larger area. This is one on the limitations of your analysis, because you look at volume changes only (see my major comment 1). You should consider extending your analysis to glacier mass balance, or rate of elevation change.

  Please see the updated text, which compares the area-averaged volume changes as suggested.

- P9L6-8: this paragraph is very short, while Fig. 8 is probably a key figure! Could you elaborate a bit? I found an unbalance with the previous paragraph (P8L30-P9L5), which is less important and much longer than this one.

  We have expanded the discussion of this figure, including adding analysis for smaller glaciers as well.

- P9L20: the 1km neighborhood is never mentioned earlier in the section 4.2.

  We now mention it in the paragraph discussing the updated Fig. 8.

- Section 4.3: consider adding a column to Table 1, which summarizes the regional totals for each method. It might also be interesting to discuss the difference between the regional estimates obtained by summing the individual glaciers versus the regional totals obtained by applying each method to the glacierized area considered as a single body of ice.

  Given that the goal is to show how the regional total changes using each method, we feel that the column showing the change is sufficient here. I'm not sure I understand the second point here - the regional estimates are achieved by summing the volume change estimates for each glacier, not by applying the methods as though the total area were a single body of ice.

- Section 4.4: why don't you use the percentage of voids of individual glaciers to study the influence of the percentage of voids on the distance to truth?

  See updated text (now section 4.5), as well as the newly-added figures which discuss this in much greater detail.

- P10L12: give the total percentage of voids for each threshold.

  See the newly-added Figure 10, which shows this.

- P10L28: again I got confused because you look at 91 individual glaciers and then you mention the global mean hypsometric as one of the best performing method... Please clarify.

  The global method uses elevation differences from all of the glaciers in a sampled area and is applied to each glacier based on that glacier's hypsometry. Hence, the global method here is only using the data from these 91 glaciers.

- P10L31: you can mention in the text that outliers are more often located near the voids, which increases their influence in a linear interpolation.

  A good point, which we have included in the text.

- P11L5-7: the order (regional volume change then individual glacier volume) is the opposite from section 4.2 and 4.3.

  Re-written to change the order.

- P11L25-26: you actually did not demonstrate this in your analysis...

  Hopefully, you agree with us that the updated analysis serves to demonstrate this.

References

Berthier, E., E. Schiefer, G. K. C. Clarke, B. Menounos and F. Rémy, 2010. Contribution of Alaskan glaciers to sea-level rise derived from satellite imagery, *Nature Geoscience*, **3**(2), 92–95.

Sensitivity of ~~geodetic~~ glacier ~~mass balance~~ volume change estimation to DEM void interpolation

Robert McNabb[1], Christopher Nuth[1], Andreas Kääb[1], and Luc Girod[1]

[1]Department of Geosciences, University of Oslo, Postboks 1047 Blindern, 0316 Oslo, Norway

Correspondence: Robert McNabb (robert.mcnabb@geo.uio.no)

Abstract. Glacier mass balance is a direct expression of climate change, ~~with~~ and changes in glacier mass has implications for sea level, ocean chemistry, oceanic and terrestrial ecosystems, and water resources. ~~Traditionally~~Recently, glacier mass balance has been estimated ~~using in-situ measurements of changes in surface height and density at select locations on the glacier surface, or by comparing changes in surface height~~ on individual glacier and regional scales using repeat, full-coverage digital elevation models (DEMs)~~, also called the geodetic method~~. DEMs often have gaps in coverage ("voids")~~based~~ , the properties of which depend on the nature of the sensor used and the surface being measured. The way that these voids are accounted for has a direct impact on the estimate of geodetic glacier mass balance, though a systematic comparison of different proposed methods has been heretofore lacking. In this study, we determine the impact and sensitivity of ~~void-filling~~ void interpolation methods on estimates of volume change. Using two spatially complete, high-resolution DEMs over Southeast Alaska, USA, we artificially generate voids in one of the DEMs using correlation values from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) scenes. We then compare 11 different ~~void-filling~~ void interpolation methods on a glacier-by-glacier and regional basis. We find that a few methods introduce biases of up to 20% in the regional results, while other methods give results very close (<1% difference) to the true, non-voided volume change estimates. Finally, we independently show using ASTER DEMs that some of best-performing methods are more robust than others, depending on the properties of the original DEMs, and therefore recommend that studies compare a few of these methods to estimate the uncertainty introduced by ~~filling~~ interpolating DEM voids.

1 Introduction

Glacier mass balance responds directly to climatic influences, and therefore long-term records of glacier mass balance reflect changes in climate. Traditional estimates of glacier mass balance have involved *in-situ* seasonal or annual measurement of ~~the changes in surface height and density~~ accumulation and ablation at select locations, and extrapolation of these sparse measurements to the entire glacier (the glaciological method; see, e.g., Cogley, 2009). This ~~provides~~ can provide a temporally dense time series for an individual glacier, but for very large glaciers or at regional scales, it is neither practical nor even possible. As of ~~June 2018~~January 2019, the World Glacier Monitoring Service has glaciological-method mass balance measurements for only 450 of the more then 200,000 glaciers worldwide ~~.~~ (WGMS, 2019; RGI Consortium, 2017), the majority of which are performed on smaller glaciers, with mass balances that tend to be more negative than the regional average (e.g., Gardner et al., 2013).

More recently, glacier mass balance ~~is~~ has been calculated over longer time spans and with larger spatial coverage by differencing remotely sensed surface elevation measurements of glaciers (e.g., Bamber and Rivera, 2007). Integrating these differences over the glacier ~~basin~~ produces an estimate of volume change. With careful consideration of the multi-annual surface change of snow, firn, and ice composition (e.g., Huss, 2013), this so-called geodetic approach provides the total mass change of a glacier. Unlike the glaciological method, the geodetic method does not require extrapolation of sparse measurements to the entire glacier surface (although in the case of laser altimtery studies, it is still necessary to extrapolate measurements to the entire glacier surface), and thereby can be used to calibrate and/or validate time series of mass balance measurements that have been obtained through the glaciological method (Elsberg et al., 2001; Zemp et al., 2010, 2013; Andreassen et al., 2016). With the current increase in the number of available, accurate Digital Elevation Models (DEMs) derived from airborne and in particular space-borne sensors, measurements of glacier mass balance using the geodetic method are and will be more prevalent, providing proper spatial accounting of the ~~glacier water resources on the planet~~scale of glacier changes worldwide.

In this study, we focus on the estimation of geodetic mass balance from ~~Digital Elevation Models~~DEMs. In general, wide coverage DEMs are created from sensors on aerial or satellite platforms falling into two categories, optical and radar. DEMs derived from optical sensors have the advantage of measuring the snow and ice surface directly, but data availability is subject to weather and light conditions, which can often be cloudy or dark in most glaciated regions around the globe. In addition, low-contrast areas on glacier surfaces, such as in the accumulation area, can often result in missing data or data voids, though this problem has been reduced with improved radiometric resolution of more modern sensors. DEMs derived from radar sensors are weather- and illumination-independent, as the active sensor acquires data even through cloud cover and polar night. However, glaciers tend to occur in areas with steep and/or rough topography, and layover and shadow can confound efforts to unwrap elevations on glaciers (e.g., Rignot et al., 2001; Shugar et al., 2010). In addition, radar signals penetrate snow and ice differently, depending on the properties of the surface, as well as the frequency of the signal; this penetration results in a spatio-temporal systematic bias in surface measurement that is still poorly understood and constrained (e.g., Rignot et al., 2001; Dall et al., 2001; Gardelle et al., 2012; Dehecq et al., 2016). DEMs derived from airborne laser scanning (e.g., Geist et al., 2005; Abermann et al., 2010; Andreassen et al., 2016) are highly accurate, spatially complete, and mostly avoid the penetration issues associated with radar-derived DEMs. Such DEMs are expensive to produce, however, and have similar requirements as optical sensors and aerial photography, i.e., clear sky or high clouds, conditions that can be difficult to find over glaciers.

In the most ideal scenarios to calculate geodetic mass balances from repeat DEM differencing, the entire glacier would be sampled systematically and with similar accuracy. In the most commonly used DEMs for glacier mass balance described above, zones of missing data (hereafter called "voids") are rather common, and may severely bias estimates depending upon how these regions are accounted for (e.g., Kääb, 2008; Berthier et al., 2018). Several different methods have been applied in the literature, and we briefly summarize them here. They include bilinear interpolation of elevation or elevation differences (e.g., Kääb, 2008); filling with an average value from a surrounding neighborhood (e.g., Melkonian et al., 2013, 2014); multiplying a mean value by glacier-covered area (e.g., Surazakov and Aizen, 2006; Paul and Haeberli, 2008; Fischer et al., 2015); and estimating elevation change as a function of elevation, ~~and either~~ integrating this curve with the glacier hypsometry ~~directly or using it to fill unsurveyed values~~ (e.g., Arendt et al., 2002, 2006; Kohler et al., 2007; Berthier et al., 2010; Kronenberg et al., 2016). In addition, we can classify these methods into "global" and "local" types, where "global" methods use data from an entire region or group of glaciers, while "local" methods fill voids using only information from an individual glacier~~basin~~, or from data closely surrounding the voids.

While various methods are used in individual studies, the sensitivity of geodetic mass balance estimates to various interpolation methods is not clear. An overarching comparison of the numerous methods is lacking, and their subsequent effects on volume change estimates at both local and regional scales. ~~In this paper, we use two~~ Using correlation values derived from optical stereo imagery, we artificially introduce voids into a high-quality, ~~radar-derived DEMs. Biases in the volume change estimates exist due to differences in seasonal timing and radar penetration; as such, the estimates presented here should not be interpreted as mass balance estimates for these glaciers. We~~ spatially-complete DEM, and difference this DEM to another spatially-complete DEM. We then apply 11 different methods to fill these artificially-produced voids~~in this spatially complete DEM difference pair~~, and compare the resulting estimates of volume change both glacier-by-glacier and regionally to determine the potential impact and sensitivity on volume change estimates. This study aims to quantify the effects of different void-handling approaches, and to suggest the void-handling methods best suited for accurate volume change estimation. As a final note, in this paper, we use two radar-derived DEMs: one derived from C-Band radar, and another derived from X- and P-Band radar. Biases in the derived volume change estimates exist due to differences in seasonal timing and radar penetration; as such, the estimates presented here should not be interpreted as mass balance estimates for these glaciers without additional corrections.

2 Data

2.1 Study Area

To test the impact of ~~void-filling~~ void interpolation methods on estimates of volume change, we chose the area surrounding Glacier Bay and Lynn Canal, Alaska, USA (Fig. 1). This area contains over 700 individual ~~glacier basins~~ glaciers (Randolph Glacier Inventory (RGI) v6.0; Pfeffer et al., 2014; RGI Consortium, 2017), with glaciers ranging from sea level to over 4000 m a.s.l., covering an area of approximately 5900 km^2. Additionally, the region is home to a wide range of glacier types, including surge-type glaciers, retreating (and advancing) tidewater glaciers, and both large and small valley glaciers. As such, it is an ideal region to estimate the effects of using spatially incomplete DEMs to estimate glacier volume changes, as it provides a diverse sample of glacier types, sizes, and altitude ranges, with a high variability of intra- and inter-glacier elevation changes.

2.2 DEMs

2.2.1 SRTM

We use the Shuttle Radar Topography Mission (SRTM) C-band global 1-arcsecond dataset as the reference DEM in this study. The SRTM was acquired in February 2000 aboard the Space Shuttle Endeavour, flying both C-band and X-band instruments (Van Zyl, 2001). This nearly global DEM is temporally consistent and therefore ideal and commonly used for geodetic mass balance estimation (e.g., Surazakov and Aizen, 2006; Larsen et al., 2007; Melkonian et al., 2013, 2014), though typically over longer time periods ($>$ 10 year separation between DEMs). We have selected this dataset, and not the US National Elevation Dataset (NED) as ~~have other studies of~~ other studies have used in the region (e.g., Arendt et al., 2002, 2006; Larsen et al., 2007; Berthier et al., 2010), as it was produced by digitizing 1948 USGS contour maps (Larsen et al., 2007) which contained large biases at higher elevations on glaciers (see, e.g., Arendt et al., 2002, *supplemental material*).

Owing to the nature of the instrument, the acquisition, and the topography in the region, there are holes/voids in the SRTM data on steep slopes due to shadowing and layover effects (e.g., Rignot et al., 2001). Filled SRTM products, such as the one distributed by CGIAR Consortium for Spatial Information (Jarvis et al., 2008), typically use the NED dataset to fill these gaps, which can introduce significant anomalies and discontinuities into the on-glacier elevations. As these holes are typically small and confined to the glacier margins in steep-sloped areas, we used the ~~un-filled~~ non-void-filled SRTM dataset and update glacier areas in our calculations (when necessary) to ignore these no-data regions, essentially assuming they do not belong to the glacier; these data gaps correspond to $<$ 2.5% voids in the SRTM DEM for our glacier outlines and study region. These original SRTM voids will thus not affect our sensitivity analysis on estimates of volume change.

2.2.2 IfSAR

As part of the Statewide Digital Mapping Initiative, the State of Alaska is producing an interferometric synthetic aperture radar (IfSAR) DEM of the entire state. The data are acquired from airborne radar operating in X-band and P-band, and are provided in a native resolution of 5 m mosaics. In our study area, flights were flown in summer 2012 and 2013. These data are available from the U.S. Geological Survey ~~, see . As of September 2017, 92% of the state has been covered through this initiative, with 57% of the statewide data available for download ()~~(USGS, 2019). The DEMs provided have a reported accuracy of 3 m, though on low-angle slopes the provided metadata indicate an accuracy closer to $\sim$1 m when compared to LiDAR swaths.

2.2.3 Glacier Outlines

We use the Randolph Glacier Inventory v6.0 data as a base to mask ~~glacier basins (RGI Consortium, 2017).~~ glaciers (RGI Consortium, 2017 . The outlines in this region are mostly based on imagery from the mid-2000s, so we have manually updated the outlines using mostly cloud-free Landsat scenes acquired in late summer in 1999 and 2001, in order to ensure our glacier outlines correspond to the SRTM date as close as possible.

As the IfSAR DEMs are only available over Alaska, and not adjacent areas in British Columbia and Yukon, we have selected only glaciers that fall 90% or more by area within Alaska. Additionally, we have removed any glaciers that fall 10% by area or more in both collection years, in order to ensure that we are using temporally consistent data to estimate volume change. ~~Finally, we~~ We also remove any glacier ~~basins~~ outlines that are smaller than 1 km^2, as for our purposes it is better to have a larger sample of on-glacier pixels to work with - voids over smaller glaciers result in more limited data to work with. Finally, we remove any glaciers for which we did not have a result for all methods (i.e., where the glacier is completely voided). This results in a total of ~~443 individual glacier basins~~ 415 individual glacier outlines used for the analysis.

