Re-review of the study by Nuimura et al.

1. General comments

In my opinion the ms has greatly improved compared to the first version (e.g. regarding structure, description of relevant details, and illustrations). With the better justification for the different definition of a glacier applied here (excluding steep parts), I can also live with the revised dataset. The authors have described and illustrated what the impacts of their specific glacier definition are compared to other datasets, also indicating that there might not be a universal right or wrong approach. I also appreciated the thoughtful point-to-point reply to all reviewers, indicating that the review effort was justified.

What I miss is the critical reflection about the impacts of the decisions taken or what happens in case of failure. The authors stay rather close to their datasets / methods and report what has been done with a limited consideration of alternatives. Below are some examples of issues that I think deserve additional discussion:

The authors explain their decision to remove steep (and partly also shadowed) glacier parts with the intended purpose of the inventory (determination of mass changes). While I assume this might work, it creates a problem when the dataset is included in global databases like GLIMS or the RGI. I agree with the authors that a related confusion of outlines in the accumulation area can also be found in current studies (see P21/22 of their first reply), but the question remains why adding yet another definition? There might be scientists investigating glaciers in a climate change context (e.g. area and length changes) or as a source of natural hazards. With the steep glacier areas removed this is likely not really possible. Furthermore, when glacier outlines vary with the extent of terrain shadow, our datasets might have a credibility problem. To avoid this, it could have been suggested that the final outlines will come with an attribute describing the intended purpose of the dataset and their limitations.

Another example is the use of contour lines for glacier delineation. Though this might work for many glaciers, not all shadowed and/or debris-covered glaciers are laterally delimited by steep rockwalls where 'clear turnoff points' (L212) can be identified. What has been done in these cases? On the other hand, it could have been mentioned that slope and curvature (as derived from a DEM) have been used in previous studies to determine glacier boundaries. And this could have resulted in a discussion of the benefits when using manual interpretation of elevation contour lines vs slope or curvature (first or second derivative) with a threshold.

The outline review process is now much better described, but for me a new question is now why analysts with field experience produce more accurate outlines than the others? Of course it is good to know how a glacier looks like in reality, but transferring this knowledge back to a 30 m pixel remote sensing perspective can also be challenging, in particular when even field-based determination of glacier boundaries fail. To my knowledge, best outlines result when different possibilities are jointly discussed with all analysts and a compromise is used.

The authors describe that they have used manual delineation for glacier mapping and excluded glaciers < 0.05 km^2. This is fine but requires a-priori knowledge about the glacier size, i.e. they have to be digitized first and then removed again? Or has - as described in the first response - the length and width of each glacier been measured beforehand to determine if it is larger or smaller than this threshold? This sounds like enormous work to avoid the use of automated mapping and a size threshold to automatically remove polygons smaller than this.

In the same sense I wonder about the conclusions (L425-430): The simple band ratio method works very well for a precise classification of ice and snow in shadow as several studies have shown. In contrast, the authors stress that only a field-experienced operator can improve the outlines without even mentioning the alternative possibility. I fully agree that mapping of ice and snow in shadow requires a careful selection of the best threshold, but it has been written already in the 2002 papers that thresholds should be selected in a way that manual editing is reduced to the largest possible extent, i.e. in shadow regions that can hardly be accurately mapped otherwise. I would suggest that the authors are at least testing this in a sub-region where precise classification suffers from shadow.

I stop here and provide some further comments in the specific comments section below. I hope the authors can address the points listed above and below to get the study acceptable.

2. Specific comments

L61: I think the new Chinese glacier inventory is available as well in the meantime. To avoid that your study is out-dated too soon, I suggest adding a note that only the old Chinese inventory was available for the RGI at the time of data compilation for GAMDAM.

L62: I suggest removing this statement as this is constantly improving. Otherwise it should be mentioned that this was the status at the time of the GAMDAM compilation and that efforts are taken to constantly improve it.

