Review of Mc Nabb et al., The Cryosphere, October 2018

Mc Nabb 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.

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).

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.

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.

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.

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.

Specific comments.

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

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

1.5. Is "based" the best word here?

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).

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

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(1–4), 138–148, 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), 105–130, doi:10.1007/s10712-016-9394-y, 2017.

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.

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?

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.

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

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.

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"?

2.31 "basin" needed after "glacier"?

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.

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.

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

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

5.3 How did the authors handle clouds in ASTER?

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.

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)

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

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.

6.8 is 'original elevation' clear enough?

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.

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

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

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

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

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.

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).

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.

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.

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

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.

9.22 "Differences" of what?

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.

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.

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.

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?

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.

10.25 was dDEM defined already? (not sure)

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?

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

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

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

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?

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

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.

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

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

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.

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)?

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.