3 Methods

We first calculate the "true" volume change by directly differencing the IfSAR and SRTM DEMs after co-registration following Nuth and Kääb (2011), and subsequently summing the elevation differences multiplied by pixel area within each glacier outline. We then calculate the regional volume change as the sum of these individual glacier volume changes. The code used to generate and interpolate the DEM voids, as well as the results for each glacier, can be found at https://github.com/iamdonovan/dem_voids.

Ordinarily, using DEMs derived from radar of different bands, especially those acquired in different seasons such as the SRTM (February) and IfSAR (typically August/September), would require a consideration of the effects of differential radar penetration in snow and ice, as well as a temporal correction accounting for the difference in season, before converting elevation changes to a mass balance value (Haug et al., 2009; Kronenberg et al., 2016). In this region, the SRTM is known to have particularly high levels of penetration that cause significant biases when used in geodetic mass balance calculations (Berthier et al., 2018). As our interest in this study is in isolating the effect of void interpolation methods on estimates of volume change, we ignore the differential penetration and temporal mismatch between our DEMs. We therefore highlight that biases will exist in the numbers provided in this study and do not recommend interpreting these relative estimates of glacier volume change.

3.1 Artificial Void Generation

In order to investigate the effects of ~~filling~~ interpolating voids, we first simulate voids in the (arbitrarily chosen) IfSAR DEM to reflect the distribution and size of voids that might be expected in DEMs derived from optical stereo sensors. Correlation masks from 99 MicMac ASTER (MMASTER)-processed stereo scenes (Girod et al., 2017) provides the basis for void simulation as low correlation areas represent failure of the stereographic reconstruction and elevation determination. We thus use areas of low correlation in the ASTER scenes to mimic voids, providing a way to ensure that our artificial voids are similar to what would normally be seen in DEMs derived from spaceborne optical stereo sensors.

We average and mosaic the 99 ASTER correlation masks together, and select a ~~mean~~ correlation threshold of ~~0.5~~ 50% to serve as the lower bound for acceptable ~~correlations~~data. This choice of threshold is based on a visual inspection of the mask produced, and the desire to mimic the ASTER data as much as possible. To further investigate the effects of interpolation method on the estimates of volume change, we also increase the threshold to ~~0.7~~70%, 80%, 90%, and 95%, comparing the differences for a select few interpolation schemes. For each threshold value, we apply the resulting mask to the IfSAR DEMs, producing voids as shown in Fig. 2.

3.2 Void FillingInterpolation

The following is a brief summary of the different methods used to fill the artificially-generated voids in the DEM and dDEM products. We have split the methods into three general categories, “constant” interpolation, “spatial” interpolation, and “hypsometric” interpolation.

3.2.1 Constant Methods

For the so-called “constant” interpolation methods, we calculate the mean (or median) elevation differences of the non-void pixels for each glacier basinoutline, then multiply this value by the area of the glacier basinoutline, thereby obtaining an average volume change for the glacier basinoutline. Examples of this method in the literature include Surazakov and Aizen (2006); Paul and Haeberli (2008); Fischer et al. (2015).

3.2.2 Spatial Methods

1. *Interpolation of elevation*. This method, applied to the DEM containing voids (here, the IfSAR DEMs), interpolates raw elevation values of the surrounding pixels to fill voids. The resulting interpolated DEM is differenced from the second DEM, followed by calculation of the volume changes. Examples of this approach can be found in Kääb (2008); Pieczonka et al. (2013); Pieczonka and Bolch (2015). Though Pieczonka and Bolch (2015) uses ordinary kriging to fill gaps in the original DEMs, we choose to use linear interpolation because the voids over the glaciers are relatively small, and bilinear interpolation for further comparison with the results of Kääb (2008).

2. *Interpolation of elevation differences*. Two original, unfilled DEMs are differenced to create a DEM difference (dDEM). Then, the voids in the dDEM are filled using bilinear interpolation. An example of this approach can be found in Kääb (2008); Zheng et al. (2018).

3. *Mean elevation difference in 1 km radius*. For each void pixel, we calculate the average elevation difference based on on-glacier pixels within a 1 km radius of the void pixel. Examples of this approach can be found in Melkonian et al. (2013, 2014).

3.2.3 Hypsometric Methods

The so-called “hypsometric” methods are based on the assumption that there is a relationship between elevation change and elevation. They can be further sub-divided into “global” and “local” approaches, depending on whether the mean is calculated using data from the entire region (i.e., “global”) or for an individual glacier basin only (i.e., “local”). We have chosen this terminology to be consistent with the terms used for other forms of interpolation. The global approach is often used by altimetry studies usually used to extrapolate measurements from only a few glaciers to a regional scale (e.g., Arendt et al., 2002; Johnson et al., 2013;

. (e.g., Arendt et al., 2002; Berthier et al., 2010; Kääb et al., 2012; Johnson et al., 2013; Nilsson et al., 2015), rather than to estimate an individual glacier's mass balance. Here, we use these methods to evaluate both individual and regional volume changes, in order to see the effects on individual glacier changes.

1. *Mean (or median) elevation difference by elevation bin*. Here, the original, unfilled DEMs are differenced, and the entire dDEM is binned according to the original elevation elevation in the earliest DEM for each pixel within the glacier outlines. The mean (or median) elevation difference for each bin is then calculated and multiplied by the area of each elevation bin to get a volume change. The sum of the volume change of each individual bin then gives the volume change for the glacier. This method is used by, e.g., Kääb (2008); Berthier et al. (2010); Gardelle et al. (2013); Papasodoro et al. (2015); Kronenberg et al. (2016); Brun et al. (2017); Dussaillant et al. (2018). If a glacier has an elevation range of 500 m or more, we use 50 m wide bins; otherwise we choose elevation bins that are 10% of the glacier elevation range. Additionally, where elevation bins are completely voided, we fill these bins using a third-order polynomial fit to the available data, so long as there is data over two thirds of the glacier elevation range.

2. *Polynomial fit to elevation difference by elevation bin*. The original, unfilled DEMs are differenced, and a polynomial is fit to the elevation differences as a function of the original elevation. This elevation curve is then integrated over the glacier hypsometery in order to calculate a volume change. Based on examples from the literature, such as Kääb (2008), we have chosen a third-order polynomial.

3.3 Uncertainties

To estimate the uncertainties in the true volume changes, we first co-register each DEM (SRTM, 2012 and 2013 IfSAR campaigns) to ICESat, using the method described by Nuth and Kääb (2011). We can then use the triangulation procedure described in Paul et al. (2017) to estimate the residual bias $\varepsilon_{\text{bias}}$ after co-registering the DEMs to each other; i.e. the uncertainty in correcting the mean bias between the DEMs. We also estimate the combined random error in elevation, $\varepsilon_{\text{rand}}$ by calculating the RMS difference of the population of dDEM pixels on stable ground. For each glacier, the error in volume change $\varepsilon_{\Delta V}$ can be estimated as:

$$\varepsilon_{\Delta V}^2 = (\varepsilon_{\Delta h} A)^2 + (\varepsilon_A \overline{\Delta h})^2, \qquad (1)$$

with A the glacier area, ε_A the error in glacier area (here conservatively assumed to be 10%; cf. Brun et al. (2017); Paul et al. (2017); Pfeffer), and $\overline{\Delta h}$ the mean elevation change on the glacier. To account for spatial autocorrelation, as well as the two sources of uncertainty in the elevation differences ($\varepsilon_{\text{bias}}$ and $\varepsilon_{\text{rand}}$), $\varepsilon_{\Delta h}$ can be written:

$$\varepsilon_{\Delta h} = \frac{\sqrt{\varepsilon_{\text{rand}}^2 + \varepsilon_{\text{bias}}^2}}{\sqrt{n/(L/r)^2}} \sqrt{\frac{\varepsilon_{\text{rand}}^2}{\sqrt{n/(L/r)^2}} + \varepsilon_{\text{bias}}^2}, \qquad (2)$$

where n is the number of pixels (i.e., measurements) that fall within the glacier outline, L is the autocorrelation distance (here assumed to be 500 m; cf. Brun et al. (2017); Magnússon et al. (2016); Rolstad et al. (2009)), and r is the pixel size (30 m).

Finally, we can combine equations (1) and (2) to obtain:

$$\varepsilon_{\Delta V} = \sqrt{\left(A\frac{\sqrt{\varepsilon_{\text{rand}}^2+\varepsilon_{\text{bias}}^2}}{\sqrt{n/(L/r)^2}}\right)^2 + \left(\varepsilon_A\overline{\Delta h}\right)^2}\sqrt{\left(A\sqrt{\frac{\varepsilon_{\text{rand}}^2}{\sqrt{n/(L/r)^2}}+\varepsilon_{\text{bias}}^2}\right)^2 + \left(\varepsilon_A\overline{\Delta h}\right)^2}. \tag{3}$$

4 Results and Discussion

4.1 Void Distribution

Fig. 3 shows the void and area frequency distributions per normalized glacier elevation binand the mean and median elevation difference per normalized elevation bin. Most glaciers (73.6%, 329 . For the 50% threshold case, most glaciers (64.4%, 268 glaciers) have a total void percentage below 20%, with only a small number (8%, 36 6.7%, 28 glaciers) having more than 40% voids. Voids are distributed similarly to glacier area with respect to normalized glacier elevation (i.e., the elevation divided by the elevation range), and most of the voids, as well as most of the glacier area, are found in the middle third of the glacier elevation range. These void and area distributions, along with the range of elevation changes, suggests that the middle third of the glacier elevation range is the most important to ensure correct estimation; that is, uncertainties introduced by interpolating over voids in the upper and lower thirds of the elevation range will be muted, owing to the typically smaller areas and percentage of voids in these ranges.

4.2 Individual GlaciersVariability of elevation change

The initial, non-voided maps of elevation differences for the 2012 and 2013 IfSAR acquisition areas are Fig. 3 also shows the mean and median elevation changes per normalized elevation bin, and the standard deviation of elevation changes. The highest elevation change variability on-glacier is in the lower portion of the glacier, where most of the dynamic change is occurring. Higher up in the accumulation area, the variability is much lower, and the mean differences are close to zero. These general patterns can be observed in the spatial distribution of elevation changes shown in Fig. 4. In general, the pattern of elevation change is elevation changes in the region are negative, especially at lower elevations, as noted in other studies (e.g., Larsen et al., 2007; Johnson et al., 2013; Melkonian et al., 2013, 2014; Berthier et al., 2018)(e.g., Larsen et al., 2007; Johnson et al., . Some exceptions include Margerie, Johns Hopkins, and Rendu Glaciers in the 2012 acquisition area, and Taku Glacier in the 2013 acquisition area (cf. see Fig. 1 for glacier locations). Margerie, Johns Hopkins, and Taku Glaciers are some of the few currently advancing tidewater glaciers in Alaska (e.g., McNabb and Hock, 2014; Motyka and Echelmeyer, 2003; Truffer et al., 2009) (e.g., Motyka and Echelmeyer, 2003; Truffer et al., 2009; McNabb and Hock, 2014), while Rendu Glacier has been previously identified as a surge-type glacier (Field, 1969). The pattern of elevation change shown on Rendu Glacier in the elevation difference mapsFig. 4, with thinning at higher elevations and pronounced thickening at lower elevations, is suggestive of a surge sometime between February 2000 and August 2012 (e.g., Raymond, 1987; Björnsson et al., 2003)(e.g., Raymond, 1987; Björnsson et al., 2 ; a tributary of Margerie Glacier also appears to have surged during this time period.

~~These contrasting patterns~~ The variability of elevation changes in the region is quite high, with a standard deviation of on-glacier elevation changes of 0.65 m a^{-1} for all glaciers in the region. This level of variability significantly smaller than other parts of Alaska (1.14 m a^{-1} for glaciers in Western Alaska; Le Bris and Paul, 2015), but significantly higher than regions in High Mountain Asia, where Brun et al. (2017) found intra-regional standard deviations in mass balance of $\sim$0.2 m.w.e a^{-1} ($\sim$0.24 m a^{-1} given their assumed density of 850$\pm$60 kg m^{3}). Compared to values estimated from ICESat (Nilsson et al., 2015) , this region is in line with Svalbard (0.7 m a $^{-1}$), higher than the Canadian Arctic (0.34/0.42 m a $^{-1}$ for North and South, respectively), and much lower than Iceland (1.14 m a $^{-1}$).

4.3 Impact of void interpolation on Individual Glacier Estimates

The variability of elevation gain and elevation loss ~~inform~~ shown in Fig. 4 informs some of the patterns shown in Fig. 5. Generally~~speaking~~, the global hypsometric methods are the farthest from the true values, which is perhaps not surprising in a region with a ~~variety of elevation change patterns such as this one. Glaciers~~ high variability of glacier elevation changes. As a result, glaciers that are far from the average ~~volume change of -0.11~~elevation change of -0.36 ~~km^{3}~~ m a^{-1} will tend to be far from the true volume change when the volume change is estimated with the regional values, as the data used do not reflect conditions at that particular glacier. ~~As a result, volume changes~~ Thus, interpolated estimates for glaciers losing much more ~~volume than the~~ than the regional average tends to be overestimated ~~, while volume changes~~ (i.e., less negative change), while interpolated estimates at glaciers that are losing less than the average, or even increasing in volume, tends to be underestimated (i.e., more negative/less positive change). Methods which use data from ~~a particular glacieroutline~~an individual glacier, or in a small area close to the particular glacier outline, tend to do a much better job of reproducing volume changes over each of these glaciers than do these global methods.

~~We also see little overall difference between the two linear interpolation methods. Kääb (2008) estimated the difference between these two methods for glaciers on Edgeøya, Svalbard, to be 1$\pm$12 m RMS. If we convert the volume change estimates to a mean elevation change for the glaciated areas, we obtain a mean difference of 0.00$\pm$0.01 m for these two methods. This difference is due to the fact that Kääb interpolated between contour lines over an entire ice cap, whereas we interpolate over much smaller areas that are confined by mountains, with much smaller differences on either side of a void. Thus, the effects of the different interpolation methods are muted. Additionally, Kääb used contour lines derived from aerial images with low contrast at higher elevations, which are likely biased as a result. We would most likely see similar results if we had used the NED DEM as reference, rather than the SRTM, as the NED was produced from similarly low-contrast aerial images.~~

The statistical summary for the difference in volume change estimates over all glaciers individually (Table 1) shows that on average, mean and median differences to the true values are generally low (<~~0.001~~0.1 ~~km^{3}~~m a^{-1}), as are root mean square (RMS) values (typically <0.2 ~~km^{3},~~ m a^{-1} with the exception of the global hypsometric methods). ~~This pattern does not hold, however, for the larger glaciers, which tend to have a much larger spread in estimates of volume change, and therefore appear to be more sensitive to the various methods.~~ The percentage of estimates that fall within the uncertainty range of the true volume change estimates for most of the methods is quite high, ~~around~~ above 95%. One notable exception is the median multiplied by glacier area method described in section 3.2.1, which aside from the global hypsometric methods, shows the fewest number of glaciers for which the interpolated value falls within the uncertainty (~~77.85~~83.89%), ~~shows~~ the largest individual overestimation at ~~2.32~~1.22 ~~km^3~~m a^{-1}, the largest mean and standard deviation (~~0.02~~0.07±~~0.15~~0.19 ~~km^3), and~~ m a^{-1}), the largest RMS difference (~~0.15~~0.20 ~~km^3~~m a^{-1}), and the worst agreement with the regional volume change estimate (~~8.3~~0.69 km^3 a^{-1} overestimation).