L122: I suggest writing "(excluding the ice sheets Greenland and Antarctica)."

L132: I would add here that also the ICIMOD Inventory is excluding very steep glaciers. Another possibility would be using the Frey et al. (2012) inventory for the western Himalaya (see link in L356 comment below).

L152: This argumentation is not really a working, as clouds and debris have to be corrected in both cases. If there is really an advantage of complete manual digitizing vs using automated classification of clean ice as a starting point for the further editing, the workload for removing the seasonal snow polygons (that are included with the automated method) should be higher than the digitizing of all clean glaciers. Is this really the case?

L163-166: I do not understand this: On the one hand basin polygons have not been derived automatically, ok. But in the next sentence basin polygons have been used to identify glacier divides. So there must be basin polygons. Where do they come from?

L193/4: '… do not occur': I would not say it that exclusive. This might be well the case for most of the glaciers, but there are also hundreds if not thousands with thick and changing ice masses on steep headwalls (e.g. hanging glaciers). They do not only provide snow avalanches but they also contribute ice (either by avalanching or by flow).

L212 and 222: 'exhibit clear turnoff points': Again, this might be true for most of the glaciers but certainly not for all. How have these been delineated?

L256: Just as a note of caution: It might be the case that the glacier parts that are finally considered in this study might no longer be seen as real glaciers.

L266: I am somewhat sceptical here. Validation might be difficult if others do not apply the same definition of a glacier. So whenever real changes of glacier extent should be followed with the outlines provided here, the same group of persons must provide the new outlines. All others would only derive virtual changes due to differences in interpretation. So I suggest writing this differently to be clear.

L277: ... and the missing parts of the accumulation region will also play a role.

L288: I assume excluding glaciers <0.05 km^2 is meant here?

L294: '… with sizes between 16 …'

L351: Field experience might not be very helpful when it comes to mapping of ice in shadow, as the related regions might be invisible from the ground.

L356: The comparison with the ICIMOD inventory might give a good agreement as they have excluded steep accumulation areas as well. Hence, the 'despite' does not really apply. Is there any chance of performing a comparison with an inventory that does not exclude steep parts a priory? Maybe the one we have created for GlobGlacier can be used instead (http://www.globglacier.ch/outlines/himalaya)?

L434-437: I think this is not the point. ICESat data are filtered in regions of steep terrain so that wrong elevation changes in this region are unconsidered anyway. Instead, we have observed that strong changes in the derived mass budgets result from errors in the ablation region, for example when the outlines have a shift, were generalized too much (e.g. including rock outcrops), or not precise enough (e.g. incl. the lateral moraine). So the general idea that precision of these mass change estimates can be increased when just removing the steep parts of the accumulation region does in my opinion not really hold.

L 438ff: I hope that area changes in this case will then not be due to changes in shadow or seasonal snow cover.

Figures and Tables

L575: Insert before 'Zero' that "Colours refer to the number of scenes used [N] and that zero (black squares) …"

Fig. 2: This caption is too long. Please move the interpretation (e.g. L585/6, L590-592) to the main text.

Fig. 6: Maybe apply some contrast stretching to these images to see something.

Fig. 8: The green outlines in the upper left part of the image look rather conservative. Would it be possible to add a Fig 8b image showing a screenshot with the outlines draped over the high-resolution imagery available in Google Earth (such as in Fig. 10b)?

Fig. 11: The text in L678-680 should be in the main text rather than in the figure caption. I do not see any glaciers in Fig. 11c.

Fig. 14: Can we agree to remove the regression equations in this figure? They make no sense in my opinion.

Fig. 16: I suggest removing the upper half of the figure, as it is not used.

Fig. 17: I suggest using a different symbol for the debris-covered glaciers (star or cross) and map them on top of the other class to be visible.

Table 2/3:

These tables are a little bit critical as the RGI will soon be updated with improved outlines, i.e. the provided numbers will no longer apply. The table captions should also mention the RGI if the values are reported.