Fig. 6 shows the elevation change over Taku Glacier, with holes filled in for the nine non-constant methods. As mentioned previously, the spatial methods (Fig. 6b-d) and the local hypsometric methods (Fig. 6h-j) show the most similarity to the original elevation changes (Fig. 6a), with some subtle differences. The hypsometric methods have the effect of smoothing out the patterns of elevation change, whereas the spatial interpolation methods tend to preserve the original spatial patterns within elevation bands. Near the dividing lines between glacier ~~basins~~outlines, discontinuities can be seen in the local hypsometric maps, compared to the gradual changes across dividing lines seen in the original elevation changes and the spatially-interpolated maps. This suggests that the choice of glacier ~~basin~~ outlines can have an impact on the resulting volume change estimates. Finally, the global hypsometric methods (Fig. 6e-g), taking data from the region, do not faithfully reproduce the anomalous elevation change patterns for Taku Glacier.

For the largest 20 glaciers in the dataset (all >100 km^2), which represent 61% of the total glacier area for the glaciers studied, as well as 68.8% of the volume change in the region (a total of -49.9 km^3), we see a number of patterns related to each of the methods. Fig. 7 shows that for these largest glaciers, most of the methods fall within $\sim\pm$~~0.5~~0.3 ~~km^3~~ m a^{-1} of the true value, with significant outliers for some methods on some glaciers. For example, each of the methods for glacier RGI60-01.27102 (Hole-in-the-Wall Glacier) are clustered quite close to the true value, with the exception of the global methods. Hole-in-the-Wall Glacier is directly adjacent to Taku Glacier, and is also slightly gaining mass, thus leading to the discrepancy with the regional averages. In general, the global methods under- or over- estimate volume change, with only a few cases where the results within the uncertainty of the true value. For another glacier, RGI60-01.21001 (Riggs Glacier), the non-global methods give a value within $\sim$~~0.15~~0.05 ~~km^3~~ m a^{-1} of the true value, while the global methods are still within $\sim$~~0.35~~0.25 ~~km^3~~m a^{-1}; the number of voids induced on this glacier are relatively small ~~overall~~ (19% of the glacier area), and the elevation ~~change pattern for this glacier is also in line with~~ changes on this glacier are also similar to the regional trends (strong elevation loss at lower elevations, small gain at higher elevations).

Fig. 8 shows a box plot of the distribution for each method~~for the glaciers shown in Fig. 7. Again, we see that the best-performing methods, based~~ , using both the largest glaciers (Fig. 8a), and all glaciers (Fig. 8b). Based on the size of the interquartile range and ~~then~~ the mean difference of each interpolated ~~volume change estimate,~~ estimate, the best-performing methods are the spatial interpolation methods, the local hypsometric ~~methods~~method, and the ~~mean dH constant method~~constant mean dH method, for both sets of glaciers. For all glaciers, the outlier range is also the smallest for the local hypsometric methods, linear interpolation of dH, constant mean dH, and the 1 km neighborhood.

Table 2 shows the differences to the true volume change for two glaciers with some of the largest deviations from the true values. The median elevation change estimate for Field Glacier has the largest overall change from the true value for the non-global methods, at ~~+2.32~~0.92 ~~km^3~~m a^{-1}, while the global ~~fits~~ methods for Taku Glacier have some of the largest negative changes, all ~~over -3.50~~below -0.50 ~~km^3~~m a^{-1}. These differences are most likely for the reasons discussed above: the data being used to estimate volume change for Taku Glacier are far more negative than reality. For Field Glacier, only the median elevation change method and the global methods perform particularly poorly; the rest are all within ±~~0.36~~0.15 ~~km^3~~ $\mathrm{m\,a^{-1}}$ of the true estimate of ~~-3.02~~-1.15 ~~km^3 (12~~$\mathrm{m\,a^{-1}}$ ($\sim$13%). As shown in Fig. 9, this is most likely because of the heavy slant towards very negative elevation changes in the elevation change distribution for Field Glacier. Those positive values of elevation change found on the glacier are relatively small as compared to the negative values, and so the median is pulled heavily towards zero, greatly underestimating the volume change.

Based on the results for individual glaciers, the best methods (i.e., that introduce the least uncertainty/bias and the estimates closest to the original, non-voided estimates) appear to be linear interpolation of elevation change, the local mean hypsometric approach, and the ~~1km~~ 1 km neighborhood approach.

4.4 Impact of void interpolation on Regional ~~Totals~~Total

While the differences to the true values, when averaged over all glaciers, tends to be close to zero, the differences in the regional estimates can vary substantially, as shown in Table 1. The methods that came closest to the "true" volume change for the region were local mean hypsometric method, linear interpolation of elevation differences, and the global mean hypsometric method, which all yielded estimates within ~~0.4~~0.3 $\mathrm{km}^3\,\mathrm{a}^{-1}$ (0.8%) of the regional total. A form of the global mean hypsometric method is one that is often used in altimetry-based studies to extrapolate measurements to unsurveyed glaciers, either using absolute or ~~relative elevation (e.g., Arendt et al., 2002; Kääb et al., 2012; Johnson et al., 2013; Larsen et al., 2015) ,~~ normalized elevation (e.g., Arendt et al., 2002; Kääb et al., 2012; Johnson et al., 2013; Larsen et al., 2015) and this result would indicate that relatively little bias is introduced to the regional estimate through this form of extrapolation, at least for these particular datasets. In this study, we have used absolute elevations, rather than normalized elevations, for the global methods. In other regions, it may be worth comparing the differences between using absolute and normalized elevations.

The next best estimates after the three closest were the local median and polynomial hypsometric methods, linear interpolation of elevation, and the 1 km ~~average~~ neighborhood method, all coming within ~~2~~0.15 $\mathrm{km}^3\,\mathrm{a}^{-1}$ (4%) of the regional total. One explanation for the ~~value for~~ overall worse performance of the elevation interpolation method ~~is as~~ versus linear interpolation of elevation change is discussed in Kääb (2008): elevations on the glacier surface are not necessarily self-similar in a given area, and elevations can vary greatly even on relatively small length scales. As for the 1 km ~~neighbourhood~~ neighborhood method, it may be that 1 km is too large of an area to try to average over for some glaciers in this region, or it may be that the ~~neighbourhood~~ neighborhood window used is including values from ~~neighbouring~~ neighboring glaciers that have very different patterns of elevation change, thus behaving more like a ~~"global"~~ "global" method in some areas.

The methods that came the farthest from the regional total were the constant median elevation change method, as discussed above, and the global polynomial hypsometric method, both over/underestimating the regional total volume change by over ~~8~~0.6 $\mathrm{km}^3\,\mathrm{a}^{-1}$, well above the uncertainty of ~~5.21~~0.4 $\mathrm{km}^3\,\mathrm{a}^{-1}$. In contrast to this, estimating volume changes using the global median hypsometric method overestimated the regional total volume change by ~~4.47~~0.37 $\mathrm{km}^3\,\mathrm{a}^{-1}$. While for an entire glacier~~basin~~, the median elevation change skews very heavily towards zero due to the asymmetry in positive and negative values of elevation change, this is not necessarily the case for an elevation bin. As noted by Kääb (2008), and borne out by the elevation change interpolation method, elevation changes tend to be rather self-similar on small spatial scales, and the median change for an elevation bin tends to be a more accurate reflection of the actual elevation change. That said, for both the local and global methods, using the mean rather than the median yields a better result on both an individual and regional basis, suggesting that the mean is more representative as a rule.

4.5 Increasing void area

~~By increasing the threshold used to induce voids from 0.5 to 0.7, we were able to estimate how sensitive~~ To estimate the sensitivity of the different methods ~~are to the size~~ to the amount of voids in the DEMs~~. As might be expected, the overall number of glaciers for which the~~ , we varied the correlation threshold from 35% to 95%. The effect of using each threshold on the mean percent void for all glaciers is shown in Fig. 10. Above a threshold of 70%, the mean void percentage per glacier increases dramatically, up to 75% voids when using a threshold of 95%. In the following analysis, we compare the total set of interpolated volume changes ~~fall within the uncertainty drops across all methods, though not completely evenly. The global methods are rather insensitive to the increase in voids, likely because the increase is muted on the regional scale as compare to the individual glacier scale~~ for all glaciers over all threshold scenarios. We have limited the number of methods discussed to those that performed the best in the 50% threshold case (described in sections 4.3 and 4.4): the constant mean method, linear interpolation of elevation changes, the 1 km neighborhood method, the global mean hypsometric method, and the local mean and median hypsometric methods.

~~For the largest 20 glaciers~~ Fig. 11 shows that the local hypsometric and spatial interpolation methods can tolerate a high void percentage (>50%) before more than 10% of the estimates fall outside of the uncertainty range; beyond 70% voids, this percentage increases dramatically. The constant mean dH method does not perform as well for lower void percentages, but it does not drop as quickly at higher void percentages as the other methods. As expected, the global mean hypsometric method is low throughout the range of void percentages, though there is not as much dependence on the void percentage as with the other methods. Even at the highest void percentages, the spatial interpolation methods perform remarkably well, with upwards of 75% of estimates falling below the uncertainty range for both linear interpolation of elevation change and the 1 km neighborhood; this may not hold for other datasets, which we discuss more in section 4.6.

Based on the results for the 50% threshold case, we compared the three methods which gave the best results on a regional basis: linear interpolation of elevation change and the global and local mean hypsometric methods, with results shown in Fig. ~~7, Fig. **??** shows the difference in volume change estimate for each of the methods. Overall, most glaciers do not have large differences for most of the methods (<0.15 km^3). The global methods tend to show the least~~ 12. The global mean hypsometric method shows little variation overall, with differences to truth generally negative and with a large standard deviation. As shown in Fig. 11, linear interpolation of elevation change and the local mean hypsometric method remain close to the true values of volume change up to around 50-60% voids, before the standard deviation increases dramatically, though less so for linear interpolation of elevation change. As the local hypsometric method requires data in a given elevation bin for interpolation, it makes sense that with higher void percentage, the interpolated estimates are further and further from the true values of volume change, while linear interpolation requires less data to provide an estimate; as long as the missing values are similar enough to the ~~constant methods and spatial methods tend to show larger, though still relatively small, differences~~non-voided values, the interpolated estimates of volume change do not deviate significantly from the true volume changes.

4.6 ASTER differences

One caveat remains when using methods such as linear interpolation of elevation differences. To illustrate this, we used ASTER DEMs acquired on 13 August 2015 over a portion of the 2012 ~~IFSAR~~ IfSAR acquisition area, and differenced these DEMs to the SRTM. The ASTER DEMs were processed using MMASTER, and along-track and cross-track biases were corrected using the 2012 IfSAR DEM (see Girod et al., 2017, for more details on these corrections).

Compared to the IfSAR DEM, the ASTER DEMs are quite noisy in the accumulation areas of glaciers, owing to the low contrast, and hence low correlation, between the original images in the ASTER scenes. As such, even after correlation masking, there is significant noise in the DEM difference map (Fig. 13a). When these values are linearly interpolated, the resulting dDEM shows clear interpolation ~~artefacts~~ artifacts and elevation changes that differ greatly from the original IfSAR/SRTM differences ~~, biasing the estimated volume changes~~ (Fig. 13b)~~.~~, biasing the estimated volume changes.

For the 91 glaciers covered by these ASTER DEMs, the other "best" estimates named above (local mean hypsometric and global mean hypsometric) ~~provide a~~ yield a total volume change estimate of $\sim$0 ~~km^3~~m a^{-1}, whereas linear interpolation of elevation differences yields a volume change of ~~3.4~~+0.2 ~~km^3~~m a^{-1}. Looking further at this, this discrepancy is almost entirely due to one glacier, Johns Hopkins - linear interpolation of elevation changes yields a volume change estimate of ~~3.6~~$\sim$1 ~~km^3~~ m a^{-1} for this glacier, while the other estimates give only $\sim$~~2~~0.6 ~~km^3.~~ m a^{-1}, in line with the IfSAR-SRTM estimate of $\sim$0.44 m a^{-1}.

Thus, we caution against using a direct linear interpolation of elevation differences to fill voids without first filtering or otherwise removing potential outliers, ~~or~~ which are often located near voids, which increases their influence in a linear interpolation. We also caution against using this approach when the distances between known values are quite large in relation to the glacier width~~; that said, the~~. The local mean hypsometric approach used by many studies performs just as well as linear interpolation of elevation differences ~~,~~ in the idealized case analyzed here, and therefore appears to be more robust against this kind of noise, and is easily implemented in place of linear interpolation.

A question then arises: when using noisy, 'real-world' data such as ASTER DEMs, is it better to keep only the most reliable values for a given DEM, potentially producing large data gaps that must be interpolated, or is it better to have a more complete DEM? Given the results presented in this section, and the results shown in section 4.5, we suggest with relatively low correlation (i.e., reliability) can still have usable data. As the local hypsometric methods tend to be more robust against noise, and can tolerate a rather high percentage of data voids (up to $\sim$60%; Fig. 12), a strategy of using lower correlation thresholds ($\sim$50-60%) in combination with the local hypsometric approach seems well-suited to making the most use of the available data.

5 Conclusions

We have compared 11 different methods for ~~filling~~ interpolating voids in DEM difference maps over glaciers, and compared the effects of these different methods on estimates of glacier volume change. ~~Three~~ Two methods, linearly interpolating elevation changes ~~, the so-called~~ and the local mean hypsometric method, ~~and the so-called~~ performed well on an individual glacier basis, producing estimates within the uncertainty of the original estimates. These two methods, as well as a third, the global mean hypsometric method, ~~performed remarkably similarly~~ also performed quite well in estimating the regional total volume change, differing from the true estimate by less than one percent~~; the first two methods also performed well on an individual glacier basis~~. For the input data we have used, linearly interpolating elevation differences tends to produce elevation change maps that look the most similar to the original maps. This may not hold, however, for voids that take up a larger portion of the glacier area, where the assumption that elevation changes are similar over small distances may be violated, and interpolation ~~artefacts~~ artifacts would introduce larger uncertainties. Additionally, this may not hold for DEMs that are noisier, especially in low-contrast areas such as the accumulation zones of glaciers; as such, we caution against adopting this method without first considering the characteristics of the DEMs beings used. In terms of individual glacier estimates, using the mean elevation change per elevation bin multiplied by the glacier hypsometry performs quite well, which perhaps explains its widespread use in studies of glacier volume change and geodetic mass balance.

Taken on average, most of the methods perform well for the glaciers in the region, with low mean, median, and ~~RMS~~ root mean square differences for all methods, though large outliers skew the differences in the regional totals. Using the median elevation change for a glacier, multiplied by the glacier area, tends to work quite poorly, owing to the asymmetrical distribution of positive and negative elevation change values; i.e., the glaciers in the region tend to have significantly more negative values of elevation change than positive values. Unless there is good reason to think the distribution of elevation changes for a particular glacier or region is more symmetrical, this method should be avoided; the same can be said for using a median hypsometric approach, which does not ~~do~~ perform as well as the mean hypsometric approaches, although this may not always be the case when there is significant noise in the original DEMs. As might be expected, using regional data to estimate the volume change of an individual glacier quite often ~~does~~ performs poorly, though the regional total volume change can be well-approximated in this way.

The bias introduced by a given method is also dependent on the size of the data voids. For the two most accurate methods on both an individual glacier and regional basis, interpolating voids of up to 50% tends to introduce small differences in the estimated volume change; most of the change is happening in the lower parts of the glacier where the void percentage is smaller. Above 60-70% voids, however, the errors introduced grow substantially, and are usually significantly higher than the uncertainty in the original datasets. This is not the case for the global interpolation methods, however, which have large errors for individual glaciers that are mostly independent of the void percentage. Thus, the void percentage does not have as pronounced an effect on the regional total estimated using the global mean hypsometric method, implying that its use in regional-scale altimetry studies is well-founded.

In summary, the effect of DEM voids on estimates of geodetic mass balance depends on the size of the voids, the magnitude and spatial pattern of changes on the glaciers, as well as the nature of the DEMs used. The choice of ~~void-filling~~ void interpolation method is important, and if not considered properly, biases many times the uncertainty of the volume change measurement can be induced, while on the regional level, biases of up to 20% can be induced. The choice of “best method” will depend on the ultimate goal of the study, as well as the nature of the voids in the DEMs and the changes of the glaciers. Interpolation methods using elevation differences from an individual glacier, or differences within a close proximity to an individual glacier (in the case of glacier complexes), tend to be the most accurate and robust. If the DEMs used have significant noise, or have large holes, however, linear interpolation may not be suitable. If attempting to estimate geodetic mass balance for unsurveyed glaciers, as is needed in many altimetry-based studies, only a global method will suffice, though the mass balance estimate for a given unsurveyed glacier should not be taken at face value. Additionally, the regional bias for such a case may be strongly biased. As each of these different methods are relatively easily ~~implented~~implemented, however, a comparison of the different methods should be attempted in order to provide a measure of the uncertainty introduced by ~~filling~~ interpolating holes in the data.

Code availability. The code used to generate and fill voids, as well as the resulting data for each glacier, can be found in a git repository at https://github.com/iamdonovan/dem_voids

Competing interests. The authors declare that no competing interests are present.

Acknowledgements. This study was funded in part by the ESA project Glaciers_cci (4000109873/14/I-NB) and the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013)/ERC grant agreement no. 320816. Alaska IfSAR DEM data are provided through State of Alaska’s Statewide Digital Mapping Initiative and the US Geological Survey’s Alaska Mapping Initiative. Both the IfSAR and the SRTM DEM are distributed through the USGS Earth Resources Observation Center. Acquisition of ASTER images ~~ws~~ was guided by NASA JPL, though the ASTER science team and the Global Land Ice Measurements from Space (GLIMS) initiative. The authors wish to thank Frank Paul and Stephen Plummer for their helpful comments on early versions of the manuscript. Nick Barrand and two anonymous reviewers provided many insightful comments which helped improve the quality of the manuscript.

References

Abermann, J., Fischer, A., Lambrecht, A., and Geist, T.: On the potential of very high-resolution repeat DEMs in glacial and periglacial environments, The Cryosphere, 4, 53–65, https://doi.org/10.5194/tc-4-53-2010, 2010.

Andreassen, L. M., Elvehøy, H., Kjøllmoen, B., and Engeset, R. V.: Reanalysis of long-term series of glaciological and geodetic mass balance for 10 Norwegian glaciers, The Cryosphere, 10, 535–552, https://doi.org/10.5194/tc-10-535-2016, 2016.

Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S., and Valentine, V. B.: Rapid Wastage of Alaska Glaciers and Their Contribution to Rising Sea Level, Science, 297, 382–386, https://doi.org/10.1126/science.1072497, http://www.sciencemag.org/cgi/content/abstract/sci;297/5580/382http://www.sciencemag.org/cgi/doi/10.1126/science.1072497, 2002.

Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S., Zirnheld, S. L., Valentine, V. B., Ritchie, J. B., and Druckenmiller, M.: Updated estimates of glacier volume changes in the western Chugach Mountains, Alaska, and a comparison of regional extrapolation methods, Journal of Geophysical Research, 111, https://doi.org/10.1029/2005JF000436, http://www.agu.org/pubs/crossref/2006/2005JF000436.shtml, 2006.

Bamber, J. L. and Rivera, A.: A review of remote sensing methods for glacier mass balance determination, Global and Planetary Change, 59, 138–148, https://doi.org/doi:10.1016/j.gloplacha.2006.11.031, 2007.

Berthier, E., Schiefer, E., Clarke, G. K. C., Menounos, B., and Rémy, F.: Contribution of Alaskan glaciers to sea-level rise derived from satellite imagery, Nature Geoscience, 3, 92–95, https://doi.org/10.1038/ngeo737, http://www.nature.com/doifinder/10.1038/ngeo737, 2010.

Berthier, E., Larsen, C. F., Durkin, W., Willis, M., and Pritchard, M.: Brief communication: Unabated wastage of the Juneau and Stikine icefields (southeast Alaska) in the early 21st century, The Cryosphere Discussions, 1, 1–12, https://doi.org/10.5194/tc-2017-272, https://www.the-cryosphere-discuss.net/tc-2017-272/, 2018.

Björnsson, H., Pálsson, F., Sigurðsson, O., and Flowers, G. E.: Surges of glaciers in Iceland, Annals of Glaciology, 36, 82–90, https://doi.org/10.3189/172756403781816365, 2003.

Brun, F., Berthier, E., Wagnon, P., Kääb, A., and Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nature Geoscience, 10, 668–673, https://doi.org/10.1038/ngeo2999, http://www.nature.com/doifinder/10.1038/ngeo2999, 2017.

Cogley, J. G.: Geodetic and direct mass-balance measurements: comparison and joint analysis, Annals of Glaciology, 50, 96–100, https://doi.org/10.3189/172756409787769744, 2009.

Dall, J., Madsen, S. N., Keller, K., and Forsberg, R.: Topography and Penetration of the Greeland Ice Sheet Measured with Airborne SAR Interferometry, Geophysical Research Letters, 28, 1703–1706, 2001.

Dehecq, A., Millan, R., Berthier, E., Gourmelen, N., Trouvé, E., and Vionnet, V.: Elevation changes inferred from TanDEM-X data over the Mont-Blanc area: Impact of the X-band interferometric bias, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9, 3870–3882, 2016.

Dussaillant, I., Berthier, E., and Brun, F.: Geodetic Mass Balance of the Northern Patagonian Icefield from 2000 to 2012 Using Two Independent Methods, Frontiers in Earth Science, 6, 8, https://doi.org/10.3389/feart.2018.00008, 2018.

Elsberg, D. H., Harrison, W. D., Echelmeyer, K. A., and Krimmel, R. M.: Quantifying the effects of climate and surface change on glacier mass balance, Journal of Glaciology, 47, 649–658, https://doi.org/10.3189/172756501781831783, http://openurl.ingenta.com/content/xref?genre=article{&}issn=0022-1430{&}volume=47{&}issue=159{&}spage=649, 2001.

Field, W. O.: Current observations on three surges in Glacier Bay, Alaska, 1965-1968, Canadian Journal of Earth Sciences, 6, 831–839, 1969.

Fischer, M., Huss, M., and Hoelzle, M.: Surface elevation and mass changes of all Swiss glaciers 1980-2010, The Cryosphere, 9, 525–540, https://doi.org/10.5194/tc-9-525-2015, 2015.

Gardelle, J., Berthier, E., and Arnaud, Y.: Impact of resolution and radar penetration on glacier elevation changes computed from DEM differencing, Journal of Glaciology, 58, 419–422, https://doi.org/10.3189/2012JoG11J175, 2012.

Gardelle, J., Berthier, E., Arnaud, Y., and Kääb, A.: Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999-2011, The Cryosphere, 7, 1263–1286, https://doi.org/10.5194/tc-7-1263-2013, 2013.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F.: A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009, Science, 340, 852–857, http://www.sciencemag.org/content/340/6134/852.short, 2013.

Geist, T., Elvehoy, H., Jackson, M., and Stötter, J.: Investigations on intra-annual elevation changes using multi-temporal airborne laser scanning data: case study Engabreen, Norway, Annals of Glaciology, 42, 195–201, 2005.

Girod, L., Nuth, C., Kääb, A., McNabb, R. W., and Galland, O.: MMASTER: improved ASTER DEMs for elevation change monitoring, Remote Sensing, 7, 704–729, https://doi.org/10.3390/rs9070704, 2017.

Haug, T., Rolstad, C., Elvehøy, H., Jackson, M., and Maalen-Johansen, I.: Geodetic mass balance of the western Svartisen ice cap, Norway, in the periods 1968–1985 and 1985–2002, Annals of Glaciology, 50, 119–125, 2009.

Huss, M.: Density assumptions for converting geodetic glacier volume change to mass change, The Cryosphere, 7, 877–887, https://doi.org/DOI 10.5194/tc-7-877-2013, 2013.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled seamless SRTM data v4, Tech. rep., International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org, 2008.

Johnson, A. J., Larsen, C. F., Murphy, N., Arendt, A. A., and Zirnheld, S. L.: Mass balance in the Glacier Bay area of Alaska, USA, and British Columbia, Canada, 1995-2011, using airborne laser altimetry, J. Glaciol., 59, 632–648, https://doi.org/10.3189/2013JoG12J101, http://openurl.ingenta.com/content/xref?genre=article&issn=0022-1430&volume=59&issue=216&spage=632, 2013.

Kääb, A.: Glacier Volume Changes Using ASTER Satellite Stereo and ICESat GLAS Laser Altimetry. A Test Study on Edgeøya, Eastern Svalbard, IEEE Transactions on Geoscience and Remote Sensing, 46, 2823–2830, 2008.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488, 495–498, https://doi.org/10.1038/nature11324, 2012.

Kohler, J., James, T. D., Murray, T., Nuth, C., Brandt, O., Barrand, N. E., Aas, H. F., and Luckman, A. J.: Acceleration in thinning rate on western Svalbard glaciers, Geophysical Research Letters, 34, 1–5, https://doi.org/10.1029/2007GL030681, 2007.

Kronenberg, M., Barandun, M., Hoelzle, M., Huss, M., Farinotti, D., Azisov, E., Usubaliev, R., Gafurov, A., Petrakov, D., and Kääb, A.: Mass-balance reconstruction for Glacier No. 354, Tien Shan, from 2003 to 2014, Annals of Glaciology, 57, 92–102, https://doi.org/10.3189/2016AoG71A032, 2016.

Larsen, C. F., Motyka, R. J., Arendt, A. A., Echelmeyer, K. A., and Geissler, P. E.: Glacier changes in southeast Alaska and northwest British Columbia and contribution to sea level rise, Journal of Geophysical Research, 112, https://doi.org/10.1029/2006JF000586, http://www.agu.org/pubs/crossref/2007/2006JF000586.shtmlhttp://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:Glacier+changes+in+southeast+Alaska+and+northwest+British+Columbia+and+contribution+to+sea+level+rise{#}0, 2007.

Larsen, C. F., Burgess, E. W., Arendt, A. A., O'Neel, S., Johnson, A. J., and Kienholz, C.: Surface melt dominates Alaska glacier mass balance, Geophysical Research Letters, 42, 5902–5908, https://doi.org/10.1002/2015GL064349, 2015.

Le Bris, R. and Paul, F.: Glacier-specific elevation changes in parts of western Alaska, Annals of Glaciology, 56, 184–192, https://doi.org/10.3189/2015AoG70A227, 2015.

Magnússon, E., Muñoz-Cobo Belart, J., Pálsson, F., Ágústsson, H., and Crochet, P.: Geodetic mass balance record with rigorous uncertainty estimates deduced from aerial photographs and lidar data – Case study from Drangajökull ice cap, NW Iceland, The Cryosphere, 10, 159–177, https://doi.org/10.5194/tc-10-159-2016, http://www.the-cryosphere.net/10/159/2016/, 2016.

McNabb, R. W. and Hock, R.: Alaska tidewater glacier terminus positions, 1948-2012, Journal of Geophysical Research, 119, 153–167, https://doi.org/10.1002/2013JF002915, 2014.

Melkonian, A. K., Willis, M. J., Pritchard, M. E., Rivera, A., Bown, F., and Bernstein, S. A.: Satellite-derived volume loss rates and glacier speeds for the Cordillera Darwin Icefield, Chile, The Cryosphere, 7, 823–839, https://doi.org/10.5194/tc-7-823-2013, http://www.the-cryosphere.net/7/823/2013/, 2013.

Melkonian, A. K., Willis, M. J., and Pritchard, M. E.: Satellite-derived volume loss rates and glacier speeds for the Juneau Icefield, Alaska, Journal of Glaciology, 60, 743–760, https://doi.org/10.3189/2014JoG13J181, 2014.

Motyka, R. J. and Echelmeyer, K. A.: Taku Glacier (Alaska, USA) on the move again: active deformation of proglacial sediments, Journal of Glaciology, 49, 50–58, 2003.

Nilsson, J., Sørensen, L. S., Barletta, V. R., and Forsberg, R.: Mass changes in Arctic ice caps and glaciers: Implications of regionalizing elevation changes, The Cryosphere, 9, 139–150, https://doi.org/10.5194/tc-9-139-2015, 2015.

Nuth, C. and Kääb, A.: Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change, The Cryosphere, 5, 271–290, https://doi.org/10.5194/tc-5-271-2011, http://www.the-cryosphere.net/5/271/2011/, 2011.

Papasodoro, C., Berthier, E., Royer, A., Zdanowicz, C., and Langlois, A.: Area, elevation and mass changes of the two southernmost ice caps of the Canadian Arctic Archipelago between 1952 and 2014, The Cryosphere, 9, 1535–1550, https://doi.org/10.5194/tc-9-1535-2015, 2015.

Paul, F. and Haeberli, W. O.: Spatial variability of glacier elevation changes in the Swiss Alps obtained from two digital elevation models, Geophysical Research Letters, 35, 1–5, https://doi.org/10.1029/2008GL034718, 2008.

Paul, F., Bolch, T., Briggs, K., Kääb, A., McMillan, M., McNabb, R., Nagler, T., Nuth, C., Rastner, P., Strozzi, T., et al.: Error sources and guidelines for quality assessment of glacier area, elevation change, and velocity products derived from satellite data in the Glaciers_cci project, Remote Sensing of Environment, 2017.

Pfeffer, W. T., Arendt, A. A., Bliss, A., Bolch, T., Cogley, J. G., Gardner, A. S., Hagen, J. O., Hock, R., Kaser, G., Kienholz, C., Miles, E. S., Moholdt, G., Mölg, N., Paul, F., Radić, V., Rastner, P., Raup, B. H., Rich, J. L., Sharp, M. J., Andreassen, L. M., Bajracharya, S., Barrand, N. E., Beedle, M. J., Berthier, E., Bhambri, R., Brown, I., Burgess, D. O., Burgess, E. W., Cawkwell, F., Chinn, T., Copland, L., Cullen, N. J., Davies, B. J., De Angelis, H., Fountain, A. G., Frey, H., Giffen, B. A., Glasser, N. F., Gurney, S. D., Hagg, W., Hall, D. K., Haritashya, U. K., Hartmann, G., Herreid, S. J., Howat, I. M., Jiskoot, H., Khromova, T. E., Klein, A., Kohler, J., König, M., Kriegel, D., Kutuzov, S., Lavrentiev, I., Le Bris, R., Li, X., Manley, W. F., Mayer, C., Menounos, B., Mercer, A., Mool, P., Negrete, A., Nosenko, G., Nuth, C., Osmonov, A., Pettersson, R., Racoviteanu, A. E., Ranzi, R., Sarikaya, M. A., Schneider, C., Sigurðsson, O., Sirguey, P., Stokes, C. R., Wheate, R., Wolken, G. J., Wu, L. Z., and Wyatt, F. R.: The randolph glacier inventory: A globally complete inventory of glaciers, Journal of Glaciology, 60, 537–552, https://doi.org/10.3189/2014JoG13J176, 2014.

Pieczonka, T. and Bolch, T.: Region-wide glacier mass budgets and area changes for the Central Tien Shan between ˜ 1975 and 1999 using Hexagon KH-9 imagery, Global and Planetary Change, 128, 1–13, https://doi.org/10.1016/j.gloplacha.2014.11.014, 2015.

Pieczonka, T., Bolch, T., Junfeng, W., and Shiyin, L.: Heterogeneous mass loss of glaciers in the Aksu-Tarim Catchment (Central Tien Shan) revealed by 1976 KH-9 Hexagon and 2009 SPOT-5 stereo imagery, Remote Sensing of Environment, 130, 233–244, https://doi.org/10.1016/j.rse.2012.11.020, http://dx.doi.org/10.1016/j.rse.2012.11.020, 2013.

Raymond, C. F.: How do glaciers surge? A review, Journal of Geophysical Research, 92, 9121–9134, https://doi.org/10.1029/JB092iB09p09121, http://www.agu.org/pubs/crossref/1987/JB092iB09p09121.shtmlhttp://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:How+Do+Glaciers+Surge?+A+Review{#}0, 1987.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0: Technical Report, https://doi.org/10.7265/N5-RGI-60, Global Land Ice Measurements from Space, Colorado, USA. Digital Media. , 2017.

Rignot, E. J., Echelmeyer, K. A., and Krabill, W. B.: Penetration depth of interferometric synthetic-aperture radar signals in snow and ice, Geophysical Research Letters, 28, 3501–3504, https://doi.org/10.1029/2000GL012484, http://doi.wiley.com/10.1029/2000GL012484, 2001.

Rolstad, C., Haug, T., and Denby, B.: Spatially integrated geodetic glacier mass balance and its uncertainty based on geostatistical analysis: Application to the western Svartisen ice cap, Norway, Journal of Glaciology, 55, 666–680, https://doi.org/10.3189/002214309789470950, 2009.

Shugar, D. H., Rabus, B. T., and Clague, J. J.: Elevation changes (1949 – 1995) of Black Rapids Glacier, Alaska, derived from a multi-baseline InSAR DEM and historical maps, Journal of Glaciology, 56, 625–634, 2010.

Surazakov, A. B. and Aizen, V. B.: Estimating volume change of mountain glaciers using SRTM and map-based topographic data, IEEE Transactions on Geoscience and Remote Sensing, 44, 2991–2995, https://doi.org/10.1109/TGRS.2006.875357, 2006.

Truffer, M., Motyka, R. J., Hekkers, M., Howat, I. M., and King, M. A.: Terminus dynamics at an advancing glacier: Taku Glacier, Alaska, Journal of Glaciology, 55, 1052–1060, https://doi.org/10.3189/002214309790794887, 2009.

USGS: USGS EROS Archive - Digital Elevation - Interferometric Synthetic Aperture Radar (IFSAR) - Alaska, https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-interferometric-synthetic-aperture-radar, accessed: 2019-01-17, 2019.

Van Zyl, J. J.: The shuttle radar topography mission (SRTM): A breakthrough in remote sensing of topography, Acta Astronautica, 48, 559–565, https://doi.org/10.1016/S0094-5765(01)00020-0, 2001.

WGMS: USGS EROS Archive - Digital Elevation - Interferometric Synthetic Aperture Radar (IFSAR) - Alaska, https://wgms.ch/data_databasestats/, accessed: 2019-01-17, 2019.

Zemp, M., Jansson, P., Holmlund, P., Gärtner-Roer, I., Koblet, T., Thee, P., and Haeberli, W. O.: Reanalysis of multi-temporal aerial images of Storglaciären, Sweden (1959-99) - Part 2: Comparison of glaciological and volumetric mass balances, The Cryosphere, 4, 345–357, https://doi.org/10.5194/tc-4-345-2010, http://www.the-cryosphere.net/4/345/2010/, 2010.

Zemp, M., Thibert, E., Huss, M., Stumm, D., Rolstad Denby, C., Nuth, C., Nussbaumer, S. U., Moholdt, G., Mercer, A., Mayer, C., Joerg, P. C., Jansson, P., Hynek, B., Fischer, A., Escher-Vetter, H., Elvehøy, H., and Andreassen, L. M.: Reanalysing glacier mass balance measurement series, The Cryosphere, 7, 1227–1245, https://doi.org/10.5194/tc-7-1227-2013, 2013.

Zheng, W., Pritchard, M. E., Willis, M. J., Tepes, P., Gourmelen, N., Benham, T. J., and Dowdeswell, J. A.: Accelerating glacier mass loss on Franz Josef Land , Russian Arctic, Remote Sensing of Environment, 211, 357–375, https://doi.org/10.1016/j.rse.2018.04.004, https://doi.org/10.1016/j.rse.2018.04.004, 2018.

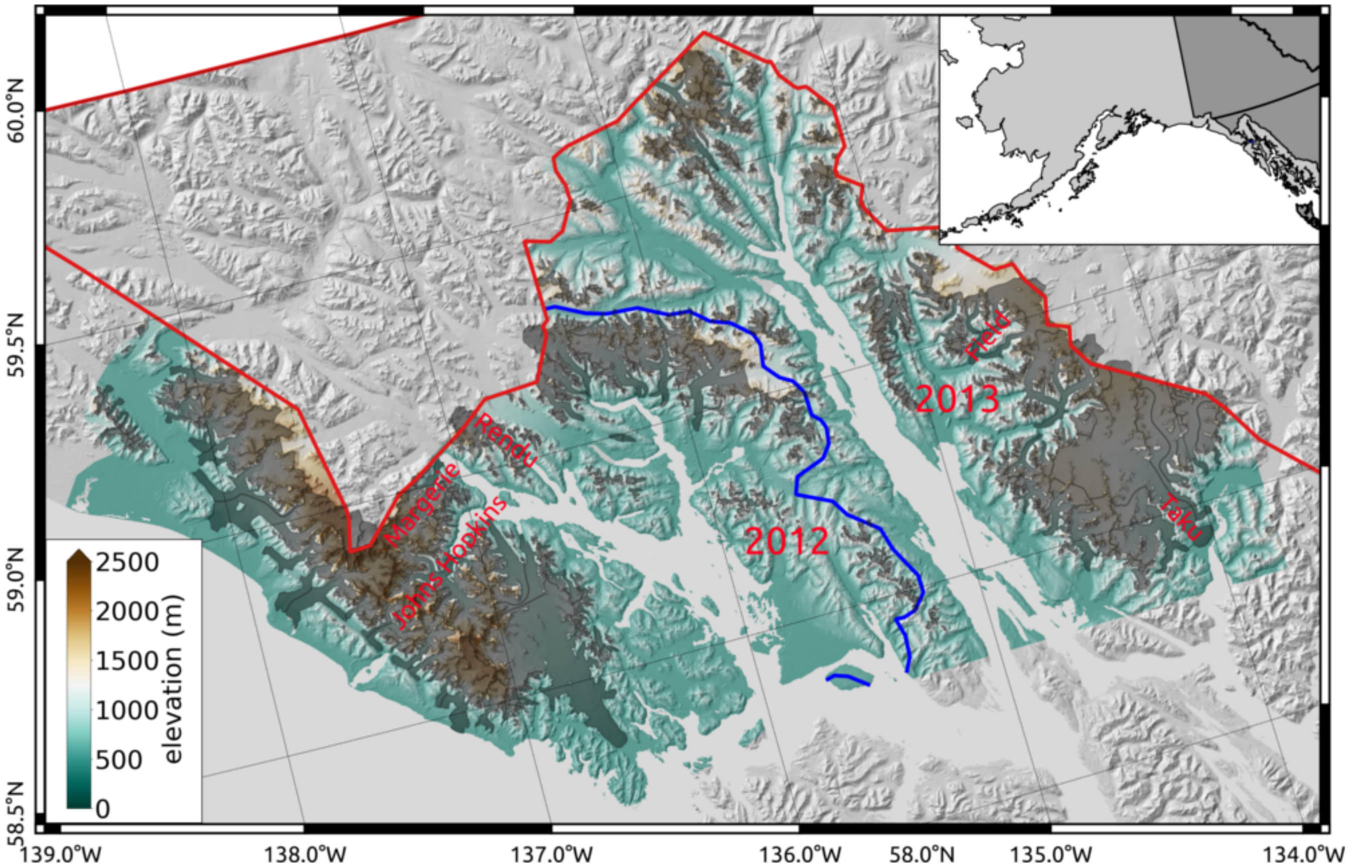

Figure 1. Study area in Southeast Alaska, USA. IfSAR DEMs are displayed overtop US NED and Canadian Digital Elevation Model (CDEM) hillshade. Blue line indicates boundary between 2012 and 2013 acquisitions. Named glaciers are discussed further in the following sections.

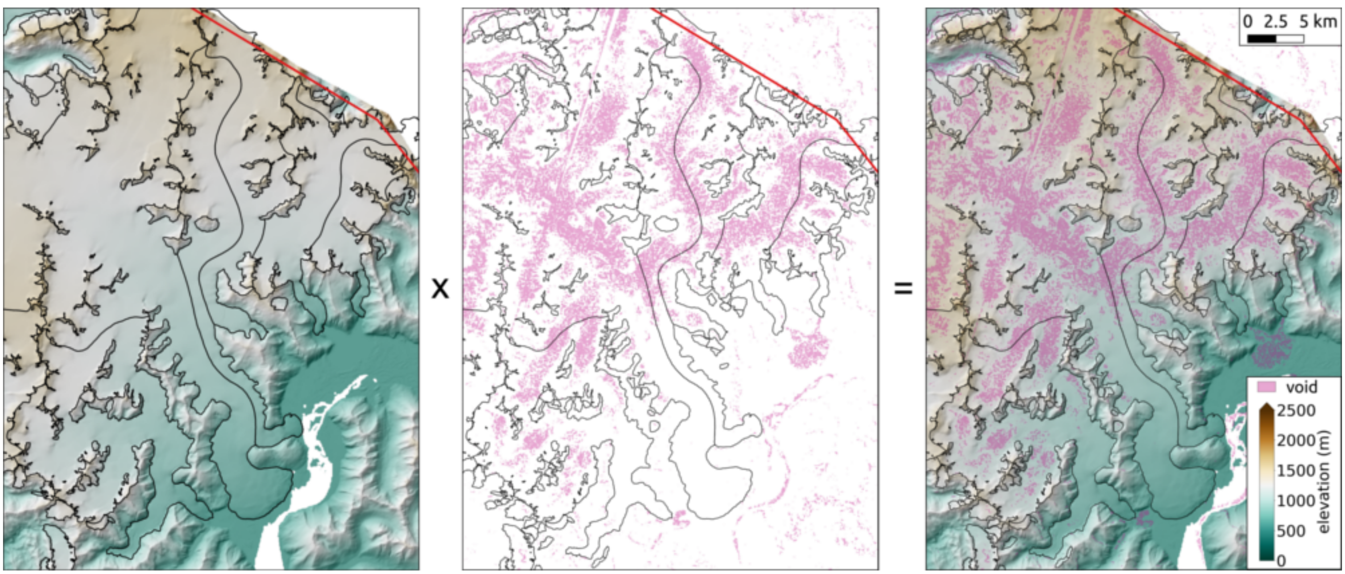

Figure 2. Example of masking procedure over Taku Glacier, with RGI outlines shown in black. The IfSAR DEM (left) is masked using the composite correlation mask from the ASTER products (middle), to produce a DEM (right) with holes (in gray) similar to expected voids in an optical DEM. In the middle panel, black represents areas where the ASTER correlation is below the chosen threshold.

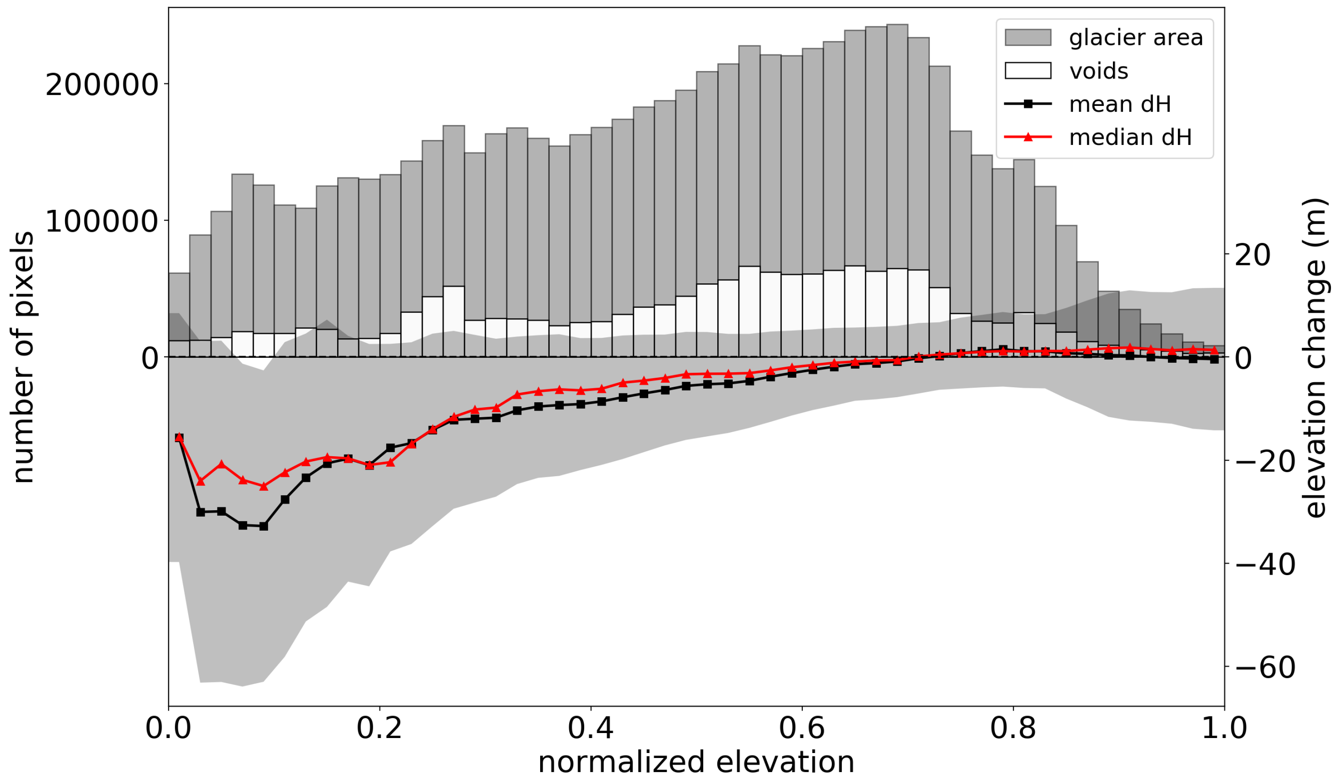

Figure 3. Distribution of glacier area and voids by normalized elevation, and mean and median elevation changes by normalized elevation. The shaded area around the mean elevation changes indicates the mean ± one standard deviation.

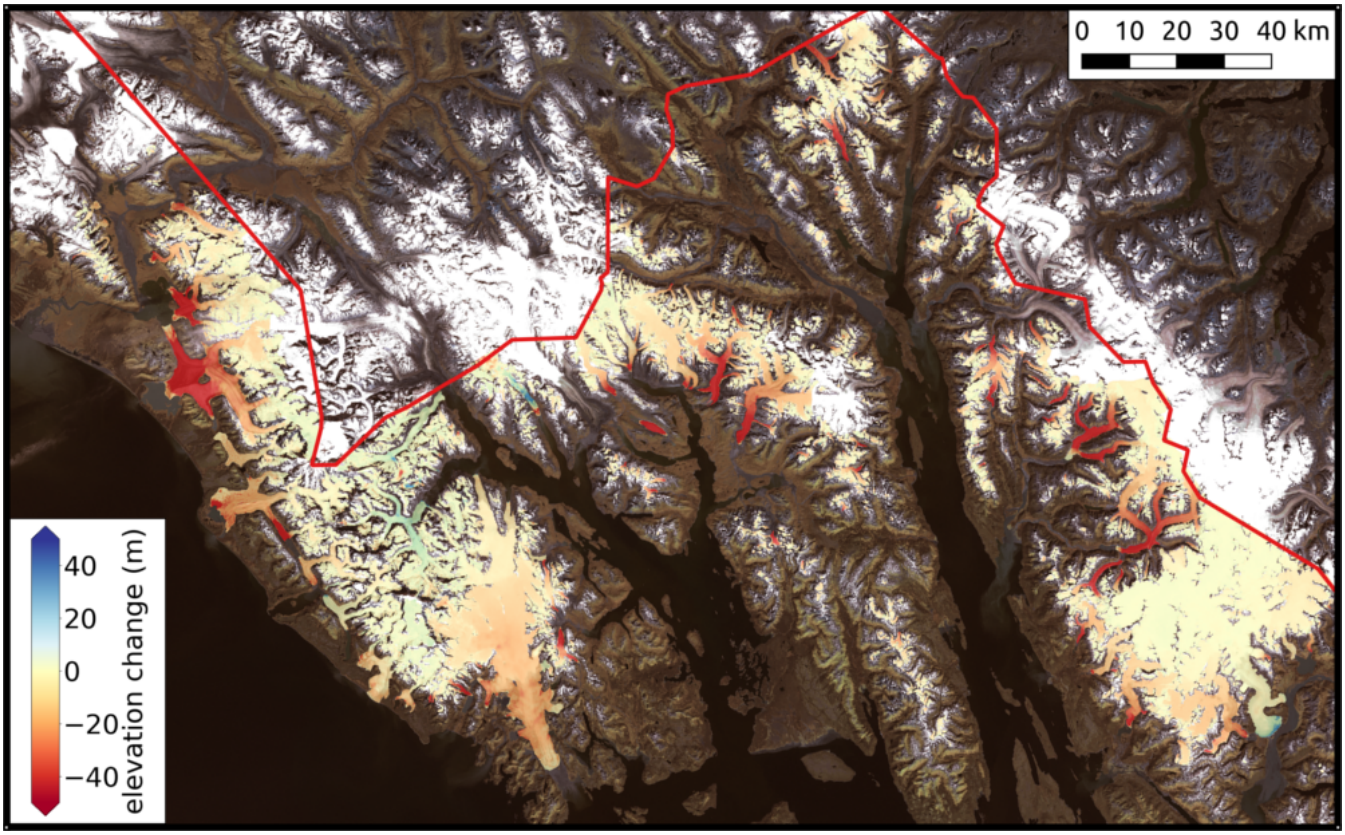

Figure 4. Non-voided, true elevation changes over the study area. Note contrasting patterns of thinning and thickening at lower elevations over glaciers labelled in Fig. 1., compared to the region in general. Background image is a mosaic of Landsat 8 scenes from 2013.

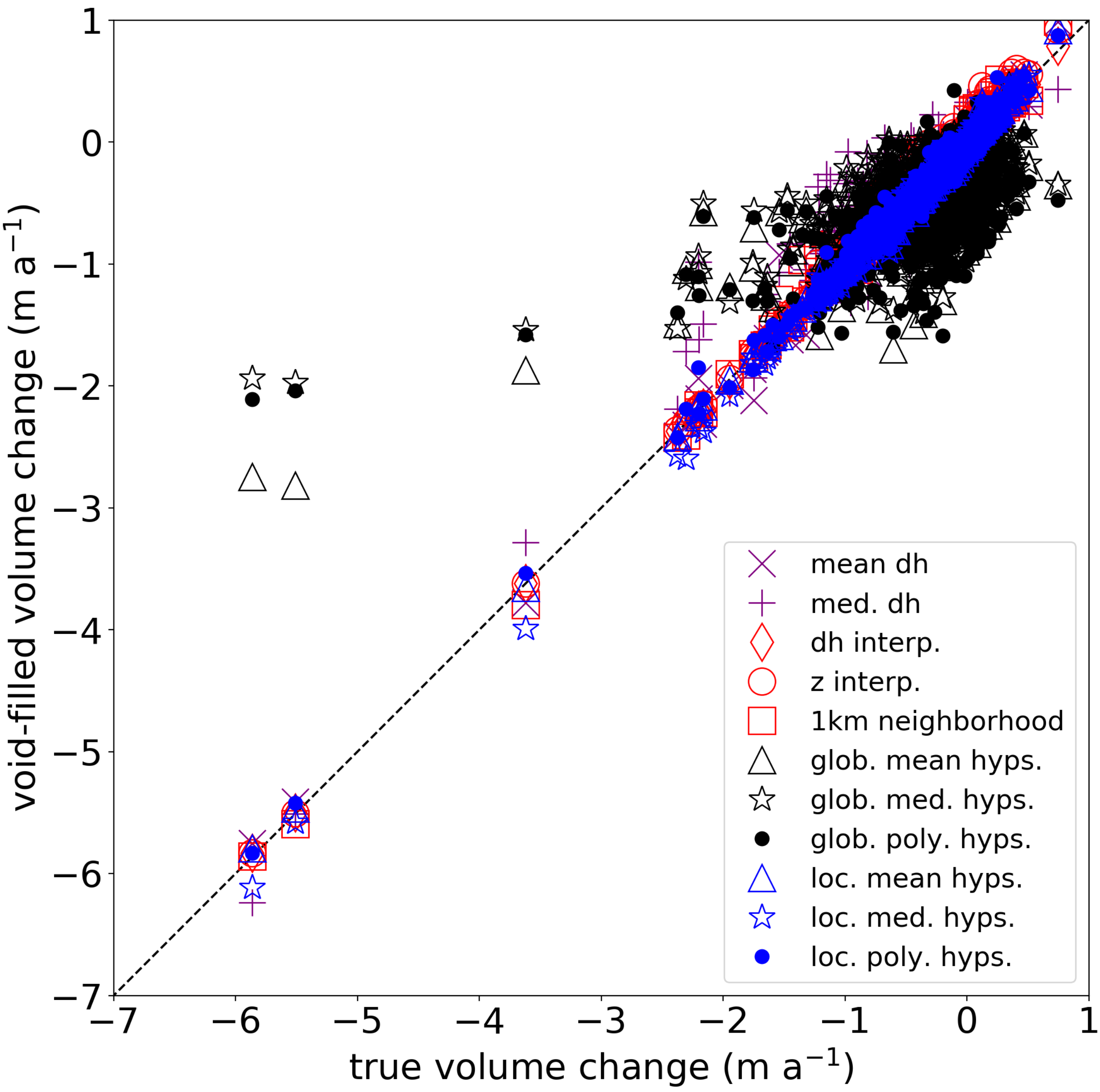

Figure 5. Void-filled volume change estimate for individual glaciers vs. true volume change.~~Inset shows detail around $x = [-1,1]$, $y = [-1,1]$.~~

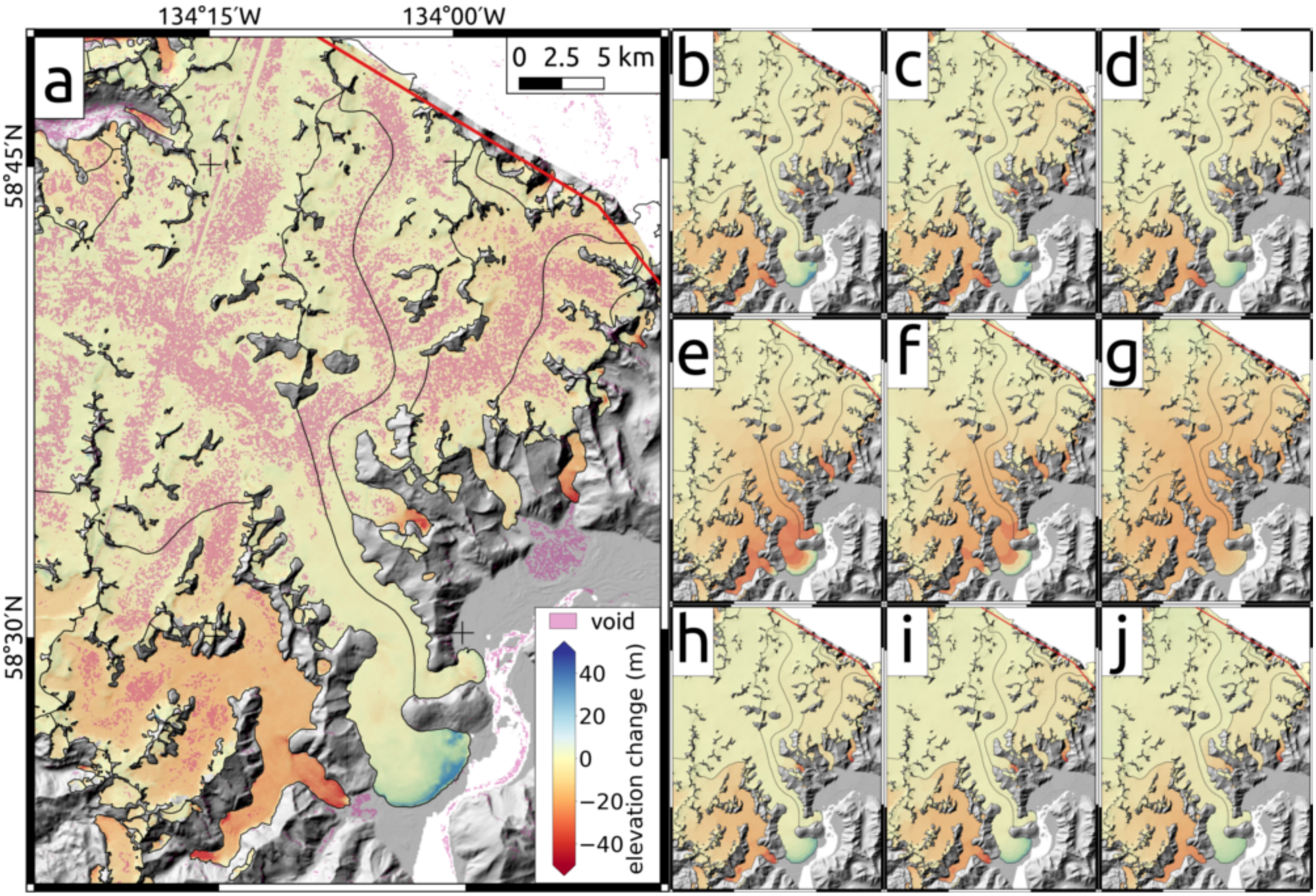

Figure 6. Final dH maps for Juneau Icefield and Taku Glacier. (a) Initial, non-voided dH; (b) dh Interpolation; (c) elevation interpolation; (d) 1 km neighbourhood; (e) global mean dH bins; (f) global median dH bins; (g) global polynomial fit; (h) local mean dH bins; (i) local median dH bins; (j) local polynomial fit. Note that the global interpolation schemes in panels e-g show primarily surface lowering, in contrast to the actual signal of no change or surface increase, as well as increased notability of individual glacier outlines in panels h-j.

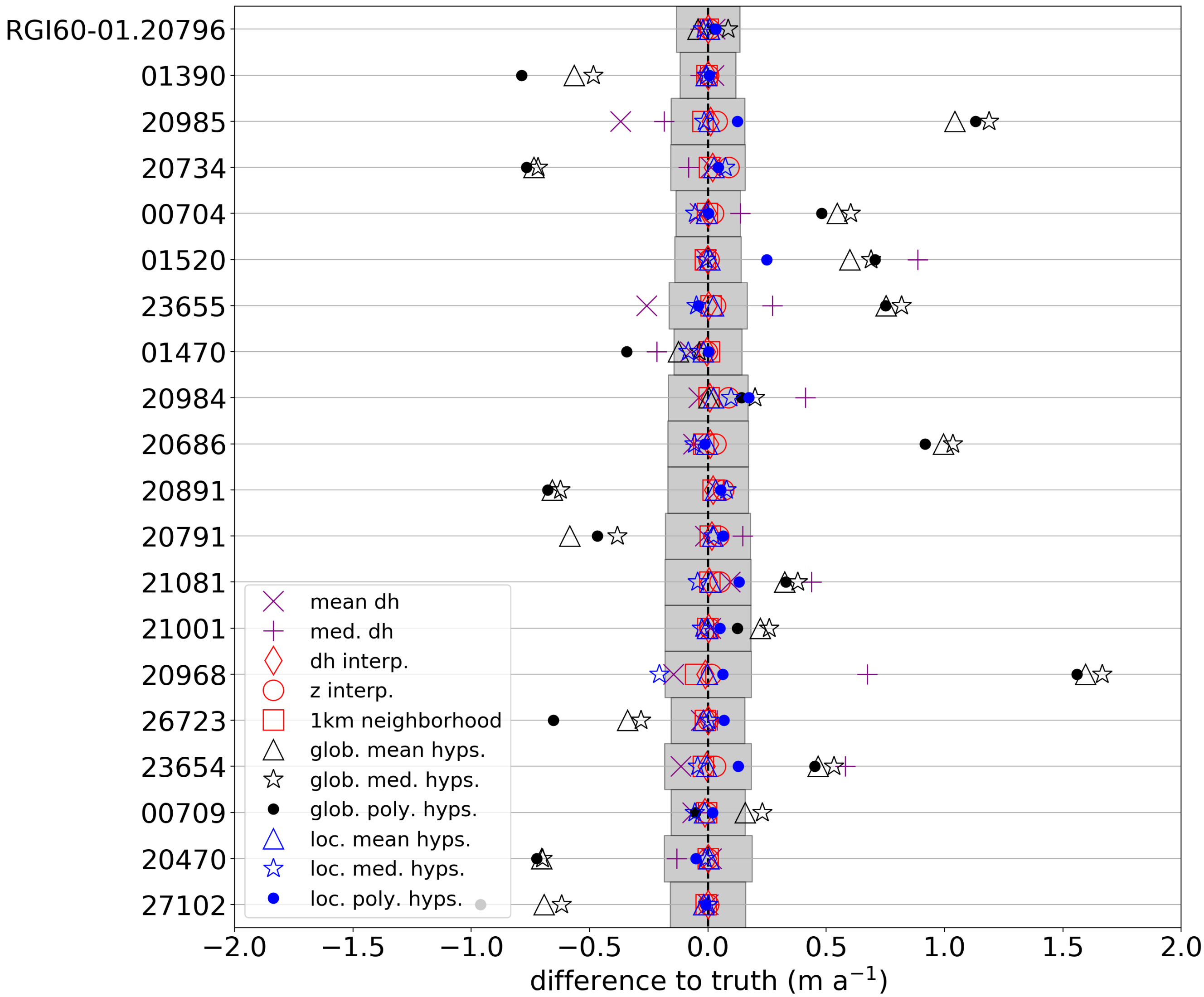

Figure 7. Comparison to true volume change for glaciers larger than 100 km^2, sorted by glacier area in descending order. Gray bars indicate uncertainty estimate for each glacier.

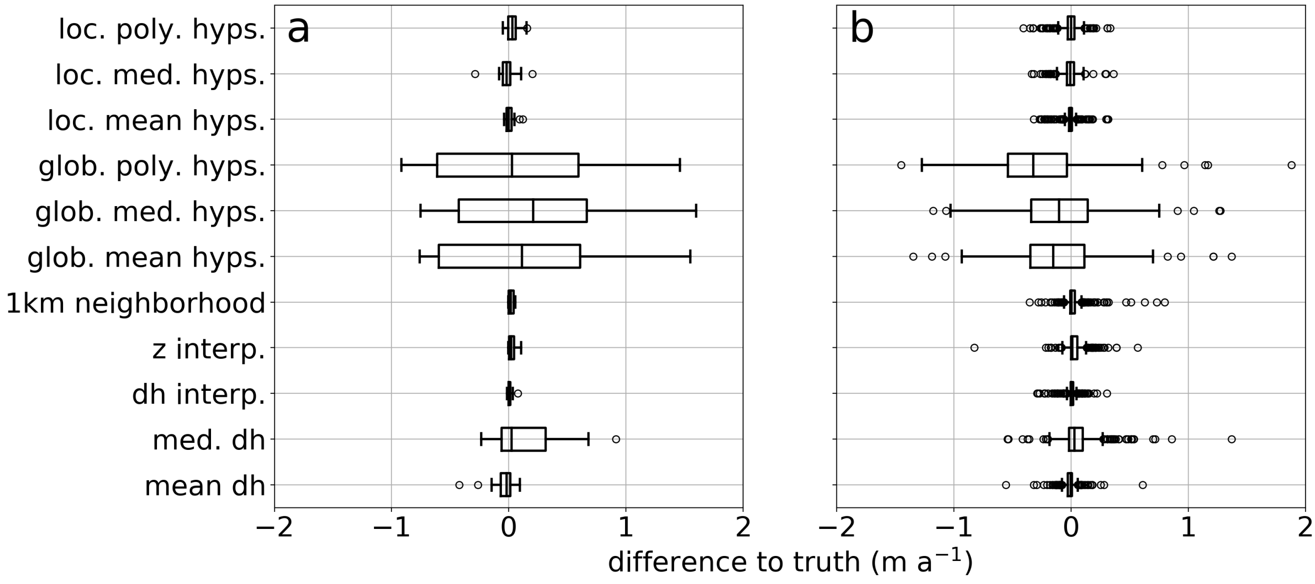

Figure 8. Difference to true volume change for each method tested, for (a) glaciers larger than 100 km^2; (b) all glaciers.

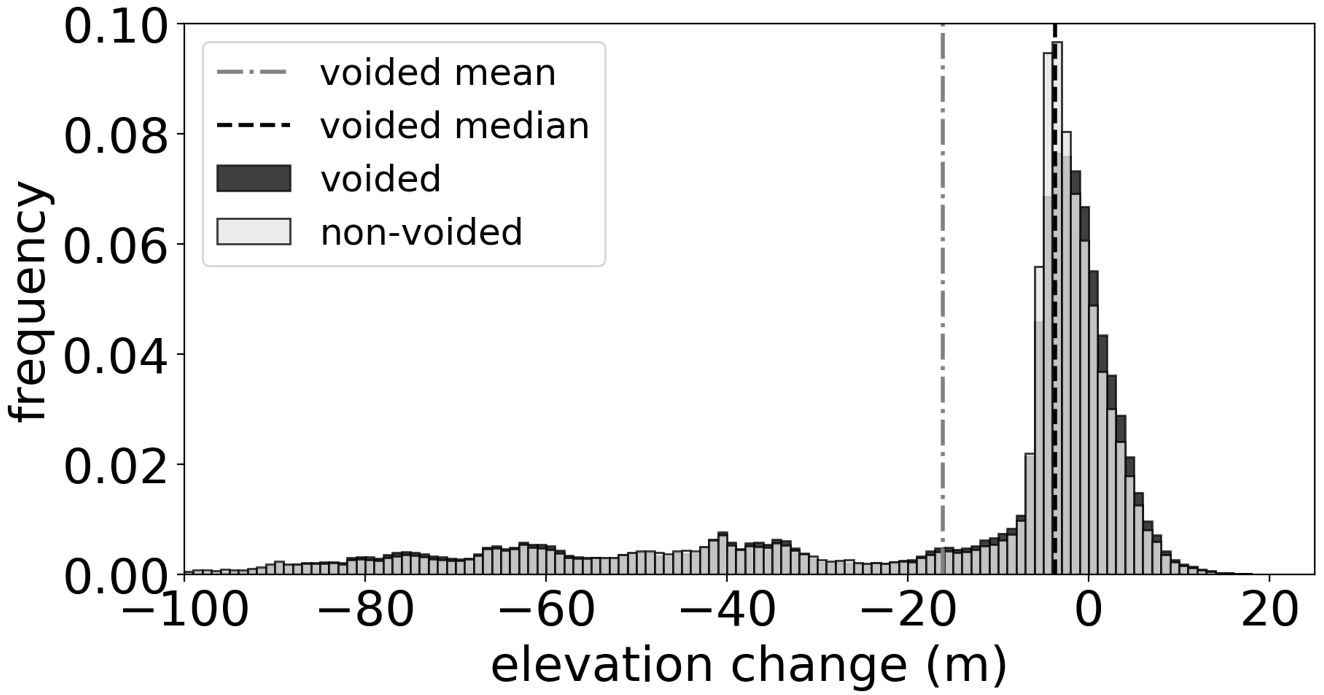

Figure 9. Distribution of elevation change values for voided and non-voided datasets over Field Glacier. Vertical lines indicate mean and median values for the voided dataset.

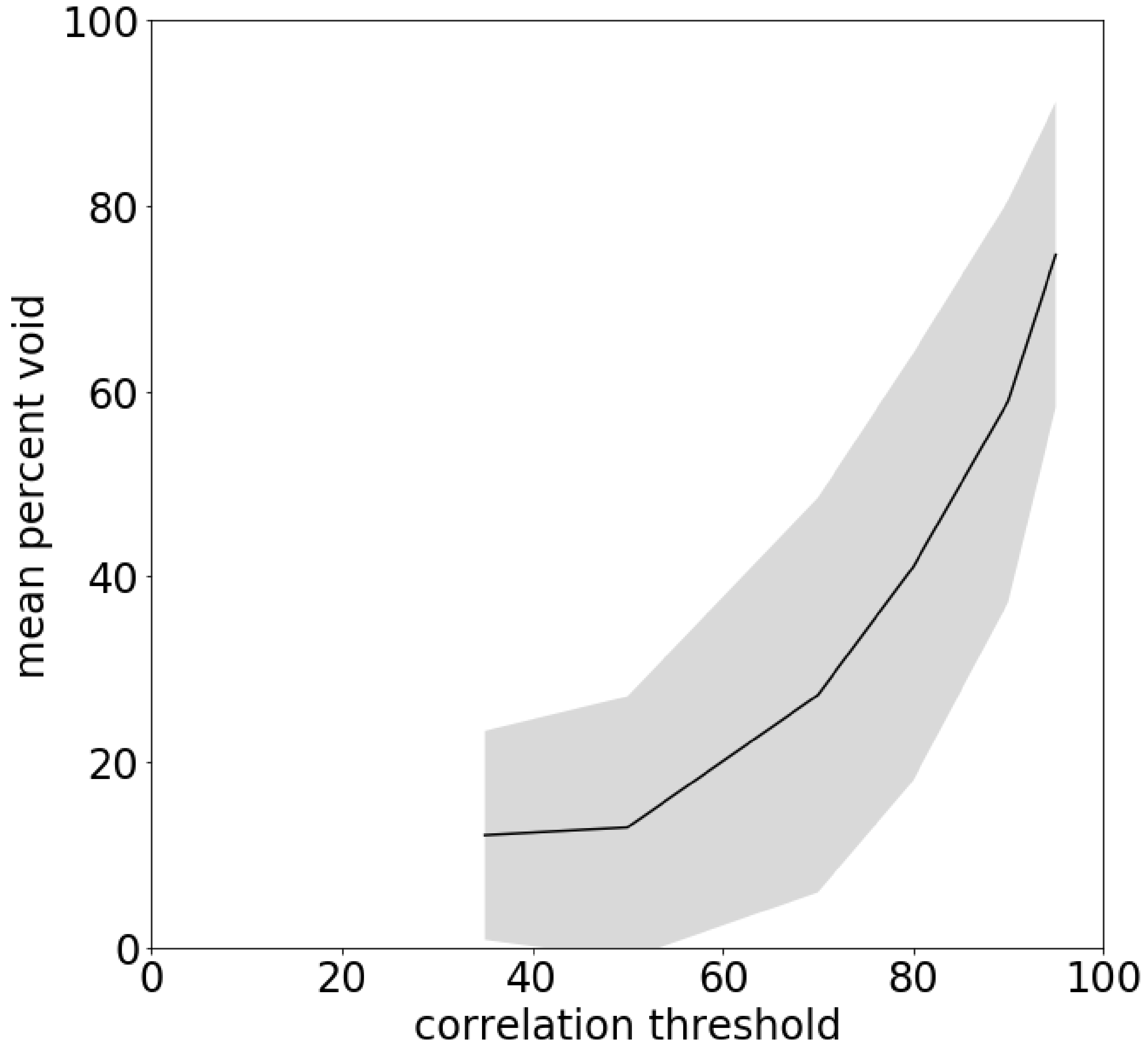

Figure 10. ~~Difference in volume change estimate~~ Mean (± standard deviation) percent void for each of the ~~largest 20 glaciers seen by changing~~ correlation ~~threshold from 50% to 70%~~thresholds investigated.

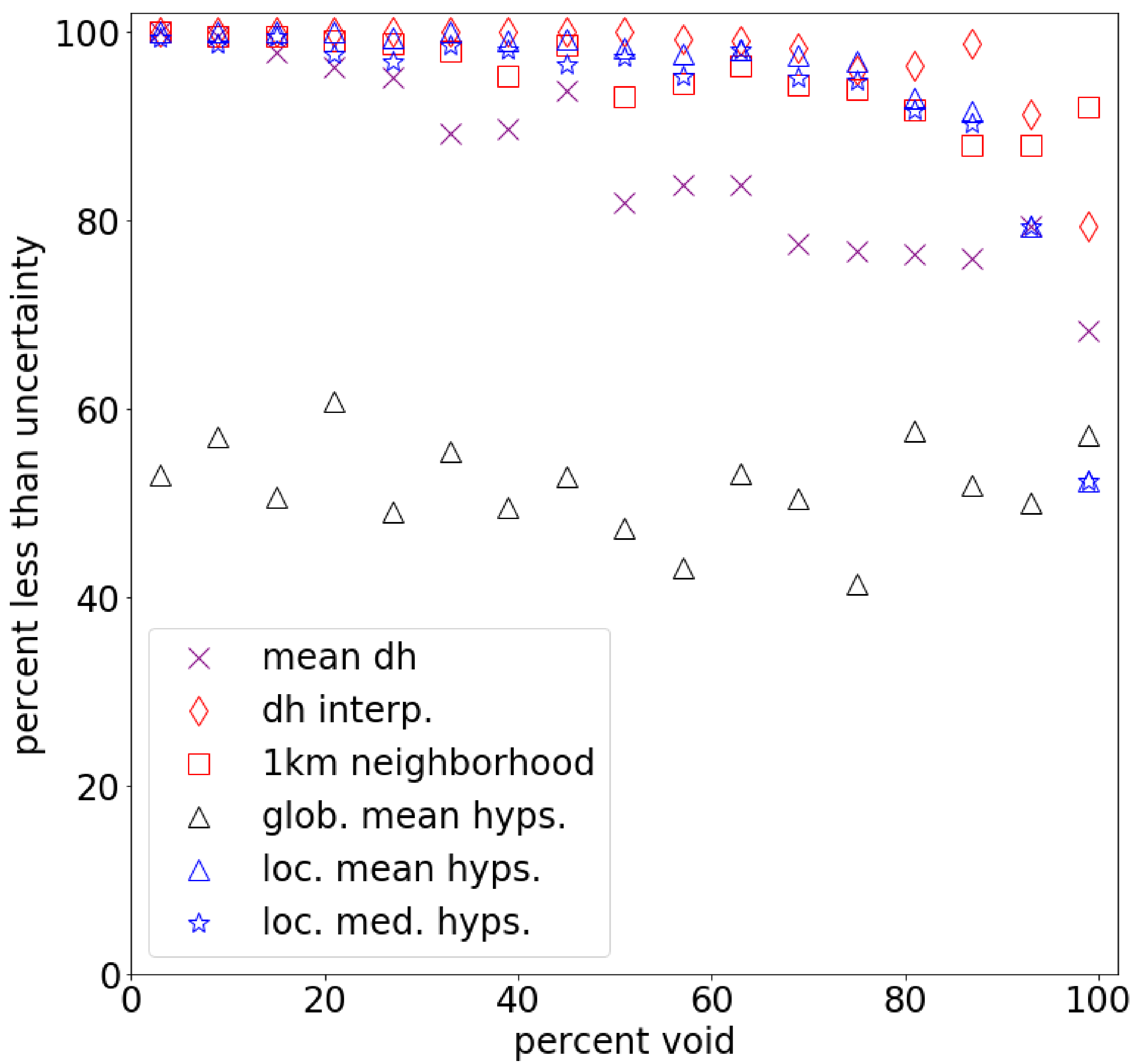

Figure 11. Percent of estimates falling within the uncertainty estimates for each glacier, as a function of the percent void.

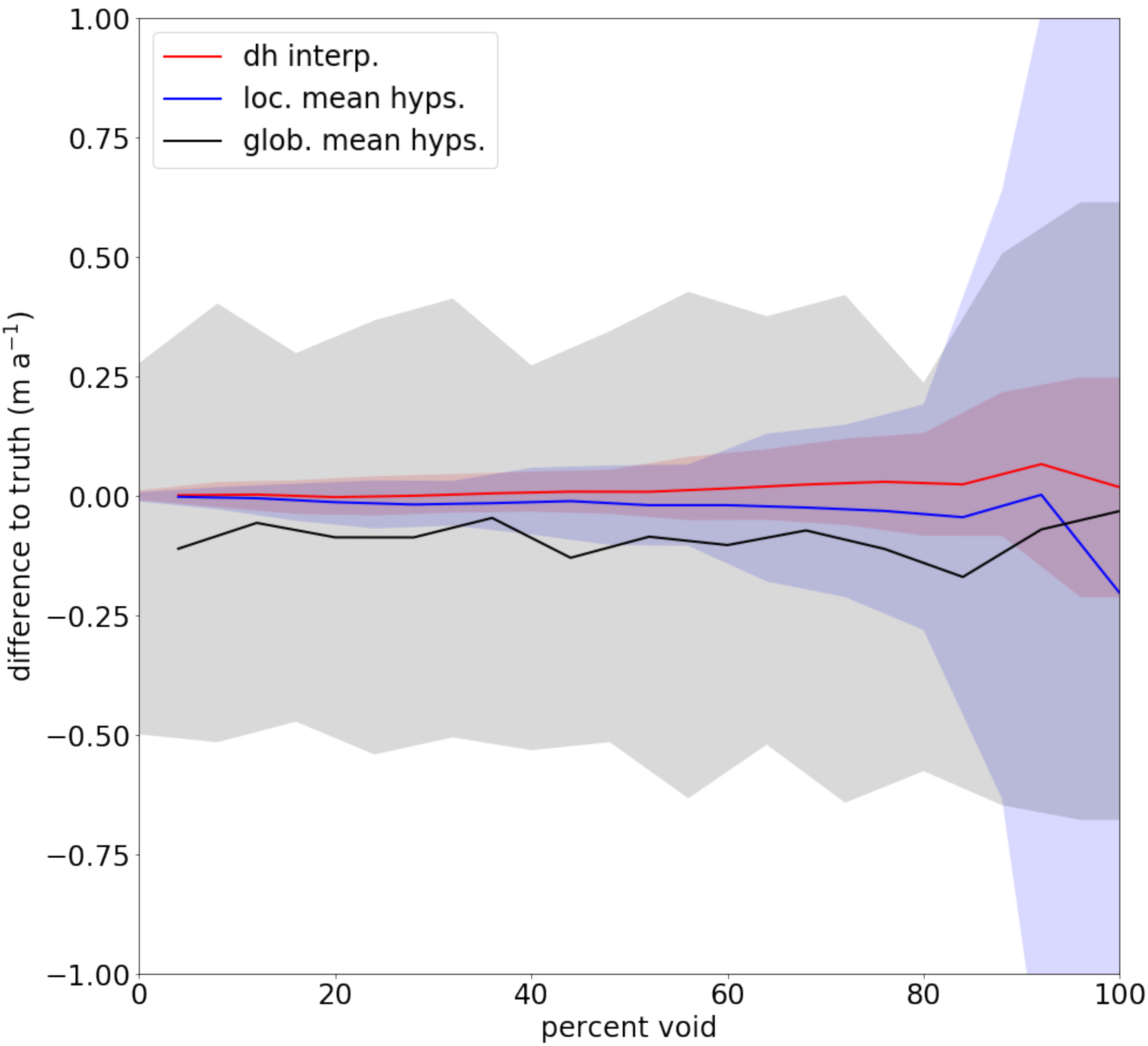

Figure 12. Difference to true mass balance as a function of void percentage, for the three best-performing methods on both an individual glacier and regional basis. Shaded region around each line indicates ± standard deviation.

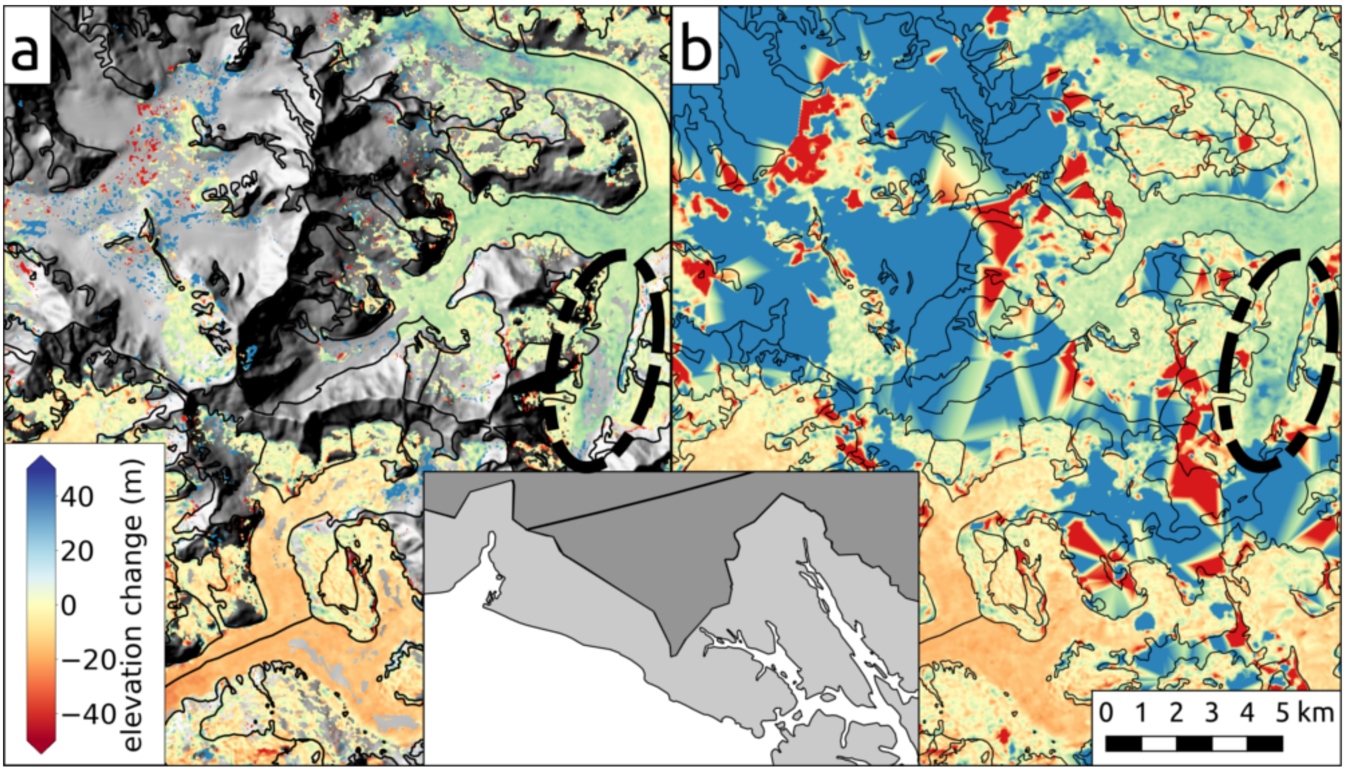

Figure 13. Elevation changes over the upper portion of Johns Hopkins Glacier, estimated by differencing an ASTER DEM acquired 13 August 2015 and the SRTM (a) with correlation-masked values left as nodata; (b) with voids filled using linear interpolation. Clear interpolation-related artefacts are seen in the sparsely-sampled accumulation area. Ellipses highlight an area over the glacier where linear interpolation performs well, with no obvious artefacts in the interpolated surface.

Table 1. Summary statistics for difference to true volume change for each method for the 415 glaciers sampled. All units in ~~km^3~~ m a^{-1}, except for "total diff", which is in units of km^3 a^{-1}, and "pct. uncert.", which indicates the percentage of glaciers for which the interpolated dV was within the uncertainty of the true volume change.

method	mean ± std	median	max	min	rms diff	total diff	pct. uncert.
~~Mean dH~~ mean dh	-0.01 ± 0.07	0.00	~~0.21~~ 0.61	~~-1.11~~ -0.37	0.07	~~-2.26~~ -0.18	~~94.41~~ 97.36
~~Median dH~~ med. dh	~~0.02~~ 0.07 ± ~~0.15~~ 0.19	~~0.00~~ 0.03	~~2.32~~ 1.22	~~-0.61~~ -0.53	~~0.15~~ 0.20	~~8.30~~ 0.69	~~77.85~~ 83.89
~~dH~~ dh interp.	0.00 ± ~~0.01~~ 0.03	0.00	0.13	~~-0.04~~ -0.16	~~0.01~~ 0.03	~~0.39~~ 0.01	~~97.99~~ 100.00
~~Z~~ z interp.	~~0.00~~ 0.03 ± ~~0.09~~ 0.06	~~0.00~~ 0.01	~~0.26~~ 0.34	~~-1.25~~ -0.20	~~0.09~~ 0.07	~~-1.85~~ 0.16	~~93.96~~ 99.52
1km neighborhood	0.00 ± ~~0.02~~ 0.08	0.00	~~0.12~~ 0.39	~~-0.02~~ -0.44	~~0.02~~ 0.08	~~1.78~~ -0.01	~~95.08~~ 99.28
~~Glob. Mean Hyps~~glob. mean hyps.	~~0.00~~ -0.08 ± ~~0.35~~ 0.45	~~0.00~~ -0.13	~~3.31~~ 3.12	~~-4.16~~ -1.10	~~0.35~~ 0.46	~~-0.40~~ 0.03	~~47.20~~ 52.64
~~Glob. Median Hyps~~glob. med. hyps.	~~0.01~~ -0.04 ± ~~0.35~~ 0.49	~~0.00~~ -0.08	~~3.65~~ 3.92	~~-3.57~~ -1.09	~~0.35~~ 0.49	~~4.47~~ 0.52	~~46.76~~ 51.92
~~Glob. Polyfit Hyps~~glob. poly. hyps.	~~-0.02~~ -0.25 ± ~~0.40~~ 0.51	~~-0.01~~ -0.32	~~3.17~~ 3.75	~~-5.38~~ -1.39	~~0.40~~ 0.57	~~-8.64~~ -0.39	~~34.90~~ 37.02
~~Loc. Mean Hyps~~loc. mean hyps.	0.00 ± ~~0.02~~ 0.03	0.00	~~0.24~~ 0.20	~~-0.17~~ -0.20	~~0.02~~ 0.03	~~0.09~~ 0.00	~~96.20~~ 100.00
~~Loc. Median Hyps~~loc. med. hyps.	~~0.00~~ -0.01 ± ~~0.04~~ 0.06	~~0.00~~ -0.01	~~0.33~~ 0.39	~~-0.41~~ -0.37	~~0.04~~ 0.07	~~-1.21~~ -0.09	~~95.97~~ 98.80
~~Loc. Polyfit Hyps~~loc. poly. hyps.	~~0.00~~ 0.01 ± ~~0.04~~ 0.06	0.00	~~0.38~~ 0.35	~~-0.17~~ -0.24	~~0.04~~ 0.06	~~1.74~~ 0.26	~~92.84~~ 97.84

Table 2. Difference to true volume change for the glaciers with the two largest changes. All units in km^3 m a^{-1}, except for pct void.

method	Taku Glacier	Field Glacier
Mean dH mean dh	0.18 0.03	-0.02 -0.01
Median dH med. dh	-0.22 -0.03	2.32 0.89
dH dh interp.	0.01 0.00	-0.00
Z z interp.	0.02 0.00	0.01 0.00
1km neighborhood	0.00 -0.01	0.11 -0.01
Glob. Mean Hyps glob. mean hyps.	-4.16 -0.57	1.59 0.60
Glob. Median Hyps glob. med. hyps.	-3.57 -0.49	1.81 0.69
Glob. Polyfit Hyps glob. poly. hyps.	-5.38 -0.79	1.71 0.71
Loc. Mean Hyps loc. mean hyps.	-0.04 -0.01	0.05 0.01
Loc. Median Hyps loc. med. hyps.	-0.03 -0.01	0.00 -0.01
Loc. Poly Hyps loc. poly. hyps.	-0.01 0.00	0.38 0.25
pct void	39.27	18